nature structural & molecular biology

Visualizing chaperone-assisted protein folding

Scott Horowitz[1,2,10], Loïc Salmon[1,2,10], Philipp Koldewey[1,2,10], Logan S Ahlstrom[3,9], Raoul Martin[1,2,9], Shu Quan[4], Pavel V Afonine[5], Henry van den Bedem[6], Lili Wang[1,2], Qingping Xu[7], Raymond C Trievel[8], Charles L Brooks III[3] & James C A Bardwell[1,2]

Challenges in determining the structures of heterogeneous and dynamic protein complexes have greatly hampered past efforts to obtain a mechanistic understanding of many important biological processes. One such process is chaperone-assisted protein folding. Obtaining structural ensembles of chaperone–substrate complexes would ultimately reveal how chaperones help proteins fold into their native state. To address this problem, we devised a new structural biology approach based on X-ray crystallography, termed residual electron and anomalous density (READ). READ enabled us to visualize even sparsely populated conformations of the substrate protein immunity protein 7 (Im7) in complex with the *Escherichia coli* chaperone Spy, and to capture a series of snapshots depicting the various folding states of Im7 bound to Spy. The ensemble shows that Spy-associated Im7 samples conformations ranging from unfolded to partially folded to native-like states and reveals how a substrate can explore its folding landscape while being bound to a chaperone.

High-resolution structural models of protein-protein interactions are critical for obtaining mechanistic insights into biological processes[1]. However, many protein-protein interactions are highly dynamic, thus making it difficult to obtain high-resolution data. Particularly challenging are the interactions of intrinsically or conditionally disordered sections of proteins with their partner proteins. Recent advances in X-ray crystallography[2] and NMR spectroscopy[3,4] have improved the ability to analyze biomolecules that exist in multiple conformations. X-ray crystallography has historically provided valuable information on small-scale conformational changes, but observing large-amplitude heterogeneous conformational changes often falls beyond the reach of current crystallographic techniques. NMR can theoretically be used to determine heterogeneous ensembles[5,6], but in practice, this proves to be very challenging.

Despite the importance of understanding how proteins fold into their native state within the cell, knowledge about this critical process remains limited. It is clear that molecular chaperones aid in protein folding. However, exactly how they facilitate the folding process is still being debated[7,8]. Structural characterization of chaperone-assisted protein folding would probably help to bring clarity to this question. Structural models of chaperone–substrate complexes have recently begun to provide information as to how a chaperone can recognize its substrate[9–18]. However, the effects that chaperones have on their substrates, and how these interactions affect the folding process remain largely unknown. For most chaperones, it is still unclear whether chaperones actively participate in and affect the folding of substrate proteins or merely provide a suitable microenvironment enabling substrates to fold on their own. Although this is a truly fundamental question in the chaperone field, its answer has eluded the research community largely because of the highly dynamic nature of the chaperone–substrate complexes.

To address this question, we investigated the ATP-independent *E. coli* periplasmic chaperone Spy. Spy prevents protein aggregation and aids in protein folding under various stress conditions, including treatment with tannin and butanol[19]. We originally discovered Spy on the basis of its ability to stabilize the protein-folding model Im7 (refs. 19–21) *in vivo* and have recently demonstrated that Im7 folds while it is associated with Spy[22]. The Spy crystal structure reveals that Spy forms a thin α-helical homodimeric cradle[19,23]. Cross-linking and genetic experiments have suggested that Spy interacts with substrates somewhere on its concave side[19,24]. By using a new X-ray crystallography–based approach to model disorder in crystal structures, we determined the high-resolution ensemble of the dynamic Spy–Im7 complex. This work provides a detailed view of chaperone-mediated protein folding and shows how substrates such as Im7 find their native fold while they are bound to their chaperones.

RESULTS

Crystallizing the Spy–Im7 complex

We reasoned that to obtain crystals of complexes between Spy (domain boundaries in **Supplementary Fig. 1**) and its substrate proteins, the best approach would be to identify crystallization conditions that yielded Spy crystals in the presence, but not the absence, of protein substrates.

