

Crystallographic Fragment Screening on the *Shigella* Type III Secretion System Chaperone IpgC

Marina Gárdonyi, Christian Hasewinkel, Johanna Wallbaum, Jan Wollenhaupt, Manfred S. Weiss, Gerhard Klebe, Klaus Reuter,* and Andreas Heine



Cite This: *ACS Omega* 2023, 8, 46051–46065



Read Online

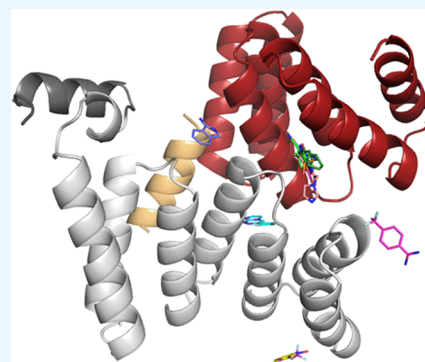
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The *Shigella* pathogenicity factor IpgC belongs to the class II of type III secretion system chaperones, whose members are characterized by a tetratricopeptide repeat (TPR) domain consisting of three and a half TPR motifs. Since IpgC is essential for *Shigella* virulence, we determined a high-resolution crystal structure of this chaperone to facilitate its use as a target for the structure-based design of anti-shigellosis compounds. The crystal structure revealed two possible homodimer assemblies, which strongly differ from the homodimer architectures so far known for IpgC and orthologues thereof. Through crystallographic fragment screening, we identified 10 small molecules that bind to IpgC and, therefore, are available for expansion to generate larger, more potent binders. A follow-up compound, based on one of our fragment hits, binds to a strictly conserved site, which overlaps with the binding site of the chaperone's substrates, IpaB and IpaC. Therefore, it constitutes a promising starting point for the design of functional IpgC inhibitors.



INTRODUCTION

Bacillary dysentery is a severe diarrheal disease, which is accompanied by the inflammation and ample destruction of the colonic epithelium. It is caused by bacteria of the genus *Shigella*, which constitute highly contagious pathogens invading and exploiting the epithelial cells of the colon by means of a type III secretion system. The latter comprises the actual type III secretion apparatus (also termed “type III translocon” or “injectisome”), the translocators IpaB and IpaC, a number of transcriptional regulators, specific chaperones, and about 25 effector proteins.¹ The ATP-dependent, syringe-like injectisome consists of the basal body spanning both the inner and outer bacterial membranes, of the external needle, and of the needle tip placed at its distal end. The invasion of the bacterium into the host cell depends on the secretion of the translocators IpaB and IpaC, which is triggered by the physical contact of the needle tip with the host cell.² Immediately after secretion, IpaB and IpaC are inserted into the host cell membrane, where they assemble to an approximately 25 Å wide pore, which remains attached to the needle tip. As a result, a continuous conduit is formed, connecting the bacterial cytoplasm with the host cytoplasm. It allows for the direct translocation of effector proteins into the host cell,³ where they act on numerous signal transduction pathways enabling the uptake of the bacterium by the host cell as well as its survival, spread, and propagation therein.

Within the bacterial cytoplasm, both IpaB and IpaC require the chaperone IpgC to prevent their early association and degradation.⁴ IpgC belongs to the class II of type III secretion

system chaperones, and accordingly, it is a member of the tetratricopeptide repeat (TPR) protein family. A number of crystal structures of IpgC and orthologues thereof have shown that the specific recognition of IpaB and IpaC by IpgC occurs through a so-called “anchor region” harboring the consensus sequence Pro/Val-Xxx-Leu-Xxx-Xxx-Pro near the translocator's N-terminus.^{5–10} In the chaperone/translocator complex, this anchor adopts an extended conformation, perfectly fitting into the cleft formed by the concave surface of the chaperone's TPR domain. However, the crystal structure of the *Aeromonas hydrophila* IpgC orthologue, AcrH, in complex with the soluble domain of the IpaB orthologue, AopB, shows that the chaperone/translocator assembly is by far not confined to this interaction. Instead, the major portion of the translocator interacts with the convex side of the chaperone's TPR domain resulting in a total protein/protein contact area of more than 3000 Å².¹⁰

As a prerequisite for the efficient secretion of IpaB and IpaC, the cytoplasmic “escort protein” Spa13 directs the IpgC/IpaB and the IpgC/IpaC complex, respectively, to the injectisome's ATPase, Spa47. This process depends on the interaction of Spa13 with IpgC, which was verified by Cherradi et al.¹¹ In

Received: September 15, 2023

Revised: October 25, 2023

Accepted: October 26, 2023

Published: November 20, 2023



addition, IpgC was shown to interact with both Spa47 and the cytoplasmic “gatekeeper”, MxiC.¹² The latter prevents the premature translocation of virulence factors and insures that the secretion of translocators occurs prior to that of effector proteins.¹³ In this context, MxiC promotes the secretion of translocators, probably involving its interaction with Spa47, immediately after activation of the injectisome. While the efficient binding of MxiC to Spa47 requires the ATPase to be present in complex with IpgC, no interaction is detectable between Spa47 and a preformed IpgC/MxiC complex.¹² However, the significance of this phenomenon in secretion hierarchy still has to be figured out.

In addition to its function as a chaperone, IpgC acts as a “co-activator” of gene expression. The expression of genes encoding “late” effectors, which are not yet required for and during the entrance process but enable the bacterium to survive, spread, and multiply within the host cell, relies on the AraC-like transcriptional activator MxiE. To become functionally active, MxiE has to form a complex with IpgC.¹⁴ Prior to the injectisome’s activation, the vast majority of IpgC is obviously captured by IpaB and IpaC and, thus, not available for any interaction with MxiE. Moreover, the “anti-activator”, OspD1, stabilized by the chaperone Spa15 is bound to MxiE, preventing it from being activated even in the presence of a small amount of free IpgC.¹⁵ Only upon contact with the host cell are IpaB, IpaC and, with short delay, OspD1 removed from the bacterial cytoplasm via secretion. As a result, both IpgC and MxiE are released for the formation of the active complex allowing the transcription of late effector genes.^{13–15}

Given the essential functions of IpgC for virulence and the fact that the inactivation of the *ipgC* gene results in an apathogenic *Shigella* mutant,⁴ we deem this protein an attractive target for the design of active agents against bacillary dysentery. Yet, the deposited crystal structures of IpgC available at the beginning of the project were of moderate resolution only (PDB IDs: 3GYZ: 2.15 Å, 3GZ1: 2.15 Å,⁵ 3GZ2: 2.65 Å,⁶ 3KS2: 3.30 Å¹⁶). Accordingly, as a precondition for rational drug design, we first established a protocol that reproducibly yields well-diffracting IpgC crystals. The structure resulting from our new crystal form, featuring space group $P3_221$ and showing a resolution of 1.58 Å, revealed two potential homodimer assemblies that had not been observed for IpgC before. A subsequently performed crystallographic fragment screening resulted in 10 hits, binding at sites well distributed over the entire structure. The expansion of one of these fragments led to a compound that binds into a cleft formed by the concave surface of the TPR domain. Since this cleft is essential for the binding of IpaB and IpaC, the discovered compound may be a promising starting point for the design of drug candidates that block the access of translocators to the chaperone.

RESULTS

High-Resolution Crystal Structure of Apo IpgC^{10–151}

While full-length recombinant IpgC appeared to be recalcitrant to crystallization, we were able to grow well-diffracting crystals of a slightly truncated IpgC variant lacking nine N-terminal and four C-terminal amino acid residues (IpgC^{10–151}). These crystals yielded diffraction data of high quality, which allowed structure determination via molecular replacement using the crystal structure of IpgC published by Lunelli et al.⁵ as a template (PDB ID: 3GYZ). The final model contains two copies of IpgC^{10–151} per asymmetric unit, hereafter termed

“chain A” and “chain B”. The structure was refined at a resolution of 1.58 Å with R and R_{free} amounting to 17.5 and 21.5%, respectively (data collection and refinement statistics are summarized in the Supporting Information, Table S1). As reported previously, the all- α -helical IpgC consists of a single tetratricopeptide repeat (TPR) domain, which comprises three and a half consecutive TPR α helix pair motifs (α helices 2 to 8)^{5,16} with α helix 8 being referred to as the “stability helix”.¹⁰ While there is no electron density assignable to the N-terminal residues, 10 to 23, of chain A, the corresponding segment of chain B is well defined. It displays an additional α helix (α helix 1) reaching from Ser10 to Asn20. Moreover, electron density is visible for the very N-terminal residue Gly9, which is, however, not present in the natural IpgC but results from the cloning procedure (Figure S1). An extended loop, which contains a short 3_{10} helix, connects α helix 1 to the TPR domain.⁵

Through size exclusion chromatography and multiangle laser light scattering, it was shown in former work that, in the absence of IpaB, IpaC, or any further interaction partner, IpgC shows a high tendency to form a homodimer.^{5,16} Indeed, our crystal structure reveals two potential, nonsymmetrical homodimer versions, which are formed by chain A and chain B, or by chain A and a symmetry-related copy of chain B, respectively (Figure 1). In the first putative dimer, α helix 5,

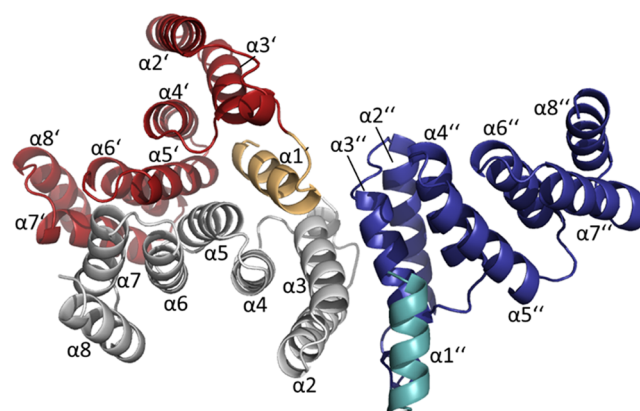
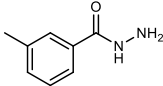
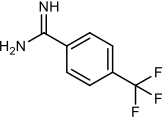
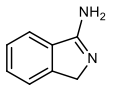
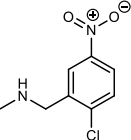
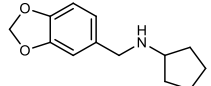
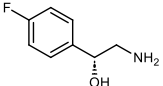
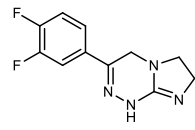
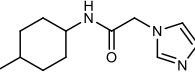
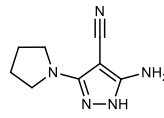
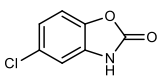


