HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2015 March 29.

Published in final edited form as:

Nat Commun. ; 5: 4976. doi:10.1038/ncomms5976.

High-resolution structure of a *Shigella* type III secretion needle by solid-state NMR and cryo-electron microscopy

Jean-Philippe Demers[1,2,#], **Birgit Habenstein**[1,#], **Antoine Loquet**[1], **Suresh Kumar Vasa**[1], **Karin Giller**[1], **Stefan Becker**[1], **David Baker**[2], **Adam Lange**[1,2,3,*], and **Nikolaos G. Sgourakis**[4,*]

[1]Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

[2]Department of Molecular Biophysics, Leibniz-Institut für Molekulare Pharmakologie (FMP), 13125 Berlin, Germany

[3]Institut für Biologie, Humboldt-Universität zu Berlin, 10115 Berlin, Germany

[4]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA

Abstract

We introduce a general hybrid approach for determining the structures of supramolecular assemblies. Cryo-electron microscopy (cryo-EM) data define the overall envelope of the assembly and rigid-body orientation of the subunits while solid-state NMR (ssNMR) chemical shifts and distance constraints define the local secondary structure, protein fold and inter-subunit interactions. Finally, Rosetta structure calculations provide a general framework to integrate the different sources of structural information. Combining a 7.7-Å cryo-EM density map and 996 ssNMR distance constraints, the structure of the Type-III Secretion System (T3SS) needle of *Shigella flexneri* is determined to a precision of 0.4 Å. The calculated structures are cross-validated using an independent dataset of 691 ssNMR constraints and STEM measurements. The hybrid model resolves the conformation of the non-conserved N-terminus, that occupies a

[*]Corresponding authors: alange@fmp-berlin.de (A.Lange) and sgourn@u.washington.edu (N.G.S.).
[#]Equally contributing authors.
Current address: Institute of Chemistry & Biology of Membranes & Nanoobjects (UMR5248 CBMN), Institut Européen de Chimie et Biologie (IECB), CNRS - Universite Bordeaux - Institut Polytechnique Bordeaux, All. Geoffroy Saint-Hilaire 33600 Pessac, France (A.L.) and Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA (N.G.S.)

Author Contributions

K.G. and S.B. expressed, purified and polymerized T3SS needles. J.P.D. recorded ssNMR experiments and prepared the processed data. J.P.D., B.H. and A. Loquet analyzed ssNMR data. J.P.D. and S.K.V. analyzed STEM images. A. Lange supervised ssNMR experiments and ssNMR and STEM data analysis. N.G.S and D.B. performed Rosetta structure calculations. J.P.D. and N.G.S. performed the cross-validation of structural models. N.G.S. and J.P.D. programmed the supplementary software. J.P.D., B.H., A. Lange and N.G.S. wrote the paper. All authors discussed the results and commented on the manuscript.

Competing financial interests

The authors declare no competing financial interests. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Accession codes

The final ensemble consisting of the top-10 ranking models according to Rosetta energy was submitted in the PDB (PDB ID 2MME).

protrusion in the cryo-EM density, and reveals conserved pore residues forming a continuous pattern of electrostatic interactions, thereby suggesting a mechanism for effector protein translocation.

INTRODUCTION

Large molecular assemblies are omnipresent in living cells and responsible for a broad spectrum of biological functions such as cellular motion (molecular motors), cellular structure (cytoskeletal filaments) and molecular transport (bacterial secretion systems). Traditionally, X-ray crystallography, solution-state Nuclear Magnetic Resonance (NMR) and cryo-electron microscopy (EM) are used to probe the atomic structures of biomolecular systems. Molecular assemblies however pose technical challenges for conventional methods[1]: finding an appropriate crystallization condition is a strenuous effort for large, multi-subunit systems; crystallization can be prevented by the presence of flexible or disordered regions or domains; and systems with non-crystallographic symmetry lack the long range order required to produce a discrete diffraction pattern. In addition, high-MW protein complexes (in excess of 100 kDa) generally do not exhibit sufficiently fast molecular tumbling to be studied by solution-state NMR spectroscopy, with notable exceptions for favorable systems[2]. The study of isolated structural domains at atomic resolution is insufficient for a complete description of the larger, biologically active system, as proteins can adopt different conformations in isolation compared to their functional complexes.

Two methods are emerging to tackle the structures of intact large assemblies towards gaining a mechanistic insight into their biological function. Technically, both solid-state NMR[3–6] (ssNMR) and cryo-EM[7,8] are not limited by the molecular size of the assembly under study. Density data from cryo-EM defines an overall envelope of the supramolecular assembly in a resolution range of 20–8 Å for standard applications and up to 6 Å for favorable systems[1]. Solid-state NMR data provide crucial information on the local structure such as the secondary structure propensity, backbone dihedral angles and interatomic distances of up to ~10 Å, which can be detected both within the individual protein domains and across inter-molecular subunit interfaces. Cryo-EM and ssNMR data are in principle highly complementary and their use in combination holds great promise for future integrative structural studies[9]. However, a general framework to integrate the different sources of structural information into high-resolution models of multimeric assemblies has been limited by 1) the different levels of resolution provided by each technique that complicates their use in computational methods, the fine-tuning of relative constraint weights and the assessment of self-consistency between these fundamentally different datasets and 2) the availability of computational methods that would determine the range of models consistent with the experimental data, as a function of the varying degrees of freedom.

We introduce a generalized hybrid approach for high-resolution structure determination of supramolecular assemblies that combines high-resolution cryo-EM density maps with solid-state NMR distance constraints. We demonstrate the power of new approach by determining the structure of the Type-III Secretion System (T3SS) needle of *Shigella flexneri* at atomic

resolution (to a precision of 0.4 Å backbone RMSD). The T3SS, or injectisome, is a supramolecular assembly found in Gram-negative bacteria, such as *Shigella*, *Salmonella*, *Escherichia*, *Pseudomonas* or *Yersinia*[10], which serves to deliver toxic effector proteins into their target eukaryotic host cell during infection[11,12]. The current study extends and improves on our previous work focusing on T3SS needles, which include the atomic model of *Salmonella typhimurium*[13–15] using ssNMR data and helical parameters from scanning transmission electron microscopy (STEM), the ssNMR resonance assignment of *Shigella flexneri* MxiH needles[16] which allowed to identify a common architecture for T3SS needles, the study of dynamics and conformational heterogeneity of MxiH needles using DNP-enhanced ssNMR[17], and solid-state NMR assignment strategies using sparsely ^{13}C-labeled samples[18–20] or highly deuterated proteins[21]. The new hybrid approach combines a 7.7 Å cryo-EM density map reconstructed using 100,000 needle segment images[22] with extensive structural information obtained from ssNMR: 162 backbone dihedral angles and 996 carbon-carbon distance constraints from proton-driven spin diffusion (PDSD[23,24]) experiments. The calculated structures are validated using 691 independent carbon-carbon and proton-proton distance constraints from ChhC and NhhC experiments[25,26]. With 12 and 8 distance constraints per residue in the calculation and validation sets respectively, the final models are among the most well resolved solid-state NMR structures to date[27,28]. Moreover, the atomic structures reveal the conformation of the previously undefined N-terminal segment (residues 1–11) and provide new insight into the protein translocation mechanism.

RESULTS

Identification of the fold and inter-molecular interfaces

We have previously achieved the NMR chemical shift assignment of MxiH needles[16] which allowed us to identify the secondary structure of MxiH subunit proteins in their assembled state. In that work, we employed solid-state NMR experiments having short mixing times (e.g. PDSD ≤100ms) to obtain intra-residue and sequential amino acid connectivity information. In the current work, we recorded a set of diverse and complementary experiments employing long PDSD mixing times (400 ms – 850 ms) that provided a large number of cross-peaks corresponding to medium and long-range distance constraints (Supplementary Table 1). Experiments were recorded on [1-^{13}C]-glucose (1-Glc) and [2-^{13}C]-glucose (2-Glc) labeled protein samples. Additional ChhC and NhhC experiments[25,26] were recorded on a uniform ^{13}C-labeled sample. In order to favor the detection of long-range cross-peaks, we recorded individual spectra with high signal-to-noise ratio.

From the 1-Glc and 2-Glc sets of spectra, we can rapidly identify pairs of amino acids, called anchor points, which are certain to be in close distance proximity using two telltale spectral features: 1) frequency-unambiguous cross-peaks have only one possible NMR assignment within a ± 0.15 ppm tolerance window; 2) network-unambiguous cross-peaks are part of an extensive network with numerous distance constraints between two amino acids. These anchor points, combined with the previously established secondary structure[16], serve to define the main fold of the MxiH subunits in the context of the needle assembly *de novo*, without using any previous structural model. In detail, amino acid pairs related by at

least two frequency-unambiguous cross-peaks or by four or more cross-peaks in a given labeling scheme are used as anchors to produce a preliminary map of the *Shigella* needle architecture (Fig. 1a). A cluster of contacts between the two antiparallel α-helices (e.g. F19-A63/Q64, T23-Y60, L26/Q27-Y57) indicates that MxiH subunits adopt a typical helix-loop-helix fold. Residues N43, P44 and L47 in the loop region make contacts with the end of the C-terminal α-helix (e.g. P44-I78, L47-I78/79) as well as with the rigid N-terminal segment (W10-N43/P44), indicating that subunit (*i*) and subunit (*i* – 11) of the helical assembly are arranged in a head-to-tail manner, forming an inter-molecular axial interface. An extensive lateral interface comprises a dense network of contacts formed between the C-terminal regions L59-K69 of subunit (*i*) and K72-R83 of subunits (*i* + 5) and (*i* + 6). Lateral contacts are also found between the loop region and both the N-terminal helix (F19-L37/A38) and the C-terminal helix (P41-L59/Y60). The rigid N-terminal extension (S2 to T11) is curved, as indicated by multiple contacts between tryptophan W10 and N-terminal residues V3-V5. Further contacts (P6-Q27/G28) indicate a lateral inter-molecular association between the N-terminal extension and the central region of the N-terminal helix that forms a minor kink at residues G22-Q24. This map confirms that the protein fold and inter-molecular interfaces identified in *Salmonella typhimurium* T3SS needles[13] are also present in *Shigella flexneri* needles.

Assignment of a large number of distance constraints

Using the preliminary map, we proceed to the disambiguation of all ssNMR distance constraint cross-peaks collected in the 1-Glc and 2-Glc spectra. The use of sparse ^{13}C labeling schemes enables very high spectral resolution, with ^{13}C line-widths ranging from 0.09 to 0.25 ppm, and the high spectral homogeneity of the T3SS needle samples leads to very small assignment uncertainty, with the standard deviation of intra-residue and sequential peak positions over all spectra of 0.04 ppm on average for ^{13}C and 0.03 ppm for ^{15}N. In order to further improve the accuracy of the chemical shift assignment for long mixing-time experiments, we re-calculated the full table of average resonance frequencies on an individual basis for each dimension of each spectrum, when enough intra-residue and sequential cross-peaks were available. A Sparky extension was created to display per spectrum resonance frequencies and a robust estimate of the average frequency (Supplementary Software 1). Our method thus allows the use of small tolerance windows (± 0.15 ppm for ^{13}C chemical shifts), which results in a drastic reduction of the number of assignment possibilities for distance cross-peaks.