[1]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan, USA. [2]Howard Hughes Medical Institute, Ann Arbor, Michigan, USA. [3]Department of Chemistry and Biophysics Program, University of Michigan, Ann Arbor, Michigan, USA. [4]State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai Collaborative Innovation Center for Biomanufacturing, Shanghai, China. [5]Lawrence Berkeley National Laboratory, Berkeley, California, USA. [6]Division of Biosciences, SLAC National Accelerator Laboratory, Stanford University, Stanford, California, USA. [7]Joint Center for Structural Genomics, Stanford Synchrotron Radiation Lightsource, SLAC National Laboratory, Menlo Park, California, USA. [8]Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan, USA. [9]Present addresses: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan, USA and Howard Hughes Medical Institute, Ann Arbor, Michigan, USA (L.S.A.) and Biophysics Graduate Group, University of California, Berkeley, Berkeley, California, USA (R.M.). [10]These authors contributed equally to this work. Correspondence should be addressed to S.H. (horowsah@umich.edu) or J.C.A.B. (jbardwel@umich.edu).

Received 12 October 2015; accepted 4 May 2016; published online 30 May 2016; doi:10.1038/nsmb.3237

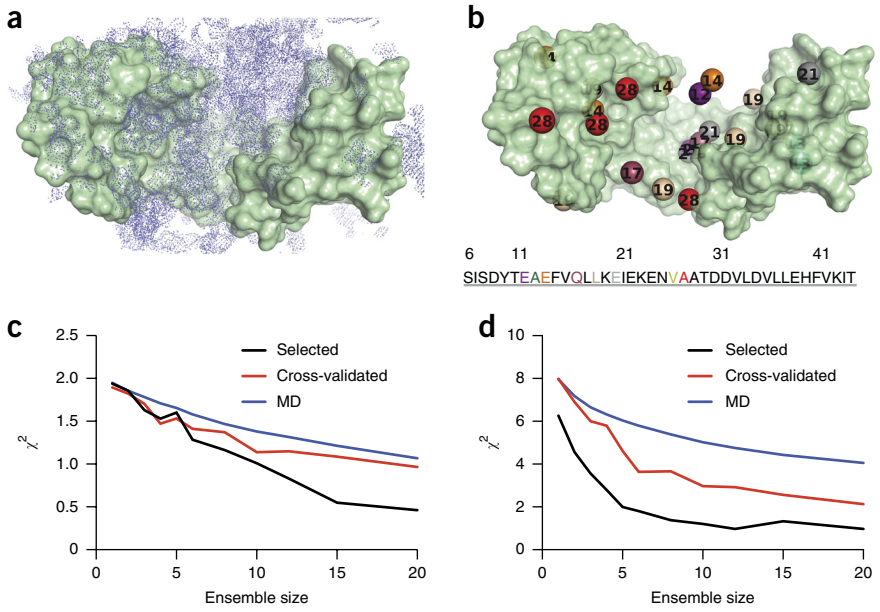

Figure 1 Crystallographic data and ensemble selection. (**a**) $2mF_o - DF_c$ omit map of residual $Im7_{6–45}$ and flexible linker electron density contoured at 0.5σ (blue dots). This is the residual density used in the READ selection. (**b**) Composites of iodine positions detected from anomalous signals by using pI-Phe substitutions, colored and numbered by sequence. Multiple iodine positions were detected for most residues. In **a** and **b**, Spy is rendered as a green surface. (**c**,**d**) Agreement of the residual $Im7_{6–45}$ electron density (**c**) and anomalous iodine signals (**d**) for ensembles of varying size generated by randomly choosing from the MD pool (blue) and from the selection procedure (black). The agreement from back-calculating a subset of data excluded from the selection procedure is shown by the red curve (cross-validation). The cost function, χ^2, decreases as the agreement to the experimental data increases and is defined in the Online Methods.