Figure 1. Two possible homodimer arrangements are observed in the crystal structure of IpgC^{10–151}. A “back-to-back” dimer of the two molecules in the crystallographic asymmetric unit is formed by chain A, which is depicted in gray, and chain B (indicated by a ’), which is shown in ocher (amino acids 9–23) and red (amino acids 24–150). In addition, a “head-to-head” dimer is formed by chain A and a symmetry mate of chain B (indicated by a ’’) which is colored in cyan (amino acids 9–23) and blue (amino acids 24–150). Protein molecules are shown in a ribbon representation.

turn $\alpha 5\alpha 6$, and α helix 6 of chain A interact with the same structural elements of chain B. In addition, α helix 1 of chain B is embedded in a groove formed by α helices 3, 4, and 5 of chain A. As calculated by the program PISA,¹⁷ this assembly leads to an interface area of 1023 Å². It may be referred to as “back-to-back” arrangement since dimer formation takes place via the convex (“back”) sides of both TPR superhelices. In total, 131 van der Waals interactions are formed across the interface of this dimer as determined by the program *Contactsym*.^{18,19} Furthermore, four subunit-bridging polar interactions, namely, one salt bridge and three additional hydrogen bonds (H bonds), frame the interface (Figure S2A,B).

Table 1. Crystallographic IpgC Binding Fragment Hits^a

| Fragment ID | Chemical structure | PDB ID | Resolution [Å] | Fragment ID | Chemical structure | PDB ID | Resolution [Å] |
|-------------|--|--------|----------------|-------------|---|--------|----------------|
| J1 |  | 7NL8 | 1.59 | J33 |  | 7AZV | 1.68 |
| J2 |  | 7AYW | 1.78 | J36 |  | 7O04 | 1.69 |
| J6 |  | 7O6S | 1.58 | J45 |  | 7NRG | 1.57 |
| J11 |  | 7B1U | 1.59 | J52 |  | 7PEO | 1.50 |
| J20 |  | 7AXY | 1.63 | J66 |  | 7NHW | 1.92 |

^aThe formula shows the *R* enantiomer of J45 as bound into the interface pocket 1. However, the racemate is provided by the Frag Xtal Screen library.

The second potential homodimer in our crystal structure features a subunit interface area of 851 Å² established by 132 subunit-bridging van der Waals interactions, three salt bridges (one of them involving the artificial N-terminal Gly9 of the recombinant IpgC^{10–151}), and one additional H bond (Figure S2C,D). The assembly of subunits therein may be referred to as “head-to-head” arrangement since it is established by interactions between the N-terminal (“head”) sections of both subunits. These include the visible N-terminal part of chain A up to α helix 3, as well as α helices 1, 2, and 3 of chain B (Figure 1). Most hydrophilic interactions between both subunits involve α helix 1 of chain B, which contacts turn $\alpha 2\alpha 3$ and the N-terminus of α helix 3 of chain A (Figure S2C). This stands in contrast to the “back-to-back” homodimer of our crystal structure, where this helix predominantly makes hydrophobic contacts with chain A.

Crystallographic Fragment Screening. To search for small molecules that bind to IpgC, we performed a crystallographic fragment screening using the Frag Xtal Screen library²⁰ documented in detail in the contribution of Hassaan et al.²¹ Each of the 96 fragments was individually soaked at a concentration of 100 mmol L⁻¹ into crystals of apo IpgC^{10–151}. Screening resulted in 10 hits with crystal structure resolutions ranging from 1.50 to 1.92 Å (Table 1) corresponding to a hit rate of about 10%. The 10 hits were found at five distinct binding sites, with three fragments, J2, J33, and J36, being observed as copies at more than one site. Difference omit maps for all observed fragment hits are shown in the Supporting

Information (Figure S3). In the following, the binding modes of individual fragments to the target protein are described.

The most frequently occupied binding pocket is formed by the amino acids connecting α helices 5 and 6 of chain A and a shallow groove composed of α helices 5' to 7' of chain B from the putative “back-to-back” dimer. Hereafter, we refer to this binding pocket as “interface pocket 1”. It is lined by the side chains of Phe97, Lys101, Asn102, Tyr104, and His133 from chain A as well as by the side chains of Ala94', Phe97', Tyr104', Val107', Thr110', Gln114', Glu125', Cys126', and Leu129' from chain B (structural elements and residues from chain B are indicated by a') (Figure 2A).

Two copies of fragment J2 bind to IpgC^{10–151}, with one of them being located in interface pocket 1 with full occupancy. With its aromatic ring, J2 forms van der Waals interactions with a number of mainly hydrophobic amino acids of the pocket in addition to an edge-to-face π – π stacking to the phenyl ring of Phe97'. Through its ammonium group, it establishes a salt bridge to the carboxylate group of Glu125' and a water-mediated interaction with both the side chain amide of Gln114' and the sulfhydryl group of Cys126'. Moreover, through its isoindoline nitrogen atom, it forms a water-mediated interaction with the hydroxyl group of Tyr104 (Figure 2B). Obviously triggered by the binding of J2, the imidazole of His133, which forms H bonds to the side chain hydroxyl groups of both Tyr104 and Tyr104' in the apo structure (Figure 2A), is rotated by about 90°.

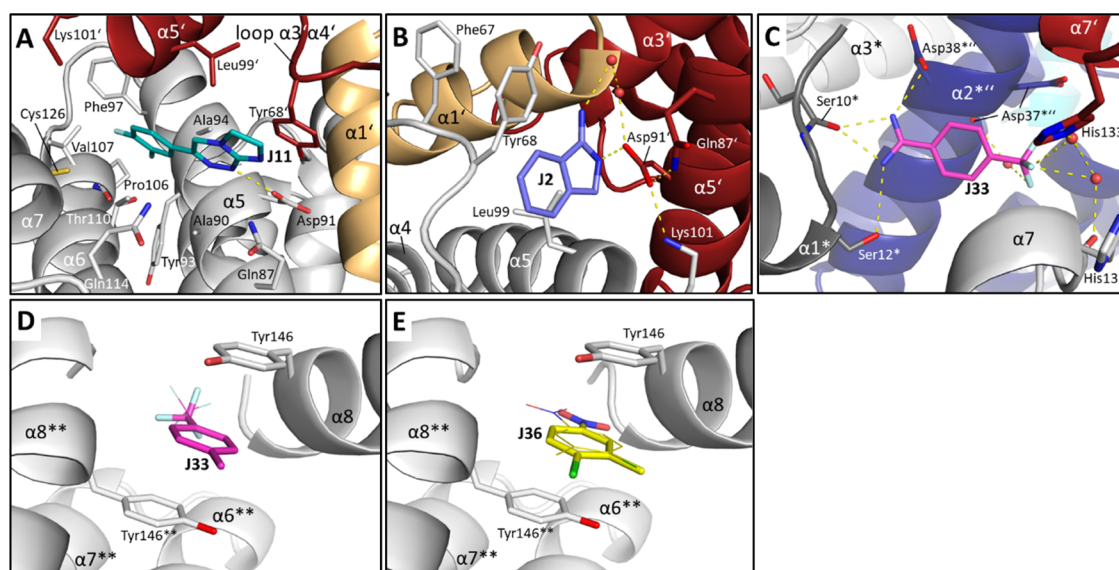


Figure 3. Binding mode of the fragments which are not accommodated by interface pocket 1. The color coding of chains A and B is identical to that in Figures 1 and 2. H bonds are indicated by dotted lines. Water molecules are shown as red spheres, and relevant amino acids are depicted as sticks. (A) Fragment J11 bound to the interface pocket 2. The carbon atoms of the ligand are colored cyan, fluorine atoms pale cyan. The side chains of amino acids that line the pocket are shown as sticks. (B) The second binding site of J2 (carbon atoms in violet) is on the surface of the “back-to-back” dimer. (C) One copy of J33 interacts with four different protein chains. Secondary structure elements and amino acid residues from a symmetry-related “head-to-head” dimer are indicated by a * (the N-terminus and α helix 1 from chain A of this dimer are colored dark gray). The carbon atoms of the ligand are colored magenta, and fluorine atoms pale cyan. (D) A second copy of J33 (carbon atoms magenta, fluorine atoms pale cyan) is sandwiched between Tyr146 of chain A and the same amino acid of chain A from a symmetry mate (Tyr146**, white). The symmetry equivalent of the fragment, caused by a 2-fold rotational axis running through it, is shown as thin lines. (E) Fragment J36 (carbon atoms yellow, chlorine atom green) binds to the same site as J33. One of both symmetry equivalents caused by the 2-fold rotational axis is shown in thin lines.

interface pocket 1, the side chains of Lys101 and Asn102 are largely invisible. The same holds true for the side chain of Glu125'. The open conformation of the pocket allows the binding of a DMSO molecule derived from the fragment solution, whose presence is confirmed by the anomalous signal of its sulfur atom. Since the placement of this DMSO strictly depends on the open conformation of interface pocket 1, its occupancy was set to 53%. With its oxygen atom, it forms a H bond to the main chain amide of Asn102 in the open conformation (Figure 2E).