Assignment ambiguity exists at two levels, chemical shift ambiguity and subunit ambiguity. Chemical shift ambiguity relates to the identification of carbon nuclei that give rise to the cross-peak. Subunit ambiguity exists as each cross-peak can arise due to either of 1 intra-subunit and 6 different inter-subunit atom pair combinations (Fig. 1b): intra-subunit, (*i*) to (*i*); inter-subunit, lateral: (*i*) to (*i* ± 5), or (*i*) to (*i* ± 6); and inter-subunit, axial: (*i*) to (*i* ± 11). The presence of multiple, distinct interfaces could potentially complicate the subunit-subunit assignment of ssNMR constraints. However, it is possible to exploit the particular architecture of helical assemblies, as the symmetry of the architecture dictates that subunits (*i* ± 5) and (*i* ± 6) are located within one axial translation (L) relative to subunit (*i*), and subunits (*i* ± 11) within two axial translations (Fig. 1b). This approach eliminates the need to

produce an additional mixed isotopically labeled sample for the purpose of interface identification[29,30]. Since, in the more general case, helical assemblies consist of layers of subunits with axial translations of 0, 1 and 2: for example, for an alternative 13-start helical assembly, subunits (*i*), (*i* ± 6/7) and (*i* ± 13) have axial translations of 0, ±1 and ±2, respectively. Our approach can thus be applied to any helical arrangement of subunits regardless of the number of subunits per layer.

As demonstrated in Fig. 1b, we can utilize the short axial translation between two amino acid pairs in the preliminary map in order to classify a distance constraint as intra-molecular, axial or lateral in the absence of a preliminary structural model of the system. Chemical shift ambiguities that do not fit any of the three inter-subunit categories are also excluded, allowing further disambiguation of cross-peaks. One type of subunit ambiguity remains however for lateral constraints, as subunit (*i* + 5) and (*i* + 6) have similar axial translation within the assembly, and is resolved in the iterative approach presented below. The distance constraints collected from the 1-Glc and 2-Glc spectra were classified as unambiguous if the chemical shift ambiguities could be adequately resolved (total of 1190 correlations), or as ambiguous cross-peaks otherwise (Fig. 2). Ambiguous correlations and correlations from the uniformly $^{13}C/^{15}N$-labeled dataset (ChhC and NhhC) were reserved be used for cross-validation of the final models (as outlined in the structure validation section).

Due to the high spectral resolution and signal-to-noise ratio of the ssNMR data, a large number of distance constraints can be found for individual nuclei. More than 12,350 cross-peaks were analyzed in long-range spectra and over 17,850 cross-peaks were analyzed in total, including short mixing time spectra. Excerpts from PDSD spectra are shown in Fig. 3a for the resonances of atoms T23Cγ_2, Q27Cβ and W10Cδ_1 for which 16, 20 and 22 distance correlations are found. Those correlations are highlighted in the final calculated structure in Fig. 3b along with additional correlations from atoms P6Cδ (12 correlations), P44Cδ (23) and Y60Cα (18). An analysis of the final list of constraints reveals that the distance correlations obtained from the two sparse ^{13}C labeling schemes are highly complementary, with <5% of constraints being shared between the 1-Glc and 2-Glc datasets (Fig. 4b). Although few cross-peaks are shared between the same atom pairs in the two datasets, the amino acid pairs giving rise to long-range correlations are highly similar (Fig. 4d), providing an independent confirmation that the resonance assignments and preliminary fold are accurate (Fig. 4a and 4c).

Iterative Structure Calculations using Rosetta

We performed a series of Rosetta structure calculations to integrate the ssNMR and cryo-EM data towards determining the final MxiH needle structure. We used an iterative assignment approach that makes use of prior knowledge alongside the structure determination process (Fig. 2). Given the previous analysis of “anchor” constraints (Fig. 1a) showing that the preliminary fold and inter-subunit arrangement are similar to the PrgI needle structure[13] for residues L12 to R83, we initialized the assignments of all ambiguous constraints from the 2-Glc dataset into the 7 interface possibilities (Fig. 1b) using the shortest distance in a homology-based model of MxiH[16]. In the resulting first-round calculations performed using the preliminary ssNMR constraints and cryo-EM density, the

lowest-energy models showed convergence to the correct needle fold and subunit arrangement; however, structural convergence in the preliminary ensemble was insufficient to yield a high-resolution structure (below 5 Å backbone RMSD). The 10 lowest-energy models were then used to refine the chemical shift assignments in the NMR peak lists and the interface assignments, (*i*) to (*i* + 5) or (*i* + 6), for lateral constraints. We performed two more iterations of structure calculations followed by assignment refinement until structural convergence to below 2.5 Å was reached in the low-energy ensemble. In this process, 194 constraints were discarded from the final constraint dataset as they were found to be inconsistent with any interface assignment; either multiple interface assignments were possible or the cross-peaks were observed over longer than expected distances (> 12 Å). For instance, the increased rigidity of the C-terminal helices[16] can enhance the rate of proton-driven spin diffusion. A final list of 996 medium- and long-range constraints was obtained from both 1-Glc and 2-Glc datasets, with the following distribution: 580 intra-subunit, 124 inter-subunit, axial and 292 inter-subunit, lateral.

Using the refined constraints, we computed 5,000 models ranked according to the weighted sum of Rosetta energy, EM correlation and ssNMR constraint score terms. The 10 top-ranking models of the final assignment round were converged to below 2.5 Å in backbone RMSD and showed a minimal number of constraint violations (1.6–3.5% using an upper limit of 12 Å), while also showing good correlation to the EM density (0.62–0.67, see Supplementary Fig. 1) and structural statistics according to Rosetta's energy function. Three models were further selected based on different criteria: 1) Rosetta energy 2) Number of constraint violations and 3) Correlation to the cryo-EM density. Each of the 3 top-ranking models was further refined 10 times in full-atom mode by optimizing the weights of both the EM density and NMR constraint score terms as described in the methods section (see also Supplementary Fig. 2). This step was done to evaluate how changes in sidechain packing and local backbone dihedrals (within 1.5 Å backbone RMSD) affect the energies of different starting models. The final, refined models show optimal geometry, good fits to the EM density and minimal constraint violations (Table 1).

Cross-validation of the structural determination approach

Towards an atomic-level validation of the final structure, we collected additional ssNMR data not used at any stage of the iterative calculations and assignment process from two sources (Table 1): ^{1}H-^{1}H correlations from uniformly-labeled samples observed in ChhC and NhhC spectra (96 atom pair correlations) and ambiguous correlations from all three isotopic labeling schemes (595 correlations). The vast majority of these constraints were satisfied in the final ensemble (validation set: 660/691, 4.5 ± 0.1 % violations, mean ± SD over the top-10 ranking models, PDB ID 2MME), therefore supporting the correctness of the proposed structure at the atomic level (Fig. 5a). Using this unbiased cross-validation approach, we tested whether the previously published structure of MxiH determined using the same cryo-EM data alone[22] (PDB ID 3J0R) was compatible with the experimental ssNMR distance constraints (Fig. 5b). Notably, the number of distance violations more than doubles (validation set: 615/682, 9.8 % violations). A pairwise comparison of all distances in the validation dataset indicates that the models determined using both the ssNMR and cryo-EM data (this study) have significantly shorter distances compared to the previous

cryo-EM structure, on average 1.22 Å shorter per distance constraint (paired difference Student t-test, 95 % confidence level), demonstrating the high-accuracy of the hybrid structural determination approach. The median distance for the validation set and pairwise comparison of distances were also used to monitor and guide the modeling procedure between successive rounds of hybrid modeling, similar to the concept of R_{free} employed in crystallography31.

Verification of sample compatibility and tilt angle

In the present application of our hybrid approach, we combine structural constraints obtained from two experimental techniques, where the employed T3SS needle samples also differ in their preparation: for the cryo-EM density map, over-expressed MxiH serotype 2 needles were extracted by shearing from the bacterial surface; for ssNMR structural constraints, serotype 6 MxiH needles were polymerized *in vitro*. In the previous sections we have demonstrated that the hybrid models produced by our approach satisfy the experimental data from both sample preparations: all features of the cryo-EM density map (correlation of 0.62 to 0.67), including the protrusion region of electron density between subunits (*i*) and (*i* – 5) (see Supplementary Fig. 1), and the independent ssNMR distance constraints. In order to confirm the compatibility of the samples employed to produce the structural data, we recommend as part of the hybrid approach the measurement of independent data allowing the validation of the final structure at the macroscopic level. We thus recorded scanning transmission electron microscopy (STEM) images of *in vitro* polymerized MxiH needles (Supplementary Fig. 3) to independently identify their helical arrangement.

The intensity of scattered electrons in a dark field image is directly related to the mass of the object. We determined the mass-per-length of the polymerized needle assemblies present in the ssNMR samples by integration of needle segments from calibrated STEM images[32,33]. Considering the molecular weight of the MxiH subunit protein (9391 Da MW), the observed mass-per-length of 2184 ± 2 Da/Å corresponds to an axial subunit displacement of 4.30 Å in the needle assembly. This value is highly consistent with the axial subunit displacement measured in the final hybrid models ensemble deposited in the PDB, 4.33 ± 0.02 Å, and confirms that *in vitro* polymerized needles adopt an 11-start helical symmetry (Supplementary Fig. 3).

When viewed from the outside of the structure, the needle filament shows a staggered pattern (Fig. 6), where subunit (*i*) interacts laterally with subunits (*i* ± 5,6) and axially with subunits (*i* ± 11). Notably, subunits (*i*) and (*i* ± 1) are not in close proximity. In the hybrid models (Fig. 6a), the subunit (*i* + 11) does not stack exactly on top of complete of subunit (*i*) as would be predicted for a number of subunits per two turns of precisely N=11.0. Instead, the number of subunits per two turns is N=11.2 and the needle filament accumulates a ‘tilt’ angle of −6.40° (positive in the counter-clockwise direction) for each subsequent 11 subunits (Fig. 6b). Recently, small changes in the tilt angle of the flagellin proto-filament and in the twist orientation have been shown to be important for the function of the flagellum in bacterial motility[34].