We therefore screened crystallization conditions for Spy with four different substrate proteins: a fragment of the largely unfolded bovine α-casein protein[19,25], wild-type (WT) *E. coli* Im7 (refs. 19,26), an unfolded variant of Im7 (L18A L19A L37A)[27], and the N-terminal half of Im7 ($Im7_{6–45}$), which encompasses the entire Spy-binding portion of Im7 (ref. 24). We found conditions in which all four substrates cocrystallized with Spy, but in which Spy alone did not yield crystals. Subsequent crystal washing and dissolution experiments confirmed the presence of the substrates in the cocrystals (**Supplementary Fig. 2**). The crystals diffracted to ~1.8-Å resolution. We used Spy–$Im7_{6–45}$ selenomethionine (SeMet) crystals for phasing with single-wavelength anomalous diffraction (SAD) experiments, and we then used this solution to build the well-ordered Spy portions of all four complexes. However, modeling of the substrate in the complex proved to be a substantial challenge because the electron density of the substrate was discontinuous and fragmented. Even the minimal binding portion of Im7 ($Im7_{6–45}$) showed highly dispersed electron density (**Fig. 1a**). We hypothesized that the fragmented density was due to multiple partially occupied conformations of the substrate bound within the crystal. Such residual density is typically not considered to be usable by traditional X-ray crystallography methods. Thus, we developed a new approach to interpret the chaperone-bound substrate in multiple conformations.

READ: a strategy to visualize heterogeneous and dynamic biomolecules

To determine the structure of the substrate portion of these Spy–substrate complexes, we conceived an approach that we term READ. We split this approach into five steps: (i) By using a well-diffracting Spy–substrate cocrystal, we first determined the structure of the folded domain of Spy and obtained high-quality residual electron density within the dynamic regions of the substrate. (ii) We then labeled individual residues in the flexible substrate with the strong anomalous scatterer iodine, which allows these residues to be located in 3D space on the basis of their anomalous density. (iii) We performed molecular dynamics (MD) simulations to generate a pool of energetically reasonable conformations of the dynamic complex and (iv) applied a sample-and-select algorithm to determine the minimal set of substrate conformations that fit both the residual and anomalous density. (v) Finally, we validated the ensemble through multiple statistical tests. Importantly, even though we labeled only a subset of the residues in the flexible substrate with iodine, the residual electron density provides spatial information on many of the other flexible residues[28,29]. These two forms of data are therefore complementary: by labeling individual residues, the residues can be located to specific points in space. The electron density then allowed us to connect the labeled residues of the substrate by confining the protein chain within regions of detectable density. In this way, the two forms of data together were able to describe multiple conformations of the substrate within the crystal. As described in detail below, we developed the READ method to uncover the ensemble of conformations that the Spy-binding domain of Im7 (i.e., $Im7_{6–45}$ (ref. 24)) adopts while it is bound to Spy. However, we believe that READ will be generally applicable to visualizing heterogeneous and dynamic complexes that have previously escaped detailed structural analysis.

Collecting READ data for the Spy–$Im7_{6–45}$ complex

To apply the READ technique to the folding mechanism used by the chaperone Spy, we selected $Im7_{6–45}$ for further investigation because NMR data suggested that $Im7_{6–45}$ is able to recapitulate unfolded, partially folded, and native-like states of Im7 (**Supplementary Fig. 3**). Moreover, binding experiments indicated that $Im7_{6–45}$ comprises the entire Spy-binding region[24]. To introduce the anomalous scatterer iodine, we replaced eight $Im7_{6–45}$ residues with the noncanonical amino acid 4-iodophenylalanine (pI-Phe). Its strong anomalous scattering[30] allowed us to track the positions of individual $Im7_{6–45}$ residues one at a time, even if the residue was found in several locations in the same crystal. We then cocrystallized Spy and the eight $Im7_{6–45}$ peptides, each of which contained an individual pI-Phe substitution at one distinct position, and collected anomalous data for all eight Spy–$Im7_{6–45}$ complexes (**Fig. 1b**, **Supplementary Tables 1** and **2** and **Supplementary Data Set 1**). In agreement with our electron density map, the majority of anomalous signals emerged in the cradle of Spy, thus implying that this region is the likely Im7-substrate-binding site. However, in agreement with the fragmented density, there were multiple iodine positions for seven of the eight substituted residues. Together, these results indicated that the Im7 substrate binds Spy in multiple conformations.