The occupancy of fragment J6 was refined to 93%. Its 1,3-benzodioxole moiety makes close van der Waals contacts to the side chain isopropyl moiety of Val107'. One oxygen atom of the heterocycle is in H bond distance to the side chain hydroxyl group of Thr110'. However, the latter might rather form a H bond to the main chain carbonyl oxygen atom of Pro106' since ether oxygen atoms are only weak H bond acceptors. Through its amino group, J6 forms a water-mediated interaction with the sulfhydryl group of Cys126'. While the side chain of Lys101 is largely disordered in both the closed and the open conformation of interface pocket 1, the side chain of Asn102 is well defined in the open conformation, which is refined to 52% occupancy. Possibly caused by the cyclopentyl moiety of this fragment, the Glu125' side chain is disordered. Moreover, the side chain imidazole of His133 adopts a second conformation in which it is rotated by approximately 90°. The side chain of Leu129' is also present in two conformations, which differ by a ca. 120° rotation about the covalent bond connecting $C\beta$ and $C\gamma$ (Figure 2F).

Fragment J45 is bound with an 85% occupancy. This fragment also partially induces the open conformation of interface pocket 1, whose occupancy was refined to 55%. Through van der Waals interactions with the protein, the para

fluoro-substituted phenyl ring is placed in the hydrophobic interior of the pocket. With its hydroxyl group, J45 forms an H bond to the main chain carbonyl oxygen atom of Lys101 within the closed conformation of turn $\alpha 5\alpha 6$ as well as a water-mediated interaction with the hydroxyl group of Tyr104. A second water-mediated interaction is formed by the ammonium group of J45 with the main chain carbonyl oxygen atom of Lys122'. Ultimately, J45 indirectly interacts with the main chain carbonyl oxygen atom of His133 via two water molecules. Both in the open and closed conformations of interface pocket 1, the side chains of Lys101 and Asn102 are largely invisible in the electron density map. Moreover, the binding of this ligand may cause the disorder of the Glu125' side chain (Figure 2G).

The occupancy of fragment J66 was also refined to 85%. Similar to the case for J45, its halogen-substituted phenyl ring is accommodated by the hydrophobic interior of the pocket. An H bond formed between the ligand's amide and the hydroxyl oxygen atom of Tyr104 constitutes the only polar interaction between J66 and the protein (Figure 2H). The side chains of Lys101 and Asn102 are visible in neither the closed nor the open conformation of interface pocket 1, whose occupancy was refined to 44%. In addition, the side chain of Glu125' is disordered upon ligand binding.

J36 also binds into the interface pocket 1, where partial electron density indicates its presence (Figure S4). Moreover, the binding of this ligand to IpgC^{10–151} causes interface pocket 1 to entirely adopt the open conformation. There is no electron density assignable to its closed conformation, which would obviously cause a clash with the bound ligand. Nonetheless, the $mF_o - DF_c$ electron density observed for J36 within interface pocket 1 turned out to be too poor to allow for the unambiguous placement of the ligand into this pocket. As a

consequence, it was not placed in this position in the structural model.

In several crystal structures of IpgC^{10–151} in complex with a fragment that is bound into the interface pocket 1, some residual mF_o-DF_c difference electron density was still observed in the binding pocket after the refinement of the ligand had been completed. We assumed that this density might be caused by a DMSO molecule bound with a low occupancy. Since DMSO may be considered as a fragment as well, we recorded a data set from a crystal of apo IpgC^{10–151}, which had been soaked in a solution containing 10% (V/V) DMSO. Indeed, at the respective site, the resultant crystal structure (PDB ID: 7PEF) reveals a DMSO molecule, which is present in two orientations mainly differing in the position of the sulfur atom. With its oxygen atom, this DMSO molecule forms several water-mediated interactions, inter alia with the side chain amide of Gln114'. The water molecule involved in this interaction mediates further interaction of the DMSO oxygen atom with the sulfhydryl group of Cys126'. Ultimately, two water molecules mediate the interaction of the DMSO oxygen with the hydroxyl group of Tyr104 and with the carboxylate group of Glu125', respectively (Figure 2I).

With 80% occupancy, fragment J11 binds into a second pocket formed by the subunits of the “back-to-back” dimer, which we refer to as “interface pocket 2”. It is framed by α helices 5 to 7 of chain A as well as by loop $\alpha 3' \alpha 4'$ and the C-terminus of α helix 5' of chain B. Its interior is lined by the side chains of Gln87, Ala90, Asp91, Tyr93, Ala94, Phe97, Pro106, Val107, Thr110, Gln114, Cys126, Tyr68', Leu99', and Lys101' (Figure 3A). One fluorine atom of J11 is at a H bond distance to the hydroxyl group of Thr110 (not shown). However, the latter is rather likely to form an H bond with the carbonyl oxygen atom of Pro106. Through its NH group, J11 forms a charge-assisted H bond to the side chain carboxylate of Asp91. The binding of J11 causes the partial disorder of the Lys101' side chain, which would clash with the ligand's difluoro-substituted phenyl ring if it was present in the same conformation as in the apo structure (Figure 5S).

While one copy of J2 is bound in interface pocket 1 with full occupancy (see above), a second copy of this fragment binds to a different site on the surface of the IpgC^{10–151} “back-to-back” dimer with an occupancy refined to 85%. Due to imine-enamine tautomerism, the isoindoline nitrogen of this J2 molecule might be endowed with a hydrogen atom, allowing the formation of an H bond to the carboxylate group of Asp91'. This carboxylate is kept in place by H bonds formed to the side chain amide of Gln87' and the side chain ammonium group of Lys101 (Figure 3B). A further interaction between the Asp91' carboxylate group and the imine nitrogen of J2 is mediated by two water molecules. In addition, the fragment forms a π - π stacking interaction (parallel displaced) with the phenyl ring of Tyr68 (distance: 5.4 Å). The resulting orientation of the Tyr68 phenyl ring enables it to form an edge-to-face interaction with the phenyl ring of Phe67 (distance: 4.9 Å). In many of our IpgC^{10–151} structures, the side chains of Phe67 and Tyr68 are disordered. Here, in contrast, both side chains are well defined in the electron density, obviously due to a stabilizing effect of the bound fragment.

In the crystal structure of IpgC^{10–151} in complex with fragment J33, two copies of the ligand are visible too. The first copy, which forms H bonds to amino acid residues from four different protein chains, was refined to an occupancy of 75%.

One of its fluorine atoms appears to form water-mediated interactions with the main chain carbonyl oxygen atom of His133 from chain A and with the imidazole moiety of His133' from chain B (Figure 3C). A second fluorine atom may form a water-mediated interaction to the main chain carbonyl oxygen atom of Asp37*' of chain B from a symmetry-related “head-to-head” dimer (marked by *). The side chain carboxylate group of the neighboring Asp38*' forms a salt bridge to the amidine function of the ligand, which is likely to be protonated and, therefore, positively charged. Remarkably, the soaking of this ligand into IpgC^{10–151} crystals enables the N-terminal residues, 9 to 23, of chain A, which are invisible in the apo structure, to adopt a defined conformation. Also here, most residues fold into α helix 1. However, facilitated by the flexible loop connecting α helix 1 to α helix 2, this helix is utterly differently oriented to the TPR domain compared with chain B (Figure S6). This allows the main chain carbonyl oxygen of Ser10* within chain A of the symmetry-related “head-to-head” dimer to form two (likely charge-assisted) H bonds to the amidine function of the bound ligand. A further weak H bond (3.5 Å) is formed between the ligand's amidine group and the hydroxyl group of Ser12* (Figure 3C).

The second copy of J33, which was refined to an occupancy of 100%, interacts with symmetry equivalents, as well. Through π - π stacking, it is sandwiched between Tyr146 of chain A and Tyr146** of chain A from a symmetry mate (indicated by a **, since it is not identical to the symmetry-related chain A involved in the formation of the first J33 binding site). Indeed, this is the only contact that this copy of J33 forms with protein molecules. Since it is located on a 2-fold rotational axis, there are two symmetry equivalents of the fragment differing in the position of their fluorine atoms. No electron density is attributable to the nitrogen atoms of the ligand's amidine function (Figure 3D).

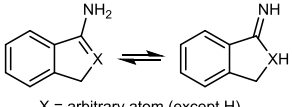
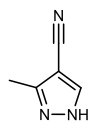
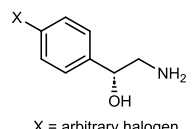
Similar to the second copy of J33, a second copy of fragment J36 is sandwiched between Tyr146 of chain A and Tyr146** of the symmetry-related chain A (as described above, the first copy of this fragment binds into the interface pocket 1). It binds with an occupancy of 100% with its methylamine group being invisible in the electron density map. Due to the 2-fold rotational axis running through it, it is present in two different orientations with 50% occupancy each (Figure 3E).

To further investigate our crystallographic hits and to identify potential additional binders that had not shown up in crystal structures, we studied the influence of the 96 compounds of the Frag Xtal Screen on the stability of our target protein. To this end, we measured the melting temperature of IpgC^{10–151} by means of the thermal shift assay.²² However, none of the tested fragments caused an increase in the melting temperature of IpgC^{10–151}, which, under the applied test conditions, amounted to 50.7 \pm 0.2 °C in the absence of any ligand. The presence of some ligands appeared to have a minor destabilizing effect on our target protein, among them some of the crystallographic fragment hits that cause interface pocket 1 to adopt the open conformation. However, the associated negative thermal shift was not considered significant.

Through Template-Based Docking to Larger Ligands.

Based on our crystallographic study, we considered J33 unsuitable for further ligand design since both copies of this fragment bind to sites that clearly originate from mere crystal contacts and, therefore, are unlikely to exist in solution. One copy of J36 is accommodated by one of these sites as well,

Table 2. Overview of Fragments for Which the Frag4Lead Workflow Was Applied

| Fragment | Modified fragment used for search | Quantity of potential follow-up compounds in the <i>MolPort</i> catalog | Quantity of molecules selected for purchase |
|----------|--|---|---|
| J2 |  <p>X = arbitrary atom (except H)</p> | 4500 | 5 |
| J20 |  | 2000 | 7 |
| J45 |  <p>X = arbitrary halogen</p> | 8500 | 5 |

while the second copy of this compound, which binds to the interface pocket 1, is insufficiently defined in the electron density map. Therefore, J36 was also not included in further efforts of ligand optimization. In contrast, we regarded the fragments which are well defined in the electron density map and which bind into interface binding pockets 1 or 2 as promising starting points for ligand optimization. In our crystal structure, both pockets are embedded in the putative “back-to-back” homodimer, which may also exist in solution. Compounds that occupy these pockets may be enlarged such that resulting molecules stabilize the “back-to-back” homodimer. This mode of action is expected to interfere with the physiological function of IpgC (see the Discussion section) and, consequently, may aid in fighting Shigellosis.