Handedness of the helical assembly

The new structure determination approach presented here further determines the handedness of the needle super-helix. In the initial PrgI needle calculations[13], the handedness of the needle could not be defined on the basis of the ssNMR and cryo-EM experimental data alone as both left- and right- handed geometries gave very similar constraint scores. Subsequent calculations on PrgI needles employing an extended number of ssNMR restraints showed that the needle filament adopts a right-handed geometry[14]. The current ssNMR data for the MxiH needle contains a large set of long-range constraints that should in principle be able to distinguish between the two alternative helical arrangements (Fig. 7). To address this question the calculations are repeated assuming a left-handed needle, inverting the *i*+5 and *i*+6 initial assignments. Notably, the right-handed calculations (Fig. 7a) lead to significantly better optimization of the NMR constraints, average of 22.5 ± 1.5 violations per model relative to 37.4 ± 3 for the left-handed calculations (Fig. 7b), that can be attributed to fewer violations of lateral constraints. A significant difference is also obtained in the average full-atom energies (−143 ± 15/−93 ± 19 Rosetta Energy units in the right- and left-handed low-energy ensembles respectively), suggesting that the right-handed structures show more favorable sidechain packing and hydrogen-bonding terms. Taken together, these results show that the current hybrid approach can precisely determine the long-range features of the structure, such as its handedness, in addition to the radius, tilt angle and helical pitch parameters consistently with all the available data.

DISCUSSION

Many important biological assemblies are not amenable to conventional atomic structure determination methods. However, precise knowledge of the structure of biological assemblies is at the foundation of a mechanistic understanding of biological processes, such as bacterial infection. Here, we demonstrate that a hybrid approach (Fig. 2) can solve the structures of molecular assemblies at atomic resolution with high accuracy, by integrating two complementary sources of experimental information, solid-state NMR and cryo-EM. The new approach has several key advantages relative to the current state-of-the-art structure determination protocols based solely on solid-state NMR[27–29]:

1. The symmetry of the system is modeled explicitly using a generalized framework[35] recently extended to include helical symmetries[13]. This allows for manipulating the internal backbone and side-chain conformation alongside the rigid-body degrees of freedom for more efficient sampling of conformational space by a) only considering conformations that are consistent with a given symmetry and b) performing a minimal number of energy and derivative calculations that are propagated among the different symmetric subunits. This allows us to model the long-range order of the needle filament together with the local atomic structure, which is needed for the use of the cryo-EM density as a calculation bias in addition to the NMR data.
2. Our approach for iterative restraint assignment alongside the structure calculation process is similar to previous NMR structure determination protocols[36,37], further enhanced by the use of a physically realistic, all-atom energy function[38] and

extended to supra-molecular assemblies with complex symmetries. By using a physically realistic all-atom energy function to model hydrogen-bonding networks and sidechain packing interactions[39], our approach does not rely as heavily on the experimental constraints, which are used only as a minor calculation bias with minimal weights to prevent over-fitting. This approach also avoids the need for inferring additional hydrogen-bond constraints not directly observed experimentally, as done in previous ssNMR studies of similar systems[27,28], and is more robust to the exact parameterization of constraint upper limits, that can depend on both the system and type of experiment and labeling scheme, as shown extensively in previous work[29].

3. The combination of the complementary cryo-EM and solid-state NMR data is crucial in obtaining a converged set of models that are further consistent with the Rosetta Energy function. In particular, the cryo-EM density map was necessary for sampling the correct needle topology and subunit arrangement in the early stages of the structure calculations, to verify the initial assignments of the NMR distance restraints (Fig. 2). Finally the 7.7 Å density[22] allows for better defining the rigid-body orientation of the subunits, and calculations performed using the ssNMR restraints alone show higher Rosetta energies relative to the hybrid models. Therefore, by taking advantage of a combination of complementary datasets, the new approach is highly suitable for challenging complexes where either technique alone would be insufficient to uniquely define a high-resolution structure.

Different biophysical techniques used in a hybrid approach are likely to have different requirements for sample conditions. For example, the cryo-EM data were obtained on samples of needles sheared from *Shigella flexneri* culture over-expressing MxiH[22] while the solid-state NMR data used here were recorded on isotopically labeled samples that were prepared by heterologous bacterial expression and polymerized in vitro. This sample preparation method is necessary in order to obtain sufficient quantities of labeled MxiH needles to record high-resolution NMR spectra[18]. Biochemical, Immuno-labeling and biophysical characterization of the needle samples done previously in our group has established that the *in vitro* reconstituted needles indeed reflect the *in vivo* needles, as outlined in detail in references [13 and 16]. Finally, the fact that we obtain convergence on a unique set of models showing good fits to both the cryo-EM density (Supplementary Fig. 1) and satisfy independent ssNMR constraints with only a minimal number of violations (Table 1 and Fig. 5) further suggests the compatibility of the two datasets at the atomic level.

The high-resolution MxiH structure determined here is consistent with the needle architecture established previously[13,14,16] consisting of 5.6 subunits per turn (N=11.2) of a 23.5 Å-pitch helix. The radius of the needle is 23.5 Å, consistent with previous EM studies[40]. Moreover, the structure of the nonconserved N-terminus (M1-T11) is now fully resolved, showing a short α-helical conformation involving residues P6 to D9 (PDKD), that occupies a “protrusion” region of the EM density (Supplementary Fig. 1). The same density region was incorrectly attributed to an intra-subunit β-sheet in a previously published model using the cryo-EM data alone[22]. The conformation of the N-terminus is defined by 132 intra-subunit constraints, including a cluster of constraints connecting V3, W10 and L15

(Fig. 4a, c). In addition, there are 46 inter-subunit constraints connecting the N-terminal residues V3-W10 of subunit (*i*) to residues D20-L34 at the N-terminal α-helix of subunit (*i* – 5), and two correlations between subunits (*i*) and (*i* – 5), V5-Y57 and P6-Y60, indicate that the beginning of the PDKD motif points inside the needle assembly (Fig. 3b, c). Finally, a cluster of 3 constraints connects T11 of subunit (*i*) to E56 located in the C-terminal (inner) α-helix of subunit (*i* – 6).

The MxiH needle structure highlights key, conserved features involved in the translocation mechanism of substrate proteins (Fig. 8). A recent functional and cryo-EM analysis of MxiH needles using designed trapped substrates showed that effector proteins are translocated in an unfolded state directly through the needle pore in a directional manner[41]. Close inspection of the electrostatic potential on the structure itself reveals that, while the needle lumen presents several charged residues to interact with unfolded substrates, these are typically compensated by conserved opposite charges, resulting in an electrostatically balanced surface (Fig. 8a, b). An inspection of the sequence conservation pattern on the MxiH needle structure identifies two clusters of highly conserved residues (Fig. 8c, d). While the first cluster forms a continuous patch on the interface between interacting subunits and plays a key role in the structural integrity of the complex (Fig. 8d, top of the structure), a second cluster of residues Lys 69, Lys 72 and Asp 73 forms a circular arrangement that decorates the needle pore (Fig. 8d, inset). The sidechains of these residues participate in a continuous pattern of symmetry-related electrostatic interactions. A previous systematic Alanine scanning mutagenesis study[42] investigated sequence-function relationships in the MxiH needle. This study showed that mutation of each of these three conserved residues to Alanine resulted in functionally defective mutants that cannot invade or lyse host cells, although they still show assembly of relatively intact needle structures that are secretion competent. While K69A and K72A resulted in altered secretion profiles of different substrates, the D73A mutant showed a “constitutively on” phenotype and secreted effector proteins in a deregulated manner without sensing an inducing cell-contact signal. These results are in agreement with the luminal localization of K69 and K72 allowing direct side chain interactions with the secreted proteins and further suggest a role of D73 in regulating substrate release from the needle. The availability of the high-resolution MxiH structure produced by the new hybrid methodology now permits the rational design of experiments towards elucidating the mechanism of translocation trough the needle pore, study of interactions between the needle and other components of the T3SS or molecules of the extracellular milieu.

METHODS

Sample preparation

Three samples of isotopic labeled *Shigella flexneri* T3SS needles were prepared following the established protocol[13,16] for expression, purification and *in vitro* polymerization of T3SS needle proteins: *E. coli* strain BL21(DE3) bacteria are transformed with a modified pET16b plasmid containing the *mxiH* gene fused at the N-terminus to a His_7 tag followed by a tobacco etch virus (TEV) protease cleavage recognition sequence. The fusion protein is expressed in minimal medium supplemented with $^{15}NH_4Cl$ as nitrogen source and either

[U-$^{13}C_6$]-glucose (uniform labeling), [1-13C]-glucose (1-Glc labeling) or [2-13C]-glucose (2-Glc labeling) as carbon source. The cells are pelleted and lysed. The fusion protein is purified from the lysate by affinity chromatography followed by reverse phase chromatography. The N-terminal His tag and TEV cleavage sequence are released after digestion by TEV protease. The MxiH subunit protein is purified by reverse phase chromatography. The polymerization of MxiH needles was carried out at 37 °C during sixteen days and produced approximately 20 mg of labeled material each time. Pellets of MxiH needles were obtained by ultra-centrifugation at 57,000 × *g* (40,000 rpm at 15°C during 30 min in a Beckman TL-100.1 rotor), washed 5× by re-suspending the pellets in fresh buffer and ultracentrifugation, and then transferred into 4.0-mm MAS rotors.

Solid-state NMR

Solid-state NMR experiments were conducted on spectrometers operating at 600 MHz, 800 MHz and 850 MHz ^{1}H Larmor frequency (Avance I and Avance III, Bruker Biospin, Germany) at MAS rates in the range 10.5–12.5 kHz. The sample temperature was maintained at 5.5°C by monitoring the temperature using the ^{1}H chemical shift of water in reference to the methyl ^{1}H signal of DSS[43]. A ramped cross-polarization with contact time of 0.7–1.2 ms was used for the initial ^{1}H-^{13}C and ^{1}H-^{15}N transfers. For ^{13}C-^{13}C correlation experiments, carbon-carbon mixing was accomplished via PDSD with the mixing times ranging from 300 ms to 850 ms. Experiments with short mixing times (50 ms or 100 ms) were also recorded to identify intra-residual and sequential correlations. Proton decoupling with a nutation frequency of 83.3 kHz was employed during evolution periods and acquisition, using either SPINAL-64[44] or SWf-TPPM[45] with a $\bar{R}\bar{R}R\bar{R}$ supercycle[46] and a tangential sweep[47] (N = 11, sweep window d = 0.25, cut-off angle tco = 55°, phase angle θ = 15°). For spectra recorded on 800 MHz and 850 MHz spectrometers, carbon-nitrogen scalar *J* couplings were removed by applying c.a. 2 kHz of waltz-16[48] decoupling on ^{15}N during acquisition and ^{13}C evolution periods. On the 600 MHz ^{1}H frequency spectrometer, spectra were recorded in double-resonance mode to increase sensitivity.