READ sample-and-select procedure

To determine the structural ensemble that $Im7_{6–45}$ adopts while it is bound to Spy, we combined the residual electron density and the

anomalous signals from our pI-Phe-substituted Spy–Im7$_{6–45}$ complexes. To generate an accurate depiction of the chaperone-substrate interactions, we devised a selection protocol based on a sample-and-select procedure used in NMR spectroscopy[4]. This procedure (READ sample-and-select algorithm diagrammed in **Fig. 2**) iteratively constructs structural ensembles and then compares them to the experimental data. During each round of the selection, a genetic algorithm alters the ensemble, and its agreement with the experimental data is reevaluated. If successful, the selection identifies the smallest group of specific conformations that best fit the residual electron density and anomalous signals.

Before performing the selection, we generated a large and diverse pool of chaperone–substrate complexes by using coarse-grained MD simulations in a pseudocrystal environment (**Fig. 2** and **Supplementary Fig. 4**). The coarse-grained simulations were based on a single-residue-resolution model for protein folding[31] and were extended here to describe Spy-Im7$_{6–45}$ binding events (Online Methods). The initial conditions of the binding simulations were not biased toward a particular conformation of the substrate or any specific chaperone-substrate interaction (Online Methods). Im7$_{6–45}$ binds and unbinds to Spy throughout the simulations. This strategy allows a wide range of substrate conformations to interact with the chaperone. From the MD simulations, we extracted ~10,000 diverse Spy–Im7$_{6–45}$ complexes to be used by the ensuing selection. Each complex within this pool comprised one Spy dimer bound to a single Im7$_{6–45}$ substrate. This pool was then used by the selection algorithm to identify the minimal ensemble that best satisfies both the residual electron and anomalous crystallographic data.