To quickly advance fragment hits to compounds with higher affinity without the need for elaborate custom synthesis, the Frag4Lead workflow has recently been developed²³ and was applied here. For fragments that bind into the interface pocket 1 or 2, we searched suitable superstructures in the *MolPort* catalog. To enhance the potential search space, the originally detected fragments were slightly modified or reduced in their chemical structure. For some fragments, however, as in the case of J11, no promising follow-up compounds could be found in the *MolPort* catalog. Based on the retrieved follow-up compounds, J2, J20, and J45 emerged as particularly promising candidates. The isoindol(in)e nitrogen of J2 was set to an arbitrary atom (except hydrogen) to enhance the search space. Only follow-up compounds that possessed an extension of J2 at the five-membered ring were considered. J20 was reduced such that only the 3-methyl-1H-pyrazole-4-carbonitrile moiety remained as a template for docking. This way, it was possible to include follow-up molecules that had a different substituent instead of the pyrrolidine ring. The fluorine atom of J45 was set to an arbitrary halogen atom (see Table 2).

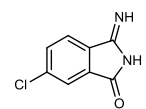
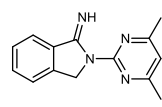
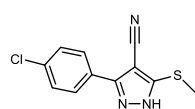
Missing side chains of the binding pocket were supplemented for docking. Either the most likely conformation was chosen or the one that caused the least steric hindrance with putatively bound molecules in the binding pocket. Glu125' located at the entrance of interface pocket 1 was most affected by this. Here, various side chain conformations were tested.

Solvent molecules were not included in docking, with a single exception. Many fragments that bind into the interface pocket 1 interact with the hydroxyl group of Tyr104 via a water molecule. This water molecule was considered for docking, as it is regarded to be essential for binding. In the case of J45, which induces a second conformation of turn $\alpha 5\alpha 6$ within chain A, both conformations were individually considered for separate docking runs.

After docking with *FlexX*, scoring with *HYDE* function, and posefiltering (detailed methodology in ref 23), final poses were evaluated in the *SeeSAR*²⁴ GUI leading to a selection of 17 compounds for purchase (overview see Tables 2 and S2).

The binding of compounds was then validated via X-ray crystallographic analysis of compound-soaked IpgC^{10–151} crystals. Ultimately, three molecules, U08, U09, and U11 were confirmed as binders via crystallography (Table 3). While both U08 and U09 are based on fragment J2, U11 is derived from the 3-methyl-1H-pyrazole-4-carbonitrile component of fragment J20 (see Table 2).

Table 3. Overview of Crystallographically Confirmed Follow-Up Compounds

| Compound name | Chemical structure | PDB ID | Resolution [Å] |
|---------------|---|--------|----------------|
| U08 |  | 8QH6 | 1.80 |
| U09 |  | 7P42 | 1.50 |
| U11 |  | 7OWV | 1.59 |

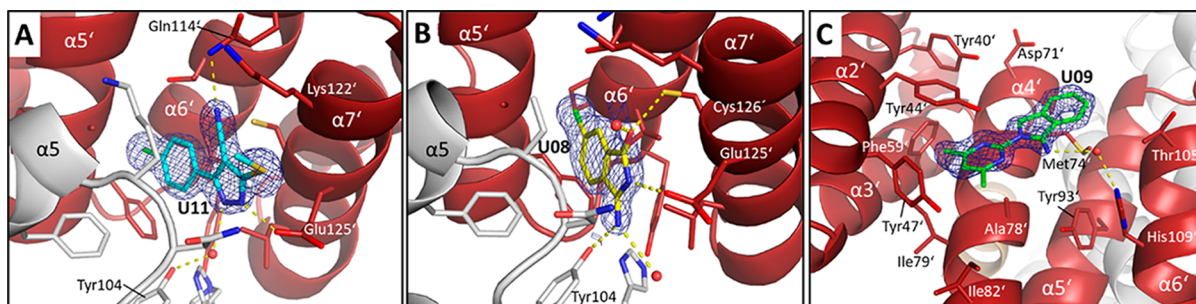


Figure 4. Relevant details from the crystal structures of IpgC^{10–151} in complex with U11, U08, and U09. The side chains of amino acids lining the binding pockets are depicted as sticks. H bonds are indicated by dotted lines. (A) Binding mode of U11, which is displayed in stick representation (carbon atoms, cyan; chlorine atom, green), within the interface pocket 1. (B) Binding mode of U08 (shown as sticks; carbon atoms are yellow, and chlorine atom green). In (A) and (B), the $2mF_o - DF_c$ density maps assignable to the ligand are contoured at a level of 1.5σ , each. Only amino acids mentioned in the main text are labeled. (C) Binding mode of U09 (shown as sticks; carbon atoms green). The $2mF_o - DF_c$ density map assignable to the ligand is contoured at a level of 1.0σ .

In the crystal structure of IpgC^{10–151} in complex with U11, the ligand is bound in interface pocket 1 with 81% occupancy. As predicted by docking, its 1*H*-pyrazole-4-carbonitrile moiety is virtually identically positioned in the pocket as that of the parent fragment J20 (Figure S7A,B). Through its nitrile function, U11 forms an H bond to the side chain amide of Gln114'. In addition, by means of the nitrogen atom at position 2 of its pyrazole moiety, it forms a water-mediated interaction with the hydroxyl group of Tyr104 (Figure 4A). Equivalent interactions between the target protein and the ligand are observed in the crystal structure of IpgC^{10–151} in complex with J20 (Figure 2C). In contrast to J20, U11 is unable to interact with the main chain carbonyl oxygen atom of Lys122', since its methyl sulfanyl group, which replaces the amino group of J20, is unable to act as a H donor. Instead, its methyl sulfanyl group makes van der Waals interactions with the aliphatic part of the Lys122' side chain, which is well defined in the electron density map of the IpgC^{10–151}·U11 crystal structure but largely invisible in the crystal structure of the IpgC^{10–151}·J20 complex. Furthermore, although the side chain carboxylate of Glu125' is disordered in the crystal structure of the IpgC^{10–151}·J20 complex, it forms an H bond to the pyrazole NH group of U11. In the IpgC^{10–151}·J20 complex, the corresponding pyrazole NH interacts with a water molecule.

Compared to the binding pose of U11 predicted by in silico-docking, the *p*-chlorophenyl moiety of the ligand is rotated by ca. 45° in the crystal structure of the IpgC^{10–151}·U11 complex. Remarkably, its position and its orientation, as observed in the crystal structure, are highly similar to those of the *p*-fluorophenyl moiety of J45 in the crystal structure of the IpgC^{10–151}·J45 complex (Figure S7C). Moreover, the chlorine-substituted phenyl moiety of fragment J66 is also bound at a very similar position (Figure 2H) proving that the hydrophobic part of the interface pocket 1 is efficiently addressed by a halogen-substituted phenyl ring.

This observation is further corroborated by the crystal structure of IpgC^{10–151} in complex with U08, which is bound to interface pocket 1 with 100% occupancy (Figure 4B). The chlorine atom attached to the phenyl ring of its isoindoline scaffold also occupies a position close to that of the chloro substituent of U11 (Figure S7D). As a result, U08 is sterically precluded from penetrating the pocket as deeply as the underlying fragment J2 (Figure S7E). An H bond that is formed between the NH group of the U08 heterocycle with

the carboxylate group of Glu125' clearly shows that the 6-chloro-3-iminoisoindolin-1-one and not the 3-amino-6-chloro-1*H*-isoindol-1-one tautomer is bound to the chaperone. A further H bond is formed by the sulfhydryl group of Cys126' and the carbonyl oxygen of the ligand. Ultimately, the 3-imino group of U08 forms a strong H bond (2.7 Å) with the hydroxyl group of Tyr104. The ligand's carbonyl oxygen as well as its 3-imino group form further H bonds to surrounding water molecules. The "subunit"-bridging H bonds generated between U08 and the two protein chains are not observed in the binding pose predicted by docking. Here, the ligand adopts a significantly different orientation, with its 3-imino group being directed to the side chain of Cys126' and its carbonyl oxygen being directed to the side chain of Tyr104 (Figure S7F).

The occupancy of U09 in the crystal structure of the IpgC^{10–151}·U09 complex was refined to 77%. Unexpectedly, in contrast to the underlying J2, this ligand is not bound by interface pocket 1. Instead, it is placed in the groove formed by the concave side of the TPR domain from chain B (Figure S8A) at a position destined to accommodate the molecular anchor region of a bound translocator (Figure S8B,C). The dimethyl pyrimidine moiety of U09 occupies a hydrophobic pocket which is lined by the side chains of Ala78', Ile79', Ile82', and a cluster of aromatic amino acids composed of Tyr40', Tyr44', Tyr47', and Phe59' (Figure 4B). Both ring systems of U09 are present in a coplanar arrangement, which is stabilized by an intramolecular H bond between the ligand's imine and one of its pyrimidine nitrogen atoms. Moreover, the phenyl ring of the U09 imino-isoindoline moiety makes van der Waals contacts to the side chain of Met74'. Except for a water-mediated interaction of its imine substituent with the imidazole NH of His109', the ligand does not form any salt bridges or H bonds with the target protein.

DISCUSSION

In addition to the structure of IpgC^{10–151} reported in the present study, another four crystal structures of IpgC, or rather slightly truncated variants thereof, have been published to now. In all of these structures, IpgC is apparently present as a homodimer, whose architecture, however, differs fundamentally depending on the exact length of the respective IpgC variant and/or the composition of the mother liquor underlying crystallization.