Processing and peak picking of ssNMR spectra

Spectra were processed using Bruker Topspin 2.1 and the NMRpipe software[49]. A quadratic sine window function is employed in both dimensions, with sine bell shift (SSB) parameters in the range 3.4–3.8, and FIDs are zero-filled up to the second-next power of two. Polynomial baseline correction was applied in both frequency domains. Spectra were analyzed with Sparky[50] (UCSF) and CcpNmr Analysis[51,52].

We carried out peak picking manually on the basis of our previously reported sequential chemical shift assignment of MxiH needles[16] (BMRB entry 18651). We first discarded peaks corresponding to artifacts such as spinning sidebands and reflections of the auto-peak diagonal and we assigned peaks corresponding to intra-residual and sequential correlations, for which cross-peaks are generally also present in short mixing time PDSD spectra (50 ms – 100 ms). Other cross-peaks correspond to medium-range correlations (between residue *j* and *k*, $2<|j-k|\leq 4$) or long-range correlations ($|j-k|\geq 5$, either intra-molecular or inter-molecular).

The average resonance frequency for a given nucleus was obtained from intra-residue and sequential cross-peaks for all spectra or for a single spectral dimension if a sufficient number of cross-peaks are available in one spectrum (>5). To be retained for analysis, a cross-peak must have a signal-to-noise ratio higher than 4.2. The noise in each spectrum is estimated by taking the average of the central 16 out of 20 trials, each trial computing the noise as median of 128 randomly sampled absolute data heights from the spectrum. The median ^{13}C-^{13}C distance detected in PDSD spectra is 6.10 Å and the 10^{th}- and 90^{th}-percentile distances are 4.37 Å and 10.08 Å, a range of distances which is expected for long mixing periods in PDSD spectra (Fig. 5a). The constraints detected in ChhC and NhhC spectra involve ^{1}H-^{1}H internuclear distances and therefore offer a higher resolution than the ^{13}C-^{13}C distances obtained from 1-Glc and 2-Glc datasets used in the structure determination process, with a median ^{1}H-^{1}H distance of 5.39 Å, and 10^{th}- and 90^{th}-percentile distances of 4.02 Å and 8.35 Å.

Assignment of chemical shift ambiguities

For each cross-peak, all resonance assignment possibilities within a tolerance window (0.15 ppm radius of the peak center for 1-Glc and 2-Glc, 0.2 ppm for uniform labeling) are considered. If only a single possibility exists, the cross-peak is classified as frequency unambiguous. The classification of remaining cross-peaks is done as follows:

The method employed here to determine the interface assignment of long-range correlations (either intra-molecular, lateral or axial interface) exploits the different axial translation of subunits (i), ($i \pm 5/6$) and ($i \pm 11$) which have 0, ±1 and ±2 axial translation indices with respect to subunit (i), respectively (Fig. 1). In particular, for a given cross-peak, assignment possibilities highly incompatible with the preliminary map established based on the symmetric arrangement of the system (Fig. 1b) are eliminated. Assignment possibilities incompatible with the isotopic labeling pattern (1-Glc and 2-Glc labeled samples) are eliminated as well. For the latter purpose, the basic labeling pattern of glucose-labeled samples was considered[53,54] as well as the additional possibility of scrambling due to subsequent rounds in the citric acid cycle or involvement of other metabolic pathways[20]. The precise labeling patterns for 1-Glc and 2-Glc are confirmed by inference from intra-residual and sequential cross-peaks. If a single assignment possibility remains, the cross-peak is classified as network unambiguous. In the case of multiple remaining possibilities, the cross-peak is classified as network unambiguous only if the average chemical shift deviation of one assignment is significantly smaller than for other possibilities (at least 0.05 ppm less); otherwise the cross-peak is classified as ambiguous (Fig. 2).

Cross-validation procedure

The validation set is constituted of cross-peaks not employed at any point of the calculation: ambiguous cross-peaks from 1-Glc and 2-Glc spectra, ambiguous and unambiguous cross-peaks from ChhC and NhhC spectra. Only cross-peaks with a signal-to-noise ratio higher than 6 and corresponding to medium- or long-range correlations were considered. In order to prevent introducing any bias in the analysis due to the prior constraint assignment, we systematically selected the chemical shift assignment and the subunit interface giving rise to the shortest inter-nuclear distance within all assignment possibilities contained in the

chemical shift tolerance window. This automatic procedure allows quantifying the agreement of a candidate structure to the experimental ssNMR distance constraints in a completely unbiased and operator-free fashion. An in-house MATLAB program was used to compile the list of shortest distances for an arbitrary PDB file of the needle assembly, yielding the distance distribution histogram (Fig. 5) and median distance.

Rosetta structure calculations

The structure calculations were performed using the Rosetta symmetric modeling framework[35] extended to helical symmetry[13]. Starting from a helical array of 29 polypeptide chains (83-residue), the fold-and-dock protocol[55] uses symmetric, backbone fragment and sidechain Monte-Carlo trials in internal coordinate space to explore internal subunit conformations, and rigid-body moves to explore the subunit arrangement that are propagated to symmetry-related subunits according to a user-defined set of transformations. While the MxiH needle solid-state NMR chemical shifts[16] were used to select 3- and 9-residue backbone fragments[56], the NMR distance constraints were used to restrain the internal degrees of freedom (backbone dihedral angles and sidechain torsions) together with the 6-rigid body degrees of freedom that define the subunit arrangement and helical parameters of the needle filament.

We made use of an iterative, structure calculation-guided approach to assign constraints into intra-subunit, and six types of inter-subunit interactions from uniform labeled samples (Fig. 2). In each round of structure calculations, we used two criteria to assign the raw NMR data (atom pairs) into the different interfaces and exclude outliers that are consistently violated in the low-energy models:

a. A constraint was included in the next round if it was satisfied in more than 30% of the previous round 10 lowest-energy models, according to an absolute upper limit of 12 Å and

b. A constraint was included in the next round if it can be consistently assigned to the same interface in more than 70% of the previous round 10 lowest-energy models.

While the initial assignments of restraints to the axial (i to $i \pm 11$) and lateral (i to $i \pm 5$ or $i \pm 6$) interfaces were based on the secondary structure and symmetry considerations alone, as outlined in the results section, to initialize the assignments of the lateral interface type, (i to $i \pm 5$ versus $i \pm 6$) we made use of limited information from our previously published homology model of the MxiH needle[16]. The structural convergence on a self-consistent network of restraints suggests that our assignment approach leads to the identification of the correct interfaces present in the biological assembly, further confirmed by independent cross-validation (Table 1). This result can be contrasted to the poor assignment and cross-validation statistics and lack of structural convergence obtained in calculations using an alternative, incorrect model[22] (PDB ID 3J0R) of the MxiH needle to initialize the lateral assignments (Fig. 5b). In the more general case, the method can handle completely *de novo* assignments through the available Rosetta software as outlined in detail in Supplementary Methods. This circumvents the need for expensive labeling schemes using uniform, diluted and mixed samples described in references [29,30], and further allows the inclusion of prior knowledge from homologous structures in the PDB, when available.

An energy term that measures correlation of the entire 29-subunit system to the 7.7 Å cryo-EM density map was used as described previously[57] to restrain the helical parameters of the needle filament. The EM density and NMR constraint weights were optimized using a grid search in a series of calculations performed with increasing weights while monitoring the total number of violations and Rosetta full-atom energy in the resulting models (Supplementary Fig. 2). NMR constraints were implemented as a flat-bottom potential with an upper limit of and an exponential penalty function. The exact values of the 12/9 Å upper limits used in the assignments and final structure calculations respectively were calibrated for optimal performance of the method, without using prior structural information. In particular, the 12 Å upper limit was optimal in clearly demarcating wrong assignments from correctly assigned restraints in the early stages of the assignment process, based on a grid search in the range (5–15 Å). The 9 Å limit was based on the observed distribution of ^{13}C-^{13}C cross-peaks in the PDSD spectra relative to the average distances sampled in preliminary structure calculations performed using a 5 Å upper limit.

Scanning Transmission Electron Microscopy (STEM)

Measurements were carried out at the STEM facility of the Brookhaven National Laboratory (Upton, NY, USA). The BNL microscope operates at a voltage of 40 keV and produces a finely focused electron beam of approximately 0.25 nm in diameter. Three sample preparations were investigated: straight MxiH needles, 3× diluted needles and 10× diluted needles, with most images obtained from the 10× dilution. Tobacco mosaic virus (TMV) particles were added for internal calibration and as a quality control for the sample preparation (Supplementary Fig. 3).

The analysis of the images was done using the PCmass 3.2 software (BNL). The default quality-of-fit parameters specified in the software (e.g. r.m.s.d. of background pixels, and r.m.s.d. image minus best-fit model) were used to determine the validity of each individual image segment. Image segments are selected for well-separated individual needles in areas of uniform background. The TMV particle segments are fitted with the “TMV Rod” model and the MxiH needle segments are fitted with the “9nm Rod-GC” model (solid rod). For each scan size (0.512 or 1.024 μm), the size of the integration radius was calibrated for the two fitting models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Hannes K. Fasshuber and Pascal Fricke for scientific discussions. We thank Brigitta Angerstein for expert technical assistance, Barth-Jan van Rossum for the preparation of an image of the T3SS needle and Joe Wall and Beth Lin (Brookhaven National Laboratory) for performing STEM experiments. This work was supported, in whole or in part, by the Max Planck Society, the Leibniz-Institut für Molekulare Pharmakologie, the Deutsche Forschungsgemeinschaft (Emmy Noether Fellowship to A. Lange), the NSERC of Canada (Postgraduate Scholarship to J.P.D.), the Fondation Bettencourt Schueller (to A. Loquet), EMBO (post-doctoral fellowships to B. H. and A. Loquet), the National Institutes of Health (1 R01 GM092802-01 to D.B.), and the European Union Seventh Framework Program under Grant Agreement 261863 (Bio-NMR). N.G.S. acknowledges funding by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (NIH).