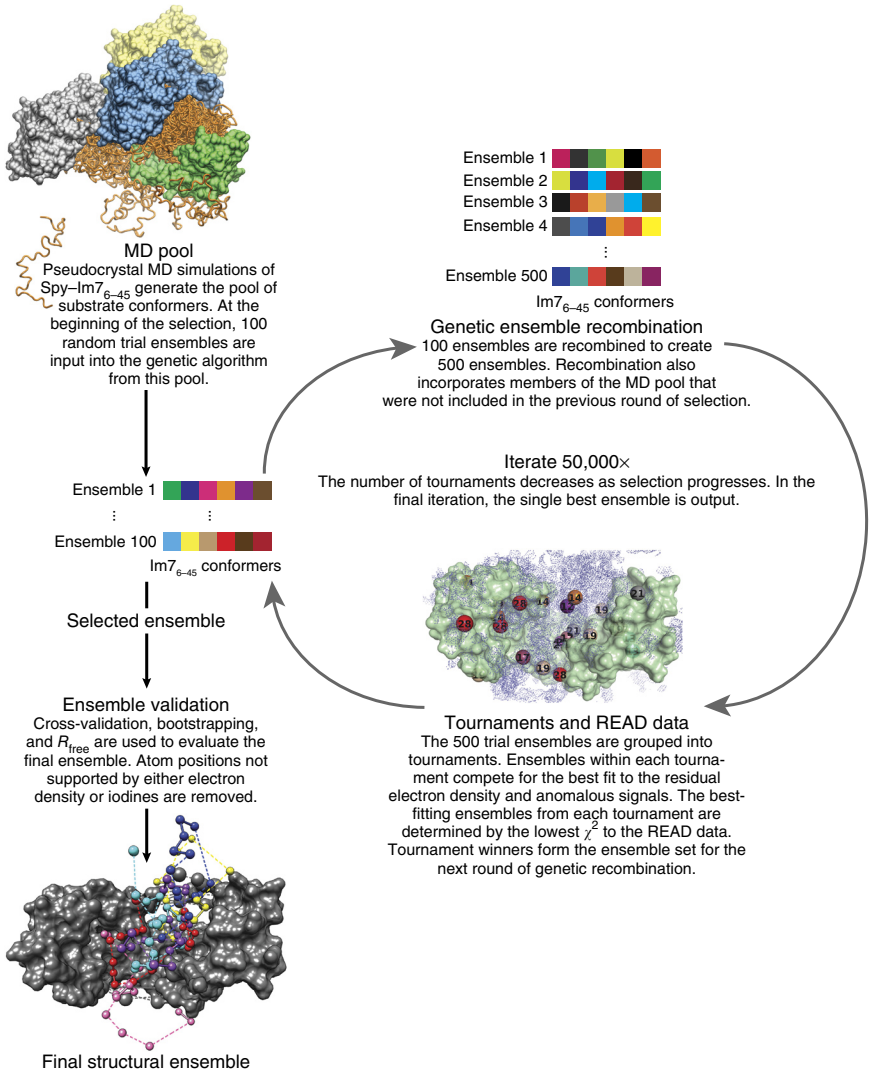

Figure 2 Flowchart of the READ sample-and-select process.

The anomalous scattering portion of the selection uses basic knowledge of pI-Phe geometry: the iodine is separated from its respective Cα atom in each coarse-grained conformer by 6.5 Å. The selection then picks ensembles that best reproduce the collection of iodine anomalous signals. Simultaneously, it uses the residual electron density to choose ensembles. To make the electron density selection practical, we needed to develop a method to rapidly evaluate the agreement between the selected subensembles and the experimental electron density during the selection procedure. To accomplish this task, we generated a compressed version of the experimental $2mF_o - DF_c$ electron density map for use in the selection. This process provided us with a target map that the ensuing selection attempted to recapitulate. To decrease the extent of 3D space to be explored, we created this compressed map by using only density from regions of space significantly sampled by Im7$_{6–45}$ in the Spy–Im7$_{6–45}$ MD simulations. For each of the ~10,000 complexes in the coarse-grained MD pool, we extracted the electron density at the Cα positions of Im7$_{6–45}$ and used it to construct an electron density map (Online Methods). We then combined these individual electron density maps from the separate conformers (**Fig. 2**) and compared them with the averaged experimental electron density map as part of the selection algorithm.

This approach allowed us to simultaneously use both the iodine anomalous signals and the residual electron density in the selection procedure. The selection resulted in small ensembles from the MD pool that best fit the READ data (**Fig. 1c**,**d**). Before analyzing the details of the Spy–Im7$_{6–45}$ complex, we first performed a series of validation tests to verify the ensemble and selection procedure (**Supplementary Note 1**, **Fig. 1c**,**d** and **Supplementary Figs. 5**–**7**). Together, these validation tests confirmed that the selection procedure and selected six-member ensemble recapitulated the experimental data. Notably, the final six-membered ensemble was the largest ensemble able to both decrease the R_{free} and pass the ten-fold cross-validation test. This ensemble depicts the conformations that the substrate Im7$_{6–45}$ adopts while it is bound to the chaperone Spy (**Fig. 3**, **Supplementary Movie 1** and **Table 1**).

Folding and interactions of Im7 bound to Spy

Our results showed that by using this READ approach, we were able to obtain structural information about the dynamic interaction of a chaperone with its substrate protein. We were