Lunelli et al.⁵ and Lokareddy et al.⁶ published three crystal structures of IpgC^{1–151}, in which the chaperone is present in its

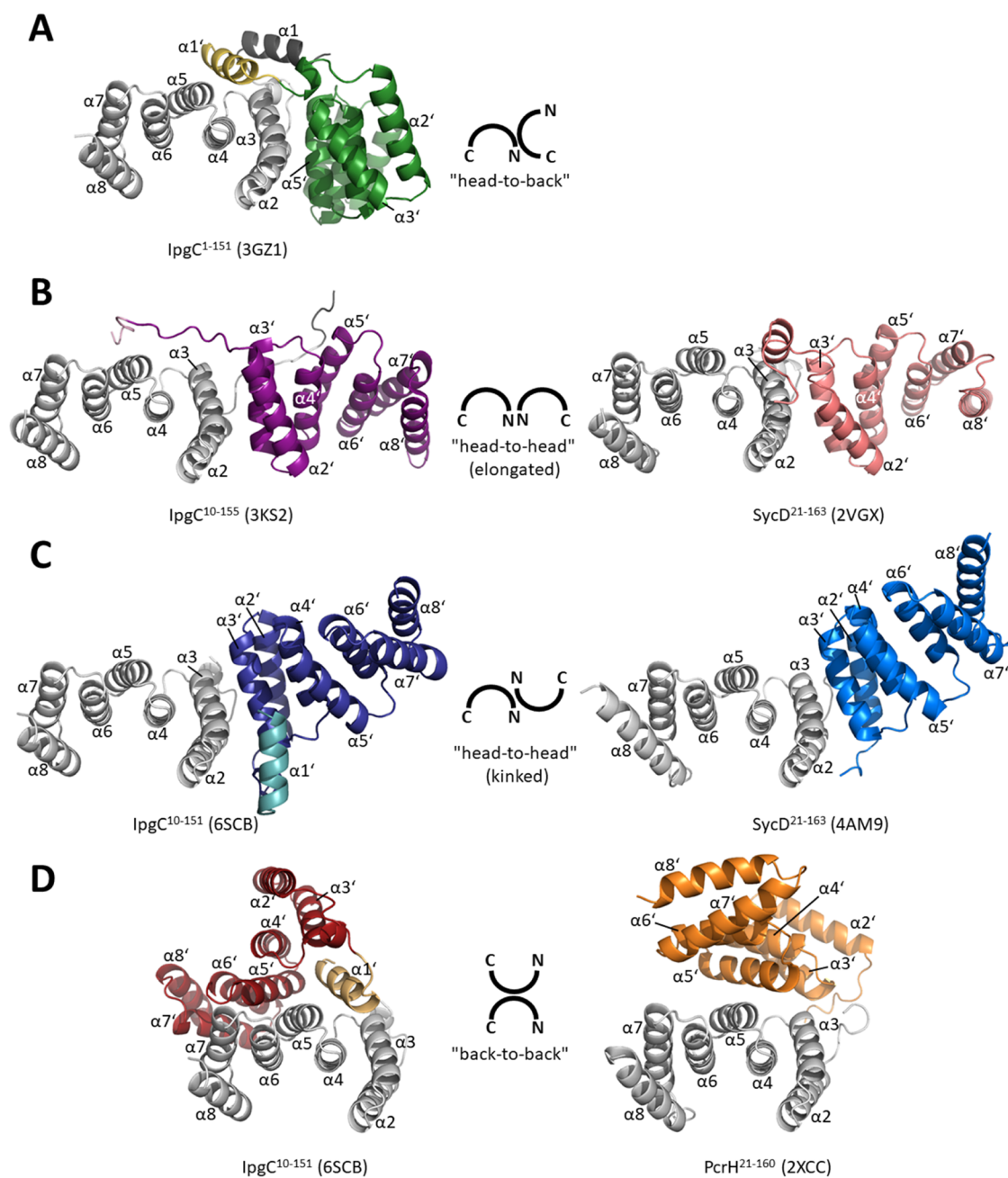


Figure 5. Different types of dimer architectures are present in the crystal structures of IpgC and orthologues thereof. Proteins are shown in ribbon representation. PDB IDs of structural models are given in parentheses. The first subunit of the dimer is colored in gray and present in the same orientation, each. The second subunit (indicated by a ') is colored, with α helix 1 (if present) being highlighted by a different color. (A) The "head-to-back" dimer of IpgC¹⁻¹⁵¹; α helix 1 of the first subunit is shown in dark gray. (B) The "head-to-head" dimer arrangement as observed in the crystal structures of IpgC¹⁰⁻¹⁵⁵ and SycD²¹⁻¹⁶³. Büttner et al. referred to this assembly as "elongated head-to-head" dimer.²⁵ (C) A different "head-to-head" arrangement, termed "kinked head-to-head" dimer by Büttner et al., as present in the crystal structures of IpgC¹⁰⁻¹⁵¹ and SycD²¹⁻¹⁶³.²⁵ (D) Two distinctly different types of "back-to-back" dimer arrangements are observed in the crystal structures of IpgC¹⁰⁻¹⁵¹ and PcrH²¹⁻¹⁶⁰.⁷

apo form (PDB ID: 3GYZ) or in complex with a peptide containing the specific anchor sequence of IpaB (PDB IDs: 3GZ1 and 3GZ2). A virtually conserved, asymmetric dimer architecture is observed in all three crystal structures, although cocrystallization with the IpaB anchor peptide leads to a slight modulation of subunit orientation in 3GZ2, in comparison with 3GYZ and 3GZ1. Disregarding α helix 1 from subunit B, this dimer is largely established by extensive contacts of the N-terminal TPR motif (α helices 2 and 3) of subunit A with the convex side (in particular α helix 5) of subunit B. Hence, it

may be referred to as a "head-to-back" assembly (Figure 5A), which radically differs from both potential dimer forms observed in our crystal structure. In all three structures determined by Lunelli et al.⁵ and Lokareddy et al.,⁶ α helix 1 is clearly visible in both subunits. However, enabled by the highly flexible loop connecting α helix 1 and α helix 2, this helix is utterly differently oriented in subunit A compared to subunit B. In the latter one, it is directed away from the TPR domain and embedded in a groove formed by α helices 3, 4, and 5 of subunit A. Remarkably, although having a different orientation

with respect to the TPR domain, this helix is identically adjusted into this groove of the dimer mate as in the “back-to-back” dimer of our crystal structure.

Barta et al. published the crystal structure of IpgC^{10–155}, in which the chaperone is also apparently present as a homodimer (PDB ID: 3KS2).¹⁶ Here, both putative subunits lack α helix 1 since there is no electron density assignable to the eight N-terminal residues, 10 to 17. Both in subunit A and subunit B, the visible residues preceding α helix 2 form an extended loop, which contacts α helix 5 on the convex side of the respective dimer mate (Figure 5B). Since the TPR domains of both subunits interact with each other via their N-terminal TPR motifs (α helices 2 and 3), this symmetric dimer features a “head-to-head” arrangement. It differs, however, substantially from the asymmetric “head-to-head” dimer present in our crystal structure (Figure 5C). Interestingly, two different “head-to-head” dimer forms are also observed in the available crystal structures of the IpgC orthologue from *Yersinia enterocolitica*, SycD. In the crystal structure of SycD^{21–163} (PDB ID: 2VGX),²⁵ this chaperone forms a “head-to-head” dimer, which is similar to the IpgC^{10–155} dimer published by Barta et al.¹⁶ (Figure 5B). In contrast, in the crystal structures of SycD^{21–163}(Ser94Glu/Tyr95Glu) (PDB ID: 2VGY)²⁵ and SycD^{21–163} in complex with an anchor sequence containing peptide (PDB ID: 4AM9),⁸ the dimer features a “head-to-head” assembly, which strongly resembles that of IpgC^{10–151} observed in our crystal structure (Figure 5C).

A further crystal structure of an IpgC orthologue, namely, PcrH^{21–160} from *Pseudomonas aeruginosa* (PDB ID: 2XCC), reveals an apparent dimer with a “back-to-back” arrangement of both subunits since they contact each other via the convex sides of their TPR domains.⁷ However, subunits are oriented to each other in a way that utterly differs from the “back-to-back” dimer of IpgC^{10–151} (Figure 5D). It should be mentioned that the PcrH homodimer was reported to be considerably less stable than those formed by IpgC and SycD.⁷

The multitude of putative homodimer architectures observed in the crystal structures of IpgC and its orthologues raises the question of which of these arrangements represents the “correct” one that exists in solution. Lunelli et al.⁵ addressed this issue by introducing two mutations, Ala94Glu and Val95Gln, in the interface of the “head-to-back” dimer present in their crystal structures. Through multiangle laser light scattering, they observed that the resulting IpgC variant was no longer able to form a homodimer, which they considered as evidence that the “head-to-back” arrangement does actually exist in solution. Yet, both Ala94 and Val95 are located in the interfaces of the “head-to-head” dimer observed in the structure of Barta et al.¹⁶ and of the “back-to-back” dimer of our crystal structure too (Figure S2B). Clearly, the Ala94Glu/Val95Gln double mutation will most likely also lead to the disruption of these dimer forms, which is why it is by no means clear which homodimer is really present in solution. Although small-angle X-ray scattering data published by Ferrari et al.²⁶ suggest that a homodimer similar to that present in the crystal structure of Barta et al.¹⁶ is predominant in solution, it is well conceivable that different subunit arrangements are at equilibrium under physiological conditions. This presumption is supported by Barta et al.,¹⁶ who reported on size exclusion chromatography experiments suggesting that the homodimer formation of IpgC may be compromised but is not completely suppressed by the Ala94Glu/Val95Gln double mutation originally introduced by Lunelli et al.⁵

The physiological significance of the IpgC homodimer is yet unknown, in particular, since it is inevitably disrupted upon the binding of a translocator. The crystal structure of the *A. hydrophila* IpgC orthologue AcrH in complex with the soluble domain of the IpaB orthologue AopB (PDB ID: 3WXX) implies that translocator binding is sterically not compatible with any of the homodimer architectures observed for IpgC.¹⁰ Clearly, the chaperone forms a heterodimer with the translocator. Moreover, in the crystal structure of the *Chlamydia pneumoniae* IpgC orthologue Scc3 in complex with the gatekeeper CopN (corresponding to MxiC of *Shigella*) (PDB ID 4NRH) the chaperone forms a heterodimer with its interaction partner, too.²⁷ Crystallographic observations are corroborated by cross-linking experiments, providing strong evidence that IpgC alone is present as a homodimer while IpgC and IpaB form a heterodimeric complex.²⁸