References

1. Karaca E, Bonvin AMJJ. Advances in integrative modeling of biomolecular complexes. Methods. 2013; 59:372–81. [PubMed: 23267861]
2. Sprangers R, Velyvis A, Kay LE. Solution NMR of supramolecular complexes: providing new insights into function. Nature Methods. 2007; 4:697–703. [PubMed: 17762877]
3. Renault M, Cukkemane A, Baldus M. Solid-State NMR Spectroscopy on Complex Biomolecules. Angewandte Chemie-International Edition. 2010; 49:8346–57.
4. Goldbourt A. Biomolecular magic-angle spinning solid-state NMR: recent methods and applications. Current Opinion in Biotechnology. 2013; 24:705–15. [PubMed: 23481376]
5. Muller, H.; Etzkorn, M.; Heise, H. Modern NMR Methodology Vol. 335 Topics in Current Chemistry. Heise, H.; Matthews, S., editors. 2013. p. 121-56.
6. Loquet A, Habenstein B, Lange A. Structural investigations of molecular machines by solid-state NMR. Accounts of Chemical Research. 2013; 46:2070–9. [PubMed: 23496894]
7. Zhou ZH. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. Current Opinion in Structural Biology. 2008; 18:218–28. [PubMed: 18403197]
8. Baker ML, Zhang J, Ludtke SJ, Chiu W. Cryo-EM of macromolecular assemblies at near-atomic resolution. Nature Protocols. 2010; 5:1697–708. [PubMed: 20885381]
9. Ward AB, Sali A, Wilson IA. Integrative Structural Biology. Science. 2013; 339:913–5. [PubMed: 23430643]
10. Troisfontaines P, Cornelis GR. Type III secretion: More systems than you think. Physiology. 2005; 20:326–39. [PubMed: 16174872]
11. Sansonetti PJ. Rupture, invasion and inflammatory destruction of the intestinal barrier by Shigella, making sense of prokaryote-eukaryote cross-talks. Fems Microbiology Reviews. 2001; 25:3–14. [PubMed: 11152938]
12. Schroeder GN, Hilbi H. Molecular pathogenesis of Shigella spp.: Controlling host cell signaling, invasion, and death by type III secretion. Clinical Microbiology Reviews. 2008; 21:134–56. [PubMed: 18202440]
13. Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. Atomic model of the type III secretion system needle. Nature. 2012; 486:276–9. [PubMed: 22699623]
14. Loquet A, Habenstein B, Chevelkov V, Vasa SK, Giller K, Becker S, Lange A. Atomic structure and handedness of the building block of a biological assembly. Journal of the American Chemical Society. 2013; 135:19135–8. [PubMed: 24328247]
15. Loquet A, Habenstein B, Demers JP, Becker S, Lange A. Structure of a bacterial nanomachine Type 3 secretion system needle. Medecine sciences: M/S. 2012; 28:926–8. [PubMed: 23171893]
16. Demers JP, Sgourakis NG, Gupta R, Loquet A, Giller K, Riedel D, Laube B, Kolbe M, Baker D, Becker S, Lange A. The Common Structural Architecture of Shigella flexneri and Salmonella typhimurium Type Three Secretion Needles. PLoS pathogens. 2013; 9:e1003245. [PubMed: 23555258]
17. Fricke P, Demers JP, Becker S, Lange A. Studies on the MxiH Protein in T3SS Needles Using DNP-Enhanced ssNMR Spectroscopy. Chemphyschem: a European journal of chemical physics and physical chemistry. 2014; 15:57–60. [PubMed: 24282046]
18. Loquet A, Lv G, Giller K, Becker S, Lange A. C-13 Spin Dilution for Simplified and Complete Solid-State NMR Resonance Assignment of Insoluble Biological Assemblies. Journal of the American Chemical Society. 2011; 133:4722–5. [PubMed: 21401039]
19. Habenstein B, Loquet A, Giller K, Becker S, Lange A. Structural characterization of supramolecular assemblies by 13C spin dilution and 3D solid-state NMR. Journal of Biomolecular NMR. 2012; 55:1–9. [PubMed: 23202987]
20. Lv G, Fasshuber HK, Loquet A, Demers JP, Vijayan V, Giller K, Becker S, Lange A. A straightforward method for stereospecific assignment of val and leu prochiral methyl groups by solid-state NMR: Scrambling in the 2-C-13 Glucose labeling scheme. Journal of Magnetic Resonance. 2013; 228:45–9. [PubMed: 23354009]

21. Chevelkov V, Giller K, Becker S, Lange A. Efficient CO-CA transfer in highly deuterated proteins by band-selective homonuclear cross-polarization. Journal of Magnetic Resonance. 2013; 230:205–11. [PubMed: 23558259]
22. Fujii T, Cheung M, Blanco A, Kato T, Blocker AJ, Namba K. Structure of a type III secretion needle at 7-A resolution provides insights into its assembly and signaling mechanisms. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:4461–6. [PubMed: 22388746]
23. Szeverenyi NM, Sullivan MJ, Maciel GE. Observation of spin exchange by two-dimensional fourier-transform ^{13}C cross-polarization magic-angle spinning. Journal of Magnetic Resonance. 1982; 47:462–75.
24. Bloembergen N. On the interaction of nuclear spins in a crystalline lattice. Physica. 1949; 15:386–426.
25. Lange A, Becker S, Seidel K, Giller K, Pongs O, Baldus M. A concept for rapid protein-structure determination by solid-state NMR spectroscopy. Angewandte Chemie-International Edition. 2005; 44:2089–92.
26. Lange A, Luca S, Baldus M. Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. Journal of the American Chemical Society. 2002; 124:9704–5. [PubMed: 12175218]
27. Shahid SA, Bardiaux B, Franks WT, Krabben L, Habeck M, van Rossum BJ, Linke D. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. Nature Methods. 2012; 9:1212–U119. [PubMed: 23142870]
28. Wang S, Munro RA, Shi L, Kawamura I, Okitsu T, Wada A, Kim SY, Jung KH, Brown LS, Ladizhansky V. Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. Nature Methods. 2013; 10:1007–12. [PubMed: 24013819]
29. Van Melckebeke H, Wasmer C, Lange A, Eiso AB, Loquet A, Bockmann A, Meier BH. Atomic-Resolution Three-Dimensional Structure of HET-s(218–289) Amyloid Fibrils by Solid-State NMR Spectroscopy. Journal of the American Chemical Society. 2010; 132:13765–75. [PubMed: 20828131]
30. Loquet A, Giller K, Becker S, Lange A. Supramolecular Interactions Probed by ^{13}C–^{13}C Solid-State NMR Spectroscopy. Journal of the American Chemical Society. 2010; 132:15164–6. [PubMed: 20932028]
31. Brunger AT. Free R-Value - A Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. Nature. 1992; 355:472–5. [PubMed: 18481394]
32. Wall JS, Hainfeld JF. Mass mapping with the scanning transmission electron microscope. Annual review of biophysics and biophysical chemistry. 1986; 15:355–76.
33. Galkin VE, Schmied WH, Schraidt O, Marlovits TC, Egelman EH. The structure of the Salmonella typhimurium type III secretion system needle shows divergence from the flagellar system. Journal of molecular biology. 2010; 396:1392–7. [PubMed: 20060835]
34. Maki-Yonekura S, Yonekura K, Namba K. Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. Nature Structural & Molecular Biology. 2010; 17:417–U50.
35. DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I. Modeling symmetric macromolecular structures in Rosetta3. PLOS One. 2011; 6:e20450. [PubMed: 21731614]
36. Herrmann T, Güntert P, Wüthrich K. Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. Journal of molecular biology. 2002; 319:209–27. [PubMed: 12051947]
37. Huang YPJ, Swapna GVT, Rajan PK, Ke HP, Xia B, Shukla K, Inouye M, Montelione GT. Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from Escherichia coli. Journal of molecular biology. 2003; 327:521–36. [PubMed: 12628255]
38. Zhang Z, Porter J, Tripsianes K, Lange OF. Robust and highly accurate automatic NOESY assignment and structure determination with Rosetta. Journal of Biomolecular NMR. 2014; 59:135–45. [PubMed: 24845473]
39. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YEA, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ,

Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods in enzymology. 2011; 487:545–74. [PubMed: 21187238]

40. Cordes FS, Komoriya K, Larquet E, Yang SX, Egelman EH, Blocker A, Lea SM. Helical structure of the needle of the type III secretion system of Shigella flexneri. J Biol Chem. 2003; 278:17103–7. [PubMed: 12571230]
41. Radics J, Königsmaier L, Marlovits TC. Structure of a pathogenic type 3 secretion system in action. Nat Struct Mol Biol. 2014; 21:82–7. [PubMed: 24317488]
42. Kenjale R, Wilson J, Zenk SF, Saurya S, Picking WL, Picking WD, Blocker A. The needle component of the type III secreton of Shigella regulates the activity of the secretion apparatus. J Biol Chem. 2005; 280:42929–37. [PubMed: 16227202]
43. Bockmann A, Gardiennet C, Verel R, Hunkeler A, Loquet A, Pintacuda G, Emsley L, Meier BH, Lesage A. Characterization of different water pools in solid-state NMR protein samples. Journal of Biomolecular NMR. 2009; 45:319–27. [PubMed: 19779834]
44. Fung BM, Khitrin AK, Ermolaev K. An improved broadband decoupling sequence for liquid crystals and solids. Journal of Magnetic Resonance. 2000; 142:97–101. [PubMed: 10617439]
45. Thakur RS, Kurur ND, Madhu PK. Swept-frequency two-pulse phase modulation for heteronuclear dipolar decoupling in solid-state NMR. Chemical Physics Letters. 2006; 426:459–63.
46. Augustine C, Kurur ND. Supercycled SWf-TPPM sequence for heteronuclear dipolar decoupling in solid-state nuclear magnetic resonance. Journal of Magnetic Resonance. 2011; 209:156–60. [PubMed: 21306928]
47. Chandran CV, Madhu PK, Kurur ND, Brauniger T. Swept-frequency two-pulse phase modulation (SWf-TPPM) sequences with linear sweep profile for heteronuclear decoupling in solid-state NMR. Magn Reson Chem. 2008; 46:943–7. [PubMed: 18666219]
48. Shaka AJ, Keeler J, Frenkiel T, Freeman R. An Improved Sequence for Broad-Band Decoupling - WALTZ-16. Journal of Magnetic Resonance. 1983; 52:335–8.
49. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPIPE: a multidimensional spectral processing system based on UNIX pipes. Journal of Biomolecular NMR. 1995; 6:277–93. [PubMed: 8520220]
50. Goddard TD, Kneller DG. SPARKY 3 v. 3.113. 2006
51. Stevens TJ, Fogh RH, Boucher W, Higman VA, Eisenmenger F, Bardiaux B, van Rossum BJ, Oschkinat H, Laue ED. A software framework for analysing solid-state MAS NMR data. Journal of Biomolecular NMR. 2011; 51:437–47. [PubMed: 21953355]
52. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins-Structure Function and Bioinformatics. 2005; 59:687–96.
53. Hong M. Determination of multiple phi-torsion angles in proteins by selective and extensive ^{13}C labeling and two-dimensional solid-state NMR. Journal of Magnetic Resonance. 1999; 139:389–401. [PubMed: 10423377]
54. Lundström P, Teilum K, Carstensen T, Bezsonova I, Wiesner S, Hansen DF, Religa TL, Akke M, Kay LE. Fractional C-13 enrichment of isolated carbons using [1-C-13]- or [2-C-13]-glucose facilitates the accurate measurement of dynamics at backbone C-alpha and side-chain methyl positions in proteins. Journal of Biomolecular NMR. 2007; 38:199–212. [PubMed: 17554498]
55. Das R, Andre I, Shen Y, Wu Y, Lemak A, Bansal S, Arrowsmith CH, Szyperski T, Baker D. Simultaneous prediction of protein folding and docking at high resolution. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:18978–83. [PubMed: 19864631]
56. Vernon R, Shen Y, Baker D, Lange OF. Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. Journal of Biomolecular Nmr. 2013; 57:117–27. [PubMed: 23975356]
57. DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D. Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta. Journal of molecular biology. 2009; 392:181–90. [PubMed: 19596339]

58. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:10037–41. [PubMed: 11517324]

59. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009; 25:1189–91. [PubMed: 19151095]

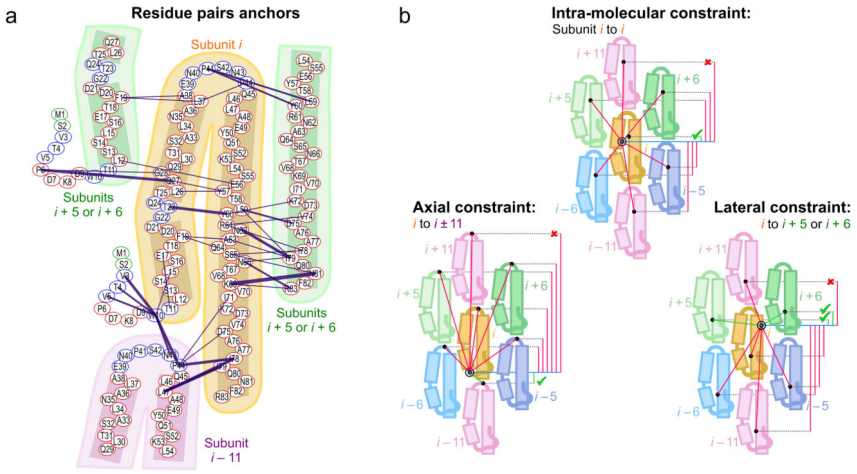

Figure 1. Preliminary map of the *Shigella flexneri* needle architecture

(a) Anchor point identification in the absence of previous structural knowledge. Proximity anchors are identified by at least two unambiguous (corresponding to well-resolved and readily assigned cross-peaks) or abundant (≥ 4) contacts in spectra of [1-^{13}C]-glucose and [2-^{13}C]-glucose labeled MxiH needles (blue lines, line thickness proportional to the number of contacts). **(b)** Strategy for assignment of ssNMR constraints arising from distinct intra-molecular and inter-molecular interfaces in helical assemblies. Pink lines indicate interface assignments of a given contact that are inconsistent with the axial displacement among symmetry-related subunits, and green lines the correct interface assignments.

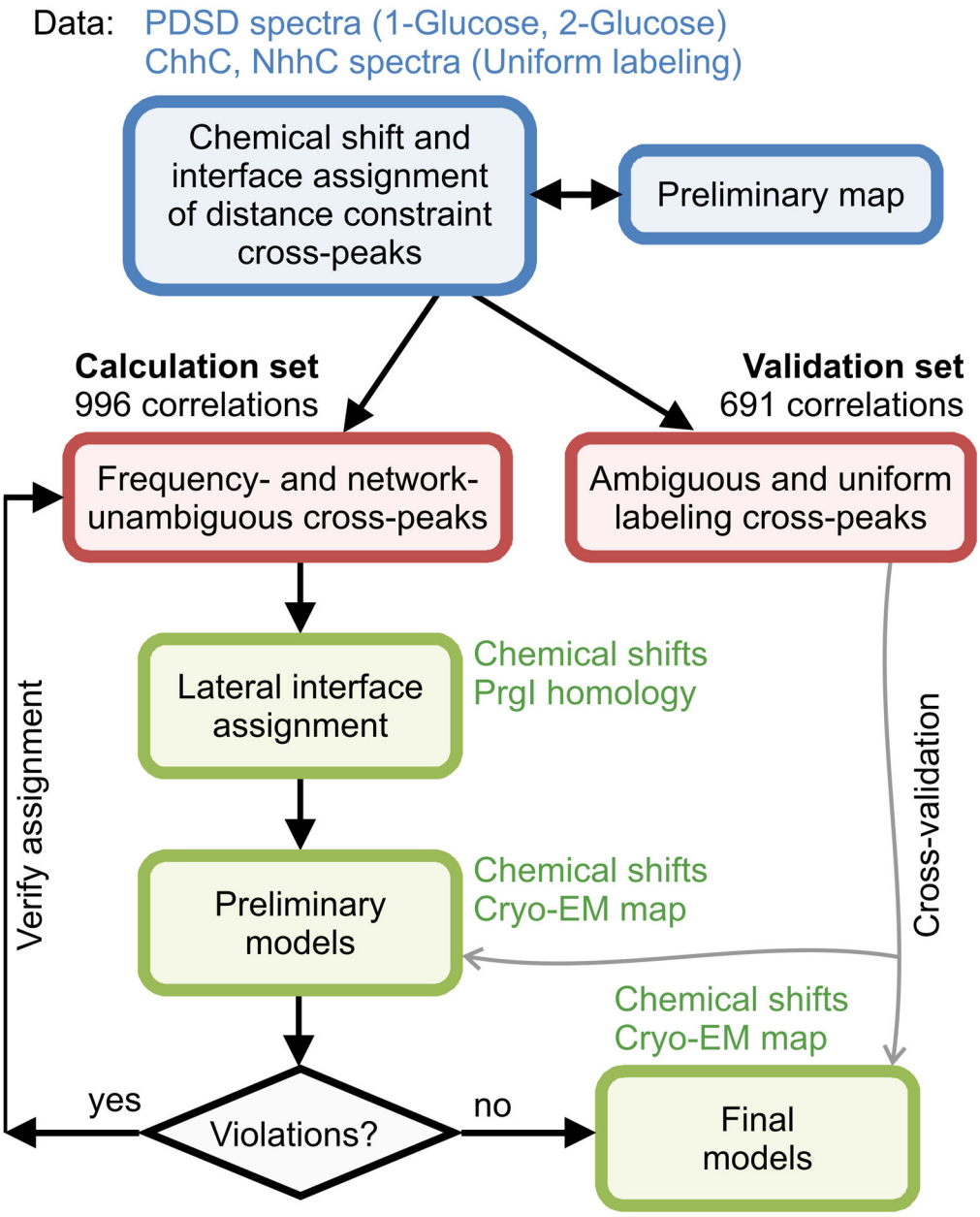

Figure 2. Flowchart of the iterative assignment/structure determination process

Flow chart presenting the assignment procedure of ssNMR distance constraints and structural modeling in the new hybrid approach. Steps of NMR data analysis are colored blue, with the resulting datasets illustrated as red rounded rectangles. Rosetta modeling steps are colored green and the additional data used at each modeling step are indicated, including the 7.7 Å cryo-EM density map[22]. The preliminary models are derived from a reduced set of distance constraints coming only from [2-^{13}C]-glucose spectra. The final models employ distance constraints from all datasets. The use of the software tools performing the different steps in the flow chart (provided as Supplementary Software 1) is outlined in detail in Supplementary Methods.

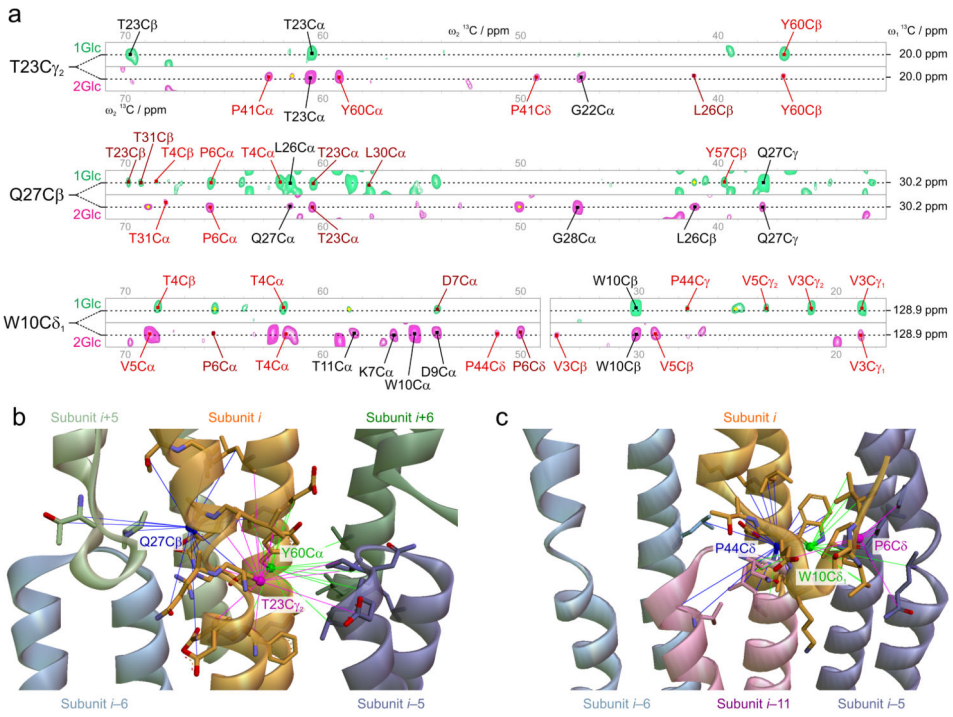

Figure 3. Identification of ssNMR contacts leading to high-resolution structural features

(a) PDSD spectra recorded using [1-^{13}C]-glucose labeled (green) and [2-^{13}C]-glucose labeled (magenta) MxiH needle samples. Strips are extracted for three indicated resonance positions, corresponding to the assignments of ^{13}C nuclei T23Cγ_2, Q27Cβ and W10Cδ_1. Assigned cross-peak labels are colored black for intra-residue, dark red for medium-range and light red for long-range atom pair contacts. Ambiguous cross-peaks are marked with a yellow rhombus. **(b, c)** Converged structural features of the T3SS needle highlighting the quality of ssNMR constraints used, for atoms in (b): Q27Cβ from subunit *i* (blue), T23Cγ_2 from subunit *i* (pink), and Y60Cα from subunit *i* (green), and in (c) P44Cδ from subunit *i–11* (blue), W10Cδ_1 from subunit *i* (green), and P6Cδ from subunit *i* (pink).