Table 1 Data collection and refinement statistics

	SeMet Spy–Im7$_{6–45}$	Spy–Im7$_{6–45}$ (PDB 5INA)	Spy–casein$_{148–177}$; substrate not modeled (PDB 5IOG)	Spy H96L–Im7 L18A L19A L37A; substrate not modeled (PDB 5IOE)	Spy H96L–WT Im7; substrate not modeled (PDB 5IOA)
Data collection					
Space group	$P4_122$	$P4_122$	$P4_122$	$P4_122$	$P4_122$
Cell dimensions					
a, *b*, *c* (Å)	42.9, 42.9, 259.3	42.9, 42.9, 260.2	43.0, 43.0, 258.2	43.1, 43.1, 258.7	43.1, 43.14, 260.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	64.82–2.44 (2.53–2.44)	30.50–1.79 (1.83–1.79)	36.88–1.77 (1.80–1.77)	30.48–1.87 (1.91–1.87)	33.21–1.87 (1.91–1.87)
R_{merge} (%)	10.6 (36)	8.2 (108)	6.2 (134)	8.4 (152)	9.6 (249)
$I/\sigma(I)$	15.1 (6.8)	7.0 (1.1)	15.3 (1.6)	13.8 (1.8)	13.2 (1.3)
Completeness (%)	100 (100)	94.0 (90.1)	99.9 (99.5)	100 (100)	96.8 (93.1)
Redundancy	15.6 (15.6)	4.3 (4.2)	8.7 (8.2)	9.6 (9.4)	8.2 (8.2)
$CC_{1/2}$		0.998 (0.689)	0.999 (0.745)	0.999 (0.676)	0.998 (0.606)
Refinement					
Resolution (Å)		1.79	1.77	1.87	1.87
No. reflections		22,583	25,052	21,505	20,838
R_{work} / R_{free}		0.22 / 0.23	0.21 / 0.24	0.22 / 0.24	0.21 / 0.25
No. atoms		1,765	1,669	1,715	1,653
Protein		1,586	1,493	1,541	1,444
Ligand/ion		30	56	60	30
Water		149	120	114	179
B factors		49.4	48.5	47.4	39.2
Protein		49.0	47.5	46.3	38.3
Ligand/ion		48.6	65.9	80.4	62.9
Water		54.2	51.9	44.5	42.1
R.m.s. deviations					
Bond lengths (Å)		0.013	0.013	0.013	0.014
Bond angles (°)		1.24	1.30	1.24	1.39

Values in parentheses are for highest-resolution shell.

particularly interested in finding answers to one of the most fundamental questions in chaperone biology: how does chaperone binding affect the structure of the substrate? By analyzing the individual structures of the six-member ensemble of Im7$_{6–45}$ bound to Spy, we observed that Im7$_{6–45}$ takes on several different conformations while it is bound. These conformations are highly heterogeneous and include unfolded, partially folded, and native-like states (**Fig. 3**). The ensemble primarily encompasses Im7$_{6–45}$ lying diagonally within the Spy cradle in several different orientations, but some conformations traverse as far as the tips or even extend over the side of the cradle (**Figs. 3** and **4a**).

We constructed a contact map of the ensemble complex, which shows the frequency of interactions for chaperone–substrate residue pairs (**Fig. 4**). We found that the primary interaction sites on Spy reside at the N and C termini (Arg122, Thr124, and Phe29) as well as on the concave face of the chaperone (Arg61, Arg43, Lys47, His96, and Met46). The Spy-contacting residues include a mixture of charged, polar, and hydrophobic residues. Unexpectedly, in the ensemble, Im7$_{6–45}$