It seems likely that the IpgC homodimer provides for the stability of the chaperone as soon as IpaB and IpaC are secreted and, thus, are removed from the bacterial cytoplasm. Accordingly, compounds that efficiently hamper homodimer formation might destabilize IpgC and, thus, affect the virulence of *Shigella*. However, a more productive strategy may be to expand molecules that bind into a pocket arising from homodimer formation in such a way that the resulting compounds strongly increase the stability of the IpgC homodimer. The identification of the J20 derivative U11 and of the J2 derivative U08 via docking represents a very first step in this direction. In particular, ligand U08, which bridges both subunits of the potential back-to-back dimer via three H bonds, seems to be an appropriate candidate for this purpose. Since the binding of translocators and probably of further IpgC interaction partners clearly requires the disruption of the putative homodimer, this approach may represent an option to severely compromise the functionality of the chaperone. An example which proves that the stabilization of a protein/protein interaction is a promising strategy in drug development was provided by Lin et al.²⁹ They presented a crystal structure in which the N-terminal domain of the MERS-CoV nucleocapsid protein is present as a homodimer, although this domain shows virtually no tendency to oligomerize in solution. Through virtual screening, the authors identified 5-benzyloxygramine as a molecule able to stabilize the non-native dimer observed in the crystal structure. This leads to an abnormal oligomerization of the nucleocapsid protein in the cell and, ultimately, to the significant antiviral activity of 5-benzyloxygramine.

As a matter of fact, compound U09 obviously constitutes the most promising outcome of the present study. Even though found by serendipity, this ligand represents our prime indication to use IpgC as a target for rational drug design. In contrast to the underlying fragment J2, which is placed in the interface pocket 1, U09 is bound to the concave surface of the chaperone. Importantly, this binding site overlaps with the binding site of the translocators' anchor sequences. As mentioned, the dimethyl pyrimidine moiety of U09 occupies a hydrophobic pocket, which is mainly lined by the side chains of Ala78 and a cluster of aromatic amino acids comprising Tyr40, Tyr44, Tyr47, and Phe59 (Figure 4B). Remarkably, both Ala78 and the aromatic cluster are strictly conserved among IpgC-like chaperones, with only Tyr44 being replaced by phenylalanine in some orthologues. Our hypothesis is strongly supported by the fact that in the crystal structures of IpgC in complex with the IpaB anchor region (PDB IDs:

3GZ1, 3GZ2),^{5,6} this pocket accommodates the side chain of the IpaB residue Leu67, which is the conserved leucine in the Pro/Val-Xxx-Leu-Xxx-Xxx-Pro consensus sequence. Furthermore, identical interactions are observed in the crystal structures of *A. hydrophila* AcrH (\equiv IpgC) in complex with the soluble domain of AopB (\equiv IpaB),¹⁰ of *Pseudomonas aeruginosa* PcrH (\equiv IpgC) in complex with the anchor region of PopB (\equiv IpaB),⁹ and of PcrH in complex with the anchor region of PopD (\equiv IpaC).⁷ Since U09 blocks this essential pocket, it appears to be a promising starting point for the rational design of compounds inhibiting the formation of IpgC/translocator complexes, which are essential for the development of *Shigella* pathogenicity.

Several other approaches, mainly based on computational methods, have been described to discover unknown sites also in interface pockets that might be competent to accommodate fragments. First of all, molecular dynamics (MD) simulations can trace the dynamic properties of macromolecules, in particular. A pioneering approach was developed in the early 1990s in the group of Karplus.^{30,31} In MCSS (Multiple Copies Simultaneous Search), several thousand random probe molecules, such as acetone, water, methanol, or benzene, are exposed to the target protein in an MD simulation. Each probe molecule is driven to an optimal position, according to the underlying force field. The probe molecules experience interaction with the protein, but they do not “feel” each other. At the end of the calculation, a frequency distribution for the probe molecules is obtained, suggesting promising binding sites for fragments. Later, many other programs were developed with a similar concept. MD can also be used to scan a protein surface for transient binding pockets that open during the simulation. MD frames showing such transient pockets can be used in a targeted docking approach to search for putative fragment binders. Other methods allow binding pockets to be retrieved from structural databases that store the macromolecules in a precalculated fashion. Ruben Abagyan's group at UCSD in San Diego has developed *Pocketome*, a comprehensive encyclopedia that can be used to search for binding pockets, especially considering related proteins in families.³² From an experimental point of view, NMR spectroscopy, in the case of full resonance assignment, provides an alternative approach to detect fragments that bind to proteins. Chemical shift perturbations make it possible to trace the region on a protein where a promising small molecule is accommodated.³³

CONCLUSIONS

In the present study, we have determined a new crystal structure of the *Shigella* virulence factor IpgC. This protein belongs to the class II of type III secretion system chaperones and is essential for the pathogenicity of this organism. The structure reveals two potential homodimer arrangements, which have not been observed for IpgC before. The high resolution of the structure provides the basis for the use of this protein as a target in structure-based drug design. Crystallographic fragment screening yielded 10 small molecules, which chiefly bind to sites established by subunit contacts of the potential homodimers. In addition, we identified a follow-up compound based on one of the fragment hits, which occupies a strictly conserved site required for the binding of the IpgC substrates, IpaB and IpaC. Since the interaction of both translocators with IpgC is a prerequisite for *Shigella* virulence,

this substance constitutes a promising starting point for the design of anti-*Shigellosis* compounds.

EXPERIMENTAL SECTION

Gene Expression and Mutagenesis, Protein Production and Purification. The reading frame encoding IpgC^{10–151} was amplified from the *Shigella flexneri* virulence plasmid pWR501 by the standard PCR method. Through the used primers, a *Bam*HI restriction site was placed 5' and a *Not*I restriction site 3' to the truncated *ipgC* gene, allowing for its insertion into the expression vector pHW1. This vector constitutes a modified version of the commercially available pGEX-4T-1, in which the target gene is fused at its 5' end to a sequence encoding an N-terminal GST tag. In pHW1, the multiple cloning site of pGEX-4T-1 has been modified and the sequence encoding the thrombin cleavage site has been removed. A sequence encoding an 8 \times histidine tag, 3' of the pHW1 multiple cloning site, was not used for the overexpression of the truncated *ipgC* gene. The sequence ATGTATCCGCGCGGC encoding the thrombin cleavage site Tyr-Pro-Arg-Gly from the gastrin-releasing peptide,³⁴ preceded by a methionine codon, was added immediately 5' of the target gene (Figure S1).

The resulting plasmid, termed pHW1-*ipgC*-9,-4S, was transformed into *Escherichia coli* BL21-CodonPlus(DE3)-RIPL cells for overexpression of the target gene. Subsequently, cells were incubated in 150 mL of LB medium containing 100 $\mu\text{g mL}^{-1}$ ampicillin and 50 $\mu\text{g mL}^{-1}$ chloramphenicol while shaking at 37 °C overnight. 50 mL of this preculture was added to 4 L LB medium containing 100 $\mu\text{g mL}^{-1}$ ampicillin and 50 $\mu\text{g mL}^{-1}$ chloramphenicol and incubated at 37 °C while shaking at 120 rpm until an OD₆₀₀ of 0.8 was reached. Then, the culture was cooled down to 16 °C and 1 mmol L⁻¹ IPTG was added to induce gene expression followed by incubation at 16 °C while shaking for 18 h. Cells were harvested by centrifugation at 4 °C and 10,000 rpm using the Avanti J-25 centrifuge and the JA-10 rotor (Beckman Coulter) and frozen at -80 °C. The frozen cell pellet was resuspended in 100 mL of PBS (140 mmol L⁻¹ NaCl, 10 mmol L⁻¹ Na₂HPO₄, 1.7 mmol L⁻¹ KH₂PO₄, 2.8 mmol L⁻¹ KCl, pH 7.3). After cell disruption with ultrasound using the Sonifier 250 (Branson) (3 \times 180 s; 3/10 at 70% output) at 4 °C, the cell extract was centrifuged using the JA-25.50 rotor at 20,000 rpm and 4 °C for 1 h. The supernatant was loaded onto an XK 16/20 column filled with 15 mL of Glutathione Sepharose 4 Fast Flow (Cytiva) equilibrated with PBS. After a washing step with 400 mL of PBS, 500 U of thrombin (CSL Behring) dissolved in 15 mL of PBS was loaded onto the column by means of a pump at a flow rate of 1 mL min⁻¹. The target protein was separated from the GST tag via thrombin cleavage overnight at 4 °C on the column. Subsequent chromatography steps were done at room temperature using the Äktaprime plus FPLC system (GE Healthcare). IpgC^{10–151} was eluted from the column with PBS at a flow rate of 1 mL min⁻¹. Then, the eluate was concentrated with a Vivaspinn 20, 5,000 MWCO PES Ultrafiltration Unit (Sartorius) to a volume of 10 mL and loaded onto an XK 16/40 column filled with 30 mL of Sephadex G-25 (Cytiva) for buffer exchange by 20 mmol L⁻¹ BisTris, pH 6.0. The IpgC^{10–151} containing fractions were concentrated to a volume of 10 mL and loaded onto a HiTrap DEAE Fast Flow column (Cytiva). Anion exchange chromatography was performed at a flow rate of 2 mL min⁻¹ using a linear gradient comprising 120 mL of 0 to 100% elution

buffer (20 mmol L⁻¹ BisTris, pH 6.0, 1000 mmol L⁻¹ NaCl). Subsequently, buffer was exchanged by 10 mmol L⁻¹ HEPES, pH 7.5, 150 mmol L⁻¹ NaCl via dialysis at 4 °C overnight using a ZelluTrans dialysis membrane, MWCO 6–8 kDa, 45 mm (Carl Roth). The dialyzed sample (ca. 10 mL) was loaded onto a HiLoad 26/60 Superdex 75 Prep grade column (Cytiva) and size exclusion chromatography was performed at a flow rate of 2 mL min⁻¹. Ultimately, the combined IpgC^{10–151} containing fractions were concentrated with a Vivaspın 20, 5000 MWCO PES Ultrafiltration Unit until a mass concentration of ca. 10 mg mL⁻¹ was reached. The protein solution was immediately used for crystallization or frozen in liquid nitrogen and stored at –80 °C until further use. After Coomassie staining, SDS PAGE of the final sample revealed a single band corresponding to IpgC^{10–151}. The yield achieved by this protocol typically amounted to ca. 35 mg per 4 L of bacterial culture.