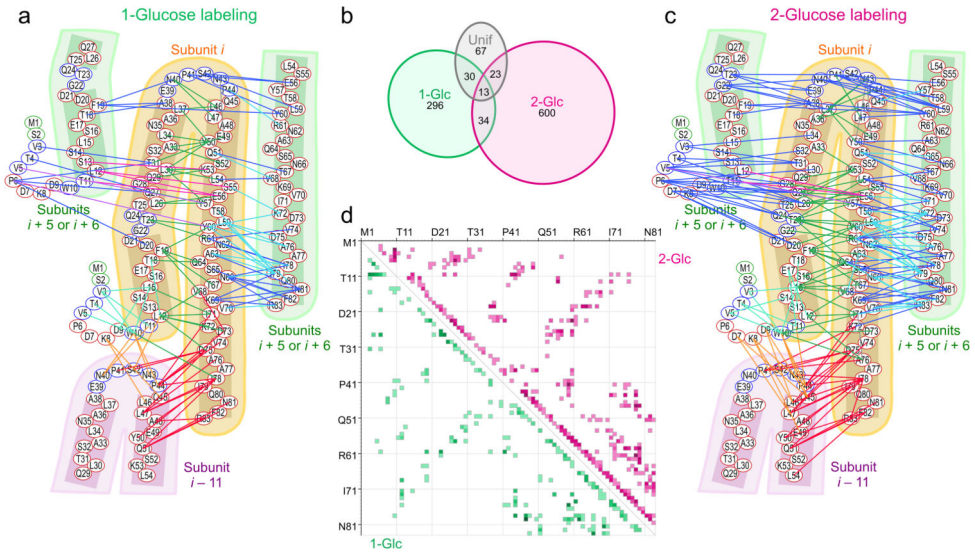

Figure 4. Long-range ssNMR constraints detected using two complementary isotopic labeling schemes

Chemical-shift unambiguous long-range constraints identified in **(a)** [1-^{13}C]-glucose and **(c)** [2-^{13}C]-glucose labeled samples. In (a), (c) interactions are color-coded according to the connecting subunits as follows: Green, cyan: intra-subunit. Blue, light blue and magenta: interface *i*, *i* ± 5 or *i*, *i* ± 6; Orange, red: interface *i*, *i* ± 11; **(b)** Partition diagram of the ssNMR constraints identified in this study considering the different ^{13}C-labeling schemes. **(d)** Residue-residue contact map of the ssNMR constraints, showing medium to long-range correlations (|*i*-*j*|>*2*) present in the data.

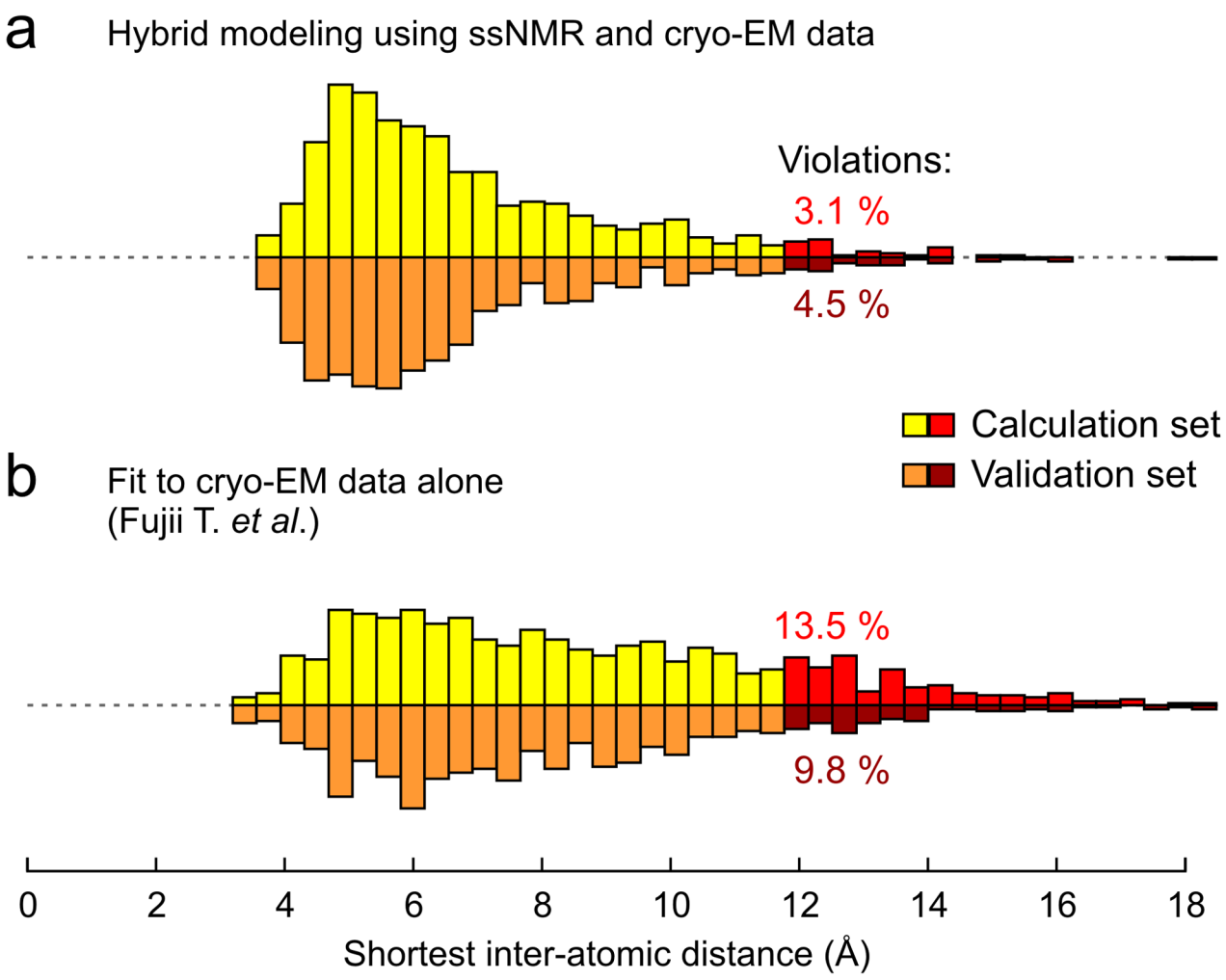

Figure 5. Cross-validation of different deposited MxiH needle models

Histogram of shortest identified distances for **(a)** a final model obtained from the hybrid structure determination approach presented here (PDB ID 2MME) and **(b)** the model reported by Fujii and coworkers by fitting of the 7.7 Å cryo-EM density map (EMD 5352) alone[22] (PDB ID 3J0R). For each observed cross-peak part of the calculation set (yellow, 996 correlations) and validation set (orange, 691 correlations), the shortest distance is calculated considering all possible chemical shift assignments within the chemical shift tolerance window (± 0.15 ppm for ^{13}C chemical shifts) and all 7 possible inter-molecular interfaces (see Fig. 1b). By considering only the shortest distance, this procedure prevents the introduction of any bias in the cross-validation statistics arising from manual peak picking of the ssNMR spectra. The validation set was not used at any time in the structure calculations. Experimentally observed correlations that have an inter-nuclear distance above 12 Å in each model are classified as violations (red histograms). The present model (a) fits the validation dataset by an average of 1.22 Å per distance constraint shorter than (b) (paired difference Student t-test, 95 % confidence level).

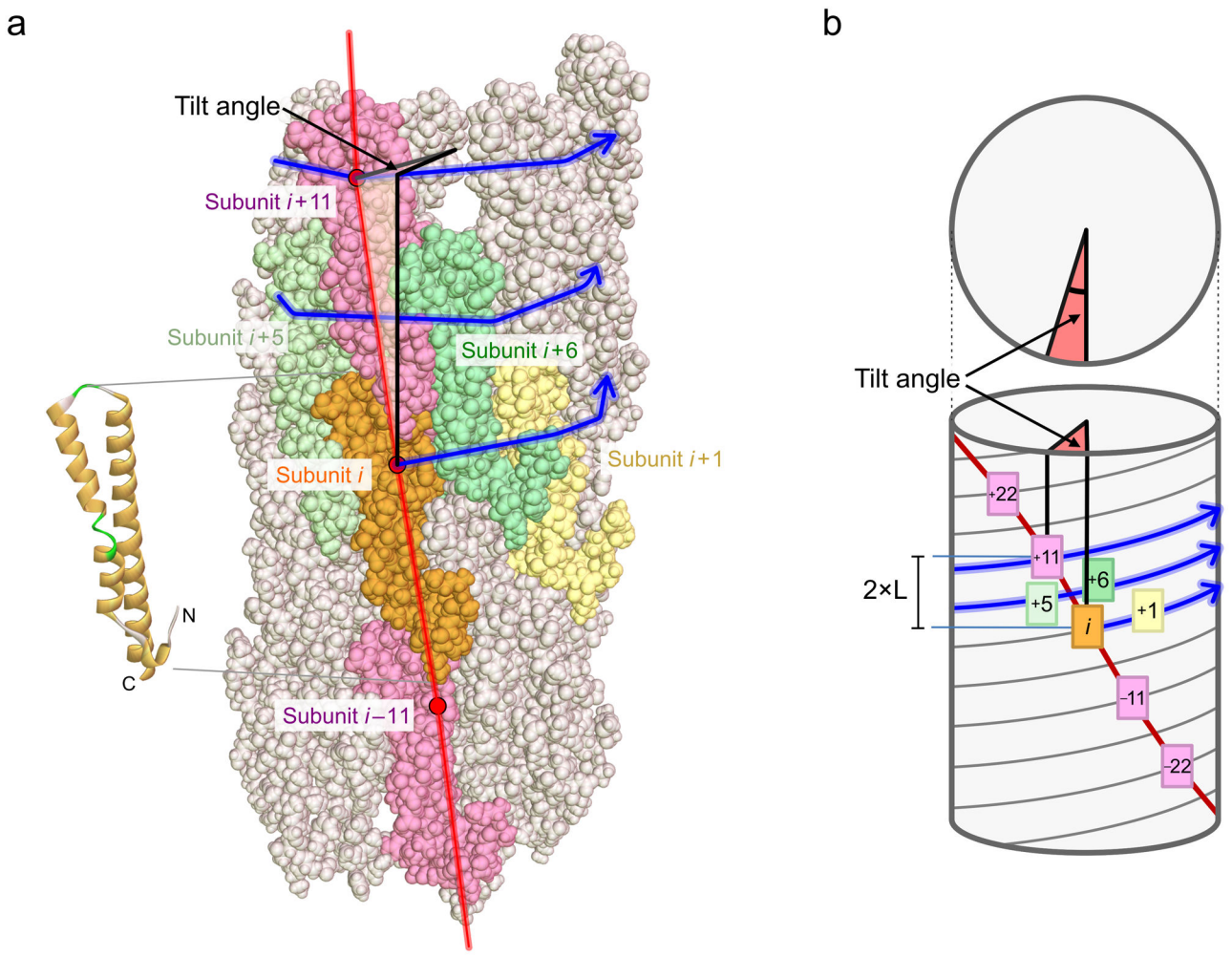

Figure 6. Tilt angle of the needle filament

(a) Side view of the hybrid needle structure colored according to subunit number. The secondary structure of subunit (*i*) is presented on the left. Blue lines connect successive subunits (*i*, *i*+1, *i*+2, *i*+3, …) and form a right-handed spiral. The helical axis of the filament is indicated as a bold black line. The red line connects subunits which share an axial interface (*i*, *i*+11, *i*+22, *i*+33, …). Subunit *i*+11 (pink) is not directly above subunit (*i*) as there are N=11.2 protein subunits per two turns of the helix (blue line). As a result, the filament has a slight tilt and the red line forms a slowly rotating left-handed spiral. **(b)** Schematic representation of the needle filament and tilt angle. The tilt angle (red) is positive in the counter-clockwise direction and can be calculated by $\alpha = (360°/N) \times (11 - N)$, where N is the number of subunits per two turns. The axial translation between subunit (*i*) and subunit (*i*+11) is indicated (2×L).