Crystallization of IpgC^{10–151} and Ligand Soaking.

Crystals were obtained with the sitting drop vapor diffusion method in the presence of 500 μL of reservoir solution, which contained 25–30% (m/V) PEG 4000 (Carl Roth), 100 mmol L⁻¹ Tris-HCl, pH 8.0, and 300 mmol L⁻¹ MgCl₂, at 25 °C. 1.5 μL of the above-described protein preparation was mixed with 1.0 μL of reservoir solution. Crystallization temperature turned out critical since at a temperature lower than 25 °C most crystals were twinned and the reproducibility of crystallization was unsatisfying. To circumvent nucleation and to accelerate the appearance of crystals, streak seeding was applied. To this end, a preformed crystal of the apo protein was touched with the tip of a horse hair, which was then dipped into a new crystallization droplet. Using the latter method, crystals reached their full size (0.4 mm in length, 0.1 mm in diameter) after two to 4 weeks of growth.

To obtain crystals of IpgC^{10–151} in complex with ligands, crystals of the apo protein were soaked in a mixture of 0.5 μL of ligand solution (1000 mmol L⁻¹ ligand in DMSO) and 4.5 μL of soaking solution [30% (m/V) PEG 4000, 100 mmol L⁻¹ Tris-HCl, pH 8.0, 300 mmol L⁻¹ MgCl₂]. Dependent on the stability of crystals in the respective fragment solution, soaking time was varied between a few minutes up to 24 h. In the case where crystals immediately cracked during soaking, a drop of the ligand solution was pipetted next to a drop of soaking solution containing the crystal. The slow increase of ligand concentration turned out gentler, with crystals being able to stay in a solution generated this way for 3 to 24 h. Prior to data collection, both crystals of apo IpgC^{10–151} and crystals soaked with a ligand were immediately vitrified in liquid nitrogen with no cryo buffer being required.

Data Collection, Processing, and Refinement. Crystals were measured at 100 K at the BESSY II on MX-beamline 14.1 or 14.2, respectively.³⁵ Data processing and scaling were done using XDSAPP2.³⁶ Molecular replacement was performed using CCP4 Phaser.³⁷ As a starting model for the apo structure, PDB ID 3GYZ⁵ was used (residues 32–151 of chain A). For all other structures, the apo structure of IpgC^{10–151} (PDB ID: 6SCB) was chosen as a search model. The program Phenix^{38,39} was used for the refinement of untwinned data sets. A 5% subset of all reflections (randomly selected) was used for the calculation of R_{free}. Five cycles of Cartesian simulated annealing were performed, followed by five refinement cycles. This was followed by alternating cycles of model building in COOT⁴⁰ and refinement in Phenix, especially refining XYZ coordinates, isotropic B-factors, and the occupation of alternative

conformations of amino acid side chains and fragments (if the occupancy was at least 20%). In addition, TLS groups were applied. For all complex structures, the same TLS groups were used. TLS groups were determined via TLSMD web server.^{41,42} In addition, non-crystallographic symmetry (NCS) was used for refinement. CCP4 Refmac⁴³ was used to refine the twinned data sets of the following PDB entries: 7PE0 and 7PEF. Twin refinement was based on intensities. For the conversion of PDB files into mmCIF files the program PDB Extract⁴⁴ was applied. During the molecule construction in COOT, single atoms of amino acid side chains were deleted if they could not be clearly identified in the 2mF_o-DF_c electron density at 1.0 σ. Water molecules were incorporated if visible in the mF_o-DF_c difference electron density at 3.0 σ. Restraints for the fragments were received from Grade Webserver.⁴⁵ Hydrogen atoms were added if this resulted in at least 0.5% lower R_{free}. The programs Procheck⁴⁶ (calculation of Ramachandran plots) and Moleman⁴⁷ (calculation of temperature B-factors) were used for structure validation. The Ramachandran plots of most structures reveal a not allowed Φ/Ψ angle combination for Lys101 from chain A and/or chain B. In these cases, Lys101 occupies the central position of a classical γ turn (Table S1).

Thermal Shift Assay. Protein unfolding was investigated via the Thermal Shift Assay.²² Fluorescence data were collected using the fluorescence of SYPRO orange (Thermo Fisher Scientific) and a QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific). The wavelength for excitation was 470 ± 15 nm, and that for emission was 575 ± 15 nm. In the temperature range between 10 and 90 °C, 657 data points were collected. MicroAmp Fast 96-Well Reaction Plates (0.1 mL) (Thermo Fisher Scientific) were used. 18.5 μL of buffer (50 mmol L⁻¹ HEPES, 150 mmol L⁻¹ NaCl, pH 7.5) with 20 × SYPRO orange (Thermo Fisher Scientific) was combined with 1 μL of protein solution (10 mg mL⁻¹ in 10 mmol L⁻¹ HEPES, 150 mmol L⁻¹ NaCl, pH 7.5) and 0.5 μL of fragment dissolved in DMSO (final concentration: 2.5 mmol L⁻¹). To record the zero value, 0.5 μL of DMSO was used instead of the fragment solution. Triple measurements were performed. Melting temperatures were determined as the maximum of the first derivative of the melting curves with data evaluation being performed with the Protein Thermal Shift software version 1.4 (Thermo Fisher Scientific).

Virtual Screening of Follow-Ups Using Frag4Lead.

Virtual Screening was performed according to the Frag4Lead workflow,²³ built within the Konstanz Information Miner (KNIME) version 3.4.4⁴⁸ with some modifications. The receptor files of starting fragments J2, J20, and J45 were prepared using the SeeSAR GUI (BioSolveIT). Potential and commercially available follow-ups were retrieved using the MolPort Chemical Search KNIME node (SIA MolPort, Latvia), by performing substructure searches in the entire MolPort catalog. After that, molecules were filtered for PAINS⁴⁹ and toxic motifs. For the resulting molecules, 3D coordinates were generated using NAOMI (BioSolveIT).⁵⁰ In the next step, template-based Docking using FlexX⁵¹ via BioSolveIT KNIME nodes was performed. Then, several post-docking filters were applied: quasi-identical poses, poses with low docking score, and poses with molecules not in the vicinity of the experimentally found fragment were removed. Residual poses were clustered by RMSD and the three best-scoring representatives were retained. Finally, molecules were scored using HYDE⁵² via BioSolveIT KNIME nodes. Sessions were

then reviewed, and an informed selection of molecules was finalized.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.3c07058>.

Construction of *ipgC*^{10–151} expression plasmid (Figure S1); supporting structures (Figures S2–S8); crystallographic data (S1), and follow-up compounds (Table S2) (PDF)

Accession Codes

Coordinates and structure factors have been deposited under following accession codes: apo IpgC^{10–151}: 6SCB; IpgC^{10–151} in complex with J1: 7NL8;... J2: 7AYW;... J6: 7O6S;... J11: 7B1U;... J20: 7AXY;... J33: 7AZV;... J36: 7O04;... J45: 7NRG;... J52: 7PE0;... J66: 7NHW;... DMSO: 7PEF;... U08: 8QH6;... U09: 7P42;... U11: 7OWV.

■ AUTHOR INFORMATION

Corresponding Author

Klaus Reuter – Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, D-35037 Marburg, Germany; orcid.org/0000-0003-3673-7971; Phone: +49-6421-28-25845; Email: reuterk@staff.uni-marburg.de; Fax: +49-6421-28-28994

Authors

Marina Gárdonyi – Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, D-35037 Marburg, Germany; Present Address: Fraunhofer-Institut für Silicatforschung, Neunerplatz 2, D-97082 Würzburg, Germany

Christian Hasewinkel – Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, D-35037 Marburg, Germany

Johanna Wallbaum – Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, D-35037 Marburg, Germany

Jan Wollenhaupt – Macromolecular Crystallography, Helmholtz-Zentrum Berlin, D-12489 Berlin, Germany; Present Address: PROTEROS Biostructures GmbH, Bunsenstr. 7a, D-82152 Planegg-Martinsried, Germany

Manfred S. Weiss – Macromolecular Crystallography, Helmholtz-Zentrum Berlin, D-12489 Berlin, Germany; orcid.org/0000-0002-2362-7047

Gerhard Klebe – Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, D-35037 Marburg, Germany; orcid.org/0000-0002-4913-390X

Andreas Heine – Institut für Pharmazeutische Chemie, Philipps-Universität Marburg, D-35037 Marburg, Germany

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsomega.3c07058>

Funding

This work was supported by the LOEWE center DRUID (project A4). The authors acknowledge the allocation of synchrotron radiation beam time and travel support to the BESSY II synchrotron provided by the Helmholtz-Zentrum Berlin, Germany. M.G., K.R., G.K., and A.H. thank BioSolveIT for free SeeSAR version 10.1 software license. J.W. and M.S.W. are grateful for support from iNEXT-Discovery, Project Number 871037, funded by the Horizon 2020 Program of the European Commission. Open-access funding was provided by the Open Access Publishing Fund of Philipps-Universität

Marburg with support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

Notes

The authors declare no competing financial interest.

Figure Preparation Structural figures were prepared using PyMOL (<http://www.pymol.org>).

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the beamline staff at BESSY II (Helmholtz-Zentrum Berlin, Germany) for outstanding support during data collection. They cordially thank T. Adam (Charité, Berlin; meanwhile Labor Berlin – Mikrobiologie und Hygiene) for providing them with virulence plasmid pWR501 DNA. Finally, they thank C. Sohn for his help during in-house X-ray data collection and S. Dörr for excellent technical assistance.

■ REFERENCES

- (1) Parsot, C. *Shigella* type III secretion effectors: how, where, when, for what purposes? *Curr. Opin. Microbiol.* **2009**, *12*, 110–116.
- (2) Murillo, I.; Martinez-Argudo, I.; Blocker, A. J. Genetic dissection of the signaling cascade that controls activation of the *Shigella* type III secretion system from the needle tip. *Sci. Rep.* **2016**, *6*, No. 27649.
- (3) Blocker, A.; Gounon, P.; Larquet, E.; Niebuhr, K.; Cabiaux, V.; Parsot, C.; Sansonetti, P. The tripartite type III secretion of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *J. Cell Biol.* **1999**, *147*, 683–693.
- (4) Ménard, R.; Sansonetti, P.; Parsot, C.; Vasselon, T. Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of *S. flexneri*. *Cell* **1994**, *79*, 515–525.
- (5) Lunelli, M.; Lokareddy, R. K.; Zychlinsky, A.; Kolbe, M. IpaB-IpgC interaction defines binding motif for type III secretion translocator. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 9661–9666.
- (6) Lokareddy, R. K.; Lunelli, M.; Eilers, B.; Wolter, V.; Kolbe, M. Combination of two separate binding domains defines stoichiometry between type III secretion system chaperone IpgC and translocator protein IpaB. *J. Biol. Chem.* **2010**, *285*, 39965–39975.
- (7) Job, V.; Mattei, P.-J.; Lemaire, D.; Attree, L.; Dessen, A. Structural basis of chaperone recognition of type III secretion system minor translocator proteins. *J. Biol. Chem.* **2010**, *285*, 23224–23232.
- (8) Schreiner, M.; Niemann, H. H. Crystal structure of the *Yersinia enterocolitica* type III secretion chaperone SycD in complex with a peptide of the minor translocator YopD. *BMC Struct. Biol.* **2012**, *12*, No. 13.
- (9) Discola, K. F.; Förster, A.; Boulay, F.; Simorre, J.-P.; Attree, L.; Dessen, A.; Job, V. Membrane and chaperone recognition by the major translocator protein PopB of the type III secretion system of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **2014**, *289*, 3591–3601.
- (10) Nguyen, V. S.; Jobichen, C.; Tan, K. W.; Tan, Y. W.; Chan, S. L.; Ramesh, K.; Yuan, Y.; Hong, Y.; Seetharaman, J.; Leung, K. Y.; Sivaraman, J.; Mok, Y. K. Structure of AcrH-AopB chaperone-translocator complex reveals a role for membrane hairpins in type III secretion system translocon assembly. *Structure* **2015**, *23*, 2022–2031.
- (11) Cherradi, Y.; Hachani, A.; Allaoui, A. Spa13 of *Shigella flexneri* has a dual role: chaperone escort and export gate-activator switch of the type III secretion system. *Microbiology* **2014**, *160*, 130–141.
- (12) Cherradi, Y.; Schiavolin, L.; Moussa, S.; Meghraoui, A.; Meksem, A.; Biskri, L.; Azarkan, M.; Allaoui, A.; Botteaux, A. Interplay between predicted inner-rod and gatekeeper in controlling substrate specificity of the type III secretion system. *Mol. Microbiol.* **2013**, *87*, 1183–1199.
- (13) Botteaux, A.; Sory, M. P.; Biskri, L.; Parsot, C.; Allaoui, A. MxiC is secreted by and controls the substrate specificity of the *Shigella flexneri* type III secretion apparatus. *Mol. Microbiol.* **2009**, *71*, 449–460.
- (14) Mavris, M.; Page, A.-L.; Tourné, R.; Demers, B.; Sansonetti, P.; Parsot, C. Regulation of transcription by the activity of the *Shigella*

- flexneri* type III secretion apparatus. *Mol. Microbiol.* **2002**, *43*, 1543–1553.
- (15) Parsot, C.; Ageron, E.; Penno, C.; Mavris, M.; Jamoussi, K.; d'Hauteville, H.; Sansonetti, P.; Demers, B. A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in *Shigella flexneri*. *Mol. Microbiol.* **2005**, *56*, 1627–1635.
- (16) Barta, M. L.; Zhang, L.; Picking, W. L.; Geisbrecht, B. V. Evidence for alternative quaternary structure in a bacterial type III secretion system chaperone. *BMC Struct. Biol.* **2010**, *10*, No. 21.
- (17) Krissinel, E.; Henrick, K. Interference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774–797.
- (18) Sheriff, S.; Hendrickson, W. A.; Smith, J. L. Structure of myohemerythrin in the azidomet state at 1.7/1.3 Å resolution. *J. Mol. Biol.* **1987**, *197*, 273–296.
- (19) Sheriff, S. Some methods for examining the interactions between two molecules. *Immunomethods* **1993**, *3*, 191–196.
- (20) Frag Xtal Screen. <https://www.jenabioscience.com/crystallography-cryo-em/screening/fragment-screen/x-fs-101-frag-xtal-screen>.
- (21) Hassaan, E.; Eriksson, P.-O.; Geschwindner, S.; Heine, A.; Klebe, G. Fragments as novel starting points for tRNA-guanine transglycosylase inhibitors found by alternative screening strategies. *ChemMedChem* **2020**, *15*, 324–337.
- (22) Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2007**, *2*, 2212–2221.
- (23) Metz, A.; Wollenhaupt, J.; Glöckner, S.; Messini, N.; Huber, S.; Barthel, T.; Merabet, A.; Gerber, H. D.; Heine, A.; Klebe, G.; Weiss, M. S. Frag4Lead: growing crystallographic fragment hits by catalog using fragment-guided template docking. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2021**, *77*, 1168–1182.
- (24) SeeSAR, Version 10.1; BioSolveIT GmbH: Sankt Augustin, Germany, 2019.
- (25) Büttner, C. R.; Sorg, I.; Cornelis, G. R.; Heinz, D. W.; Niemann, H. H. Structure of the *Yersinia enterocolitica* type III secretion translocator chaperone SycD. *J. Mol. Biol.* **2008**, *375*, 997–1012.
- (26) Ferrari, M. L.; Charova, S. N.; Sansonetti, P. J.; Mylonas, E.; Gazi, A. D. Structural insights of *Shigella* translocator IpaB and its chaperone IpgC in solution. *Front. Cell. Infect. Microbiol.* **2021**, *11*, No. 673122.
- (27) Archuleta, T. L.; Spiller, B. W. A gatekeeper chaperone complex directs translocator secretion during type three secretion. *PLoS Pathog.* **2014**, *10*, No. e1004498.
- (28) Adam, P. R.; Patil, M. K.; Dickenson, N. E.; Choudhari, S.; Barta, M.; Geisbrecht, B. V.; Picking, W. L.; Picking, W. D. Binding affects the tertiary and quaternary structures of the *Shigella* translocator protein IpaB and its chaperone IpgC. *Biochemistry* **2012**, *51*, 4062–4071.
- (29) Lin, S.-M.; Lin, S.-C.; Hsu, J.-N.; Chang, C.-K.; Chien, C.-M.; Wang, Y.-S.; Wu, H.-Y.; Jeng, U.-S.; Kehn-Hall, K.; Hou, M.-H. Structure-based stabilization of non-native protein-protein interactions of Coronavirus nucleocapsid proteins in antiviral drug design. *J. Med. Chem.* **2020**, *63*, 3131–3141.
- (30) Cafilisch, A.; Miranker, A.; Karplus, M. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* **1993**, *36*, 2142–2167.
- (31) Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 32–58.
- (32) Kufareva, I.; Ilatovskiy, A. V.; Abagyan, R. Pocketome: An Encyclopedia of small-molecule binding sites in 4D. *Nucleic Acids Res.* **2012**, *40*, D535–D540.
- (33) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discovery* **2016**, *15*, 605–619.
- (34) Chang, J. Y. Thrombin specificity. Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate. *Eur. J. Biochem.* **1985**, *151*, 217–224.
- (35) Mueller, U.; Förster, R.; Hellmig, M.; Huschmann, F. U.; Kastner, A.; Malecki, P.; Pühringer, S.; Röwer, M.; Sparta, K.; Steffien, M.; Ühlein, M.; Wilk, P.; Weiss, M. S. The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *Eur. Phys. J. Plus* **2015**, *130*, 141–150.
- (36) Sparta, K. M.; Krug, M.; Heinemann, U.; Mueller, U.; Weiss, M. S. XDSAPP2.0. *J. Appl. Crystallogr.* **2016**, *49*, 1085–1092.
- (37) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
- (38) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213–221.
- (39) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2019**, *75*, 861–877.
- (40) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486–501.
- (41) Painter, J.; Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 439–450.
- (42) Painter, J.; Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **2006**, *39*, 109–111.
- (43) Winn, M. D.; Murshudov, G. N.; Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **2003**, *374*, 300–321.
- (44) Yang, H.; Guranovic, V.; Dutta, S.; Feng, Z.; Berman, H. M.; Westbrook, J. D. Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1833–1839.
- (45) Smart, O. S. Grade Web Server. <https://grade.globalphasing.org/cgi-bin/grade/server.cgi>.
- (46) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (47) Kleywegt, G. J.; Zou, J. Y.; Kjeldgaard, M.; Jones, T. A. *International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules*; Rossmann, M. G.; Arnold, E., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; Vol. 353–356, pp 366–367.
- (48) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meil, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*; Preisach, C.; Burkhardt, H.; Schmidt-Thieme, L.; Decker, R., Eds.; Springer: Berlin, Heidelberg, 2008.
- (49) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.
- (50) Urbaczek, S.; Kolodzik, A.; Fischer, J. R.; Lippert, T.; Heuser, S.; Groth, I.; Schulz-Gasch, T.; Rarey, M. NAOMI: On the almost trivial task of reading molecules from different file formats. *J. Chem. Inf. Model.* **2011**, *51*, 3199–3207.

(51) Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 41–54.

(52) Reulecke, I.; Lange, G.; Albrecht, J.; Klein, R.; Rarey, M. Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the HYDE scoring function. *ChemMedChem* **2008**, *3*, 885–897.