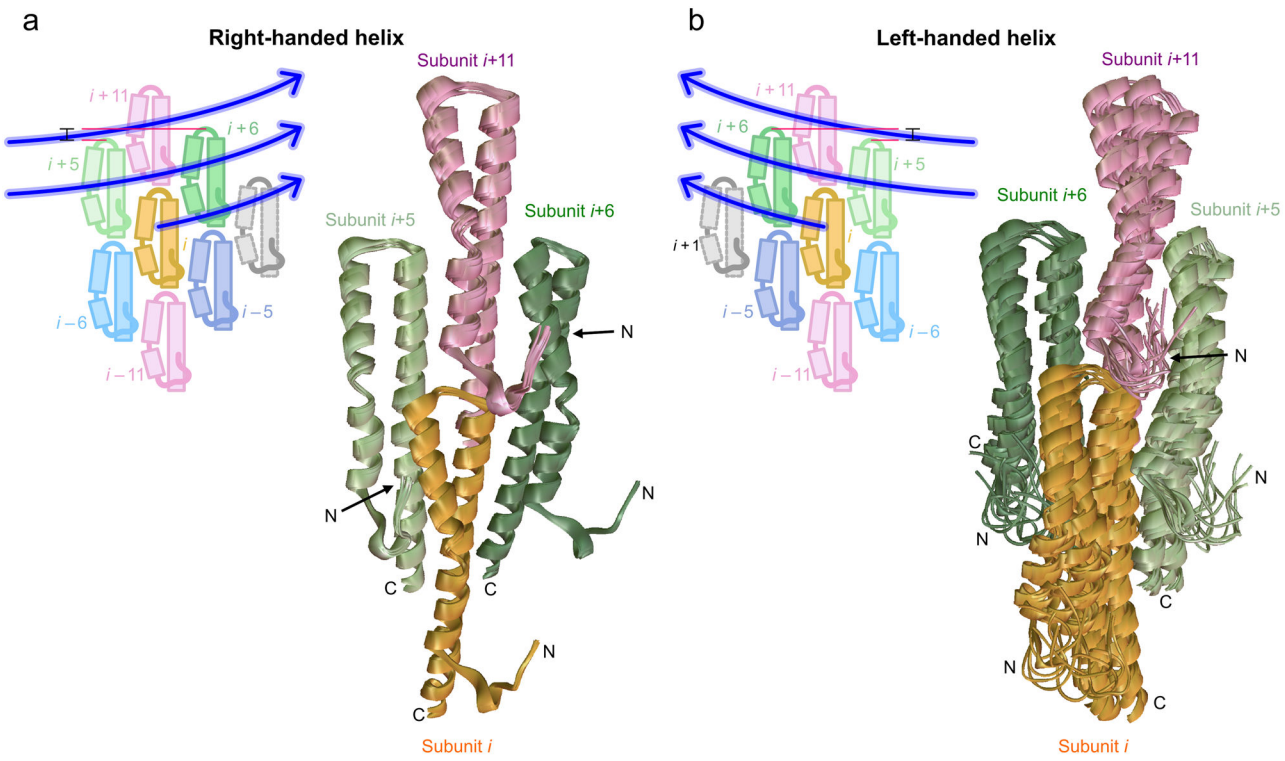

Figure 7. Superposition of right-handed and left-handed structural ensembles

The 10 top-ranking models ensemble produced by the hybrid structure determination approach considering **(a)** a right-handed helix and **(b)** a left-handed helix. The basic tetramer formed by the symmetric subunits (*i*, orange), (*i*+5, light green), (*i*+6, dark green) and (*i*+11, pink) is presented and N- and C- termini are indicated. The relative position of protein subunits is explained schematically in the top-left corner of **(a)** and **(b)** for the two handedness arrangements. In a right-handed helix, the subunit (*i*+5), closer to (*i*) in terms of axial translation, is located to the left of (*i*) while subunit (*i*+6), further away from (*i*) in terms of axial translation, is located to the right of it. Blue lines connect successive subunits as in Fig. 6. The right-handed ensemble shows much higher convergence compared to the left-handed ensemble. The right-handed ensemble presents less NMR constraint violations and has a more favorable full-atom energy, unequivocally identifying it as the correct handedness.

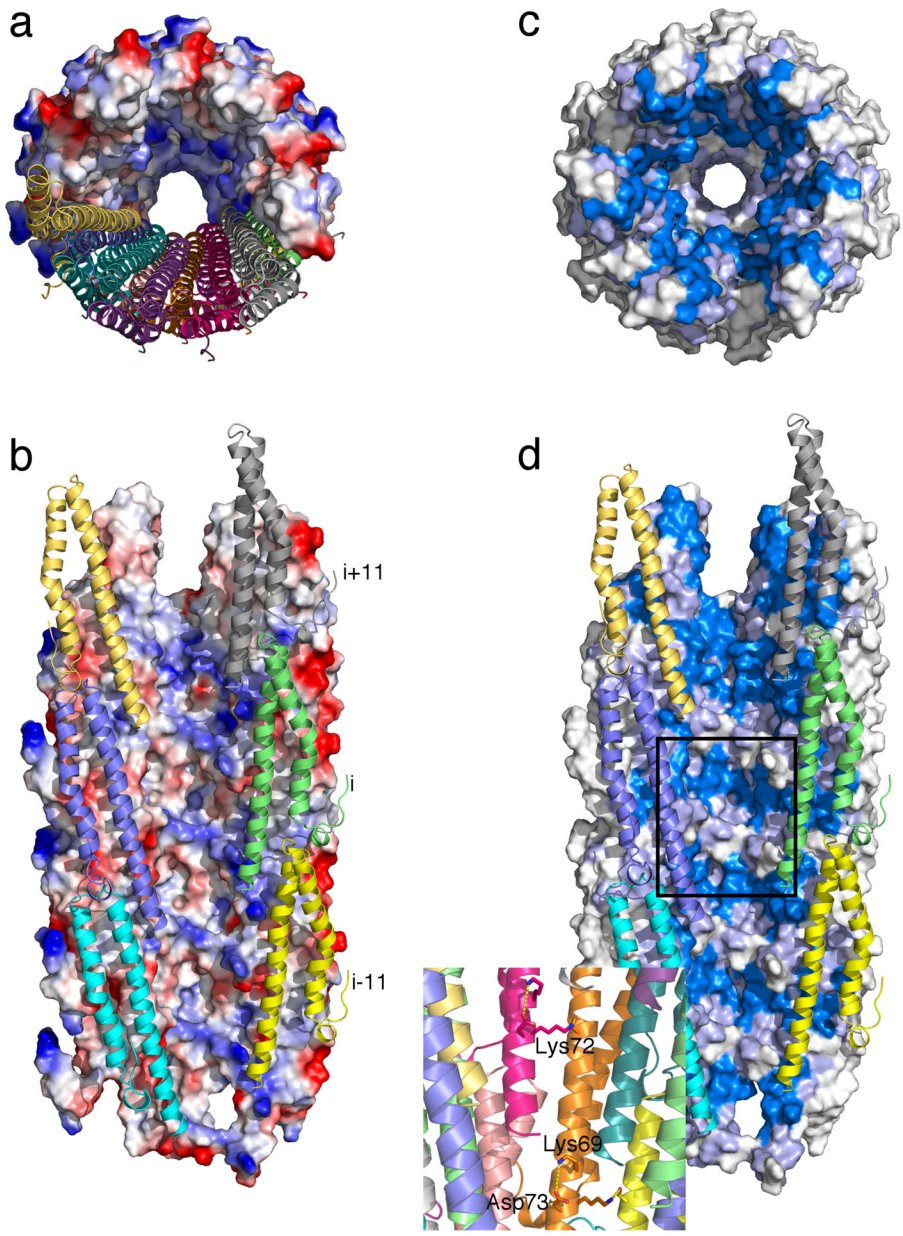

Figure 8. Conserved charged residues decorate the MxiH needle pore

(a) Top and **(b)** side view of the hybrid needle structure showing the surface electrostatic potential computed using APBS[58] and contoured linearly in the range ±5 kT/e. The helical arrangement of subunits is shown as cartoons for one-half of the full 29-subunit system as a guide. In (b, d) three 11-start protofilaments have been removed for simplicity. **(c, d)** Same view as in (a, b), with surface coloring according to the sequence conservation index calculated from the alignment of 102 unique needle sequences using Jalview[59], measured on a 1–10 scale. Marine color indicates high (7–10), light purple medium (4–6) and white low (1–3) sequence conservation index. The boxed region in (d) is enlarged in the inset,

indicating the pattern of highly conserved charged residues on the helical structure, as discussed in the main text.

Table 1

NMR constraint and refinement statistics

NMR distance & dihedral constraints	
Total distance constraints (calculation set)	996
Intra-subunit	580
Medium-range ($2<\|j-k\| \leq 4$)	391
Long-range ($\|j-k\| \geq 5$)	189
Inter-subunit	
Lateral: (i) to ($i \pm 5$)	188
Lateral: (i) to ($i \pm 6$)	104
Axial: (i) to ($i \pm 11$)	124
Hydrogen bonds	0
Total distance constraints (validation set)*	691
Uniform unambiguous	96
1-Glc ambiguous	230
2-Glc ambiguous	221
Uniform ambiguous	144
Total dihedral angle constraints**	162
φ	81
ψ	81
Structure statistics	
Violations (mean and s.d.)	22.5/31.0+
Distance constraints (Å)	0.9/0.4
Dihedral angle (°)	0/0
max. distance constraint violation (Å)	2.2
max. dihedral angle violation (°)	0
Deviations from idealized geometry	
Bond lengths (Å)	0.01
Bond angles (°)	0.7
Impropers (°)	0
Average pairwise r.m.s. deviation (Å)++	
Heavy	0.7
Backbone	0.4

*Not used in the calculations, but reserved as a validation set.

**Used to bias the selection of 3mer and 9mer backbone fragments from the PDB[56], as outlined in methods.

+Total Violations are reported separately for the calculation/validation sets.

++Computed over all members of the NMR ensemble superimposed to the coordinates of the first model.