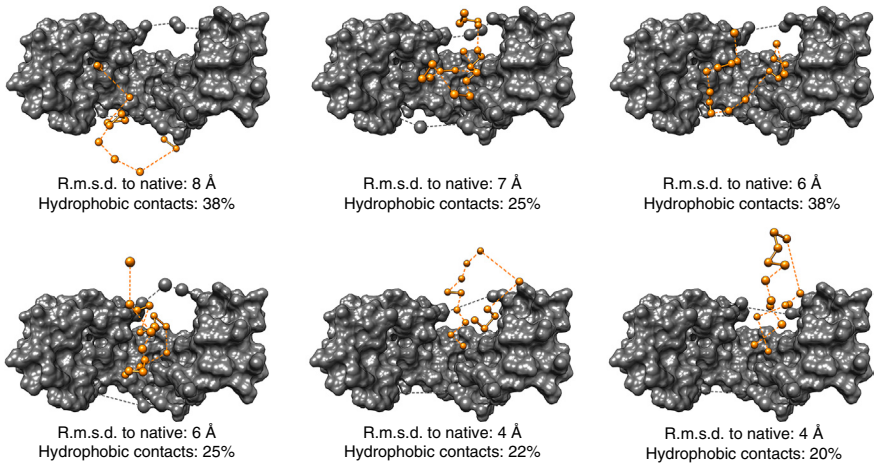

Figure 3 Spy–Im7$_{6–45}$ ensemble, arranged by r.m.s. deviation (r.m.s.d.) from the native state of Im7$_{6–45}$. Although the six-membered ensemble from the READ selection should be considered only as an ensemble, for clarity, the individual conformers are shown separately. Spy is depicted as a gray surface, and the Im7$_{6–45}$ conformer is shown as orange balls. Atoms that were not directly selected in the READ procedure or whose position could not be justified on the basis of agreement with the residual electron density were removed, thus leading to noncontiguous sections. Dashed lines connect noncontiguous segments of the Im7$_{6–45}$ substrate. Residues of the Spy flexible linker region that fit the residual electron density are shown as large gray spheres. Shown below each ensemble member is the r.m.s. deviation of each conformer from the native state of Im7$_{6–45}$, as well as the percentage of contacts between Im7$_{6–45}$ and Spy that are hydrophobic.

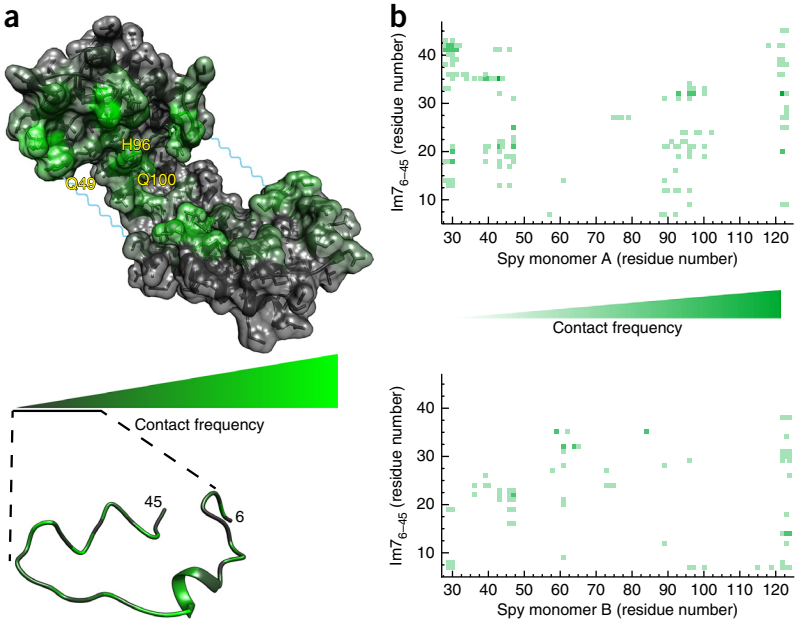

Figure 4 Contact maps of the Spy–Im7$_{6–45}$ complex. (**a**) Spy–Im7$_{6–45}$ contact map projected onto the bound Spy dimer (above) and Im7$_{6–45}$ (below) structures. For clarity, Im7$_{6–45}$ is represented in a single conformation. The frequency plotted is calculated as the average contact frequency from Spy to every residue of Im7$_{6–45}$ and vice versa. Because the residues involved in contacts are more evenly distributed in Im7$_{6–45}$ than Spy, its contact map was amplified. (**b**) Detailed contact maps of Spy–Im7$_{6–45}$. Contacts to the two Spy monomers are depicted separately. The flexible linker region of Spy (residues 47–57) is not represented in the 2D contact maps.

interacts with only 38% of the hydrophobic residues in the Spy cradle, but it interacts with 61% of the hydrophilic residues in the cradle. This distribution suggests the importance of both electrostatic and hydrophobic components in binding the Im7$_{6–45}$ ensemble. With respect to the substrate, we observed that nearly every residue in Im7$_{6–45}$ is in contact with Spy (**Fig. 4a**). However, despite this uniformity, we observed that regions of Im7$_{6–45}$ preferentially interact with different regions in Spy (**Fig. 4b**). For example, the N-terminal half of Im7$_{6–45}$ binds more consistently in the Spy cradle, whereas the C-terminal half predominantly binds the outer edges of Spy's concave surface.

As expected, we found that as Im7$_{6–45}$ progresses from the unfolded to the native state, its interactions with Spy shift accordingly. Whereas the least folded Im7$_{6–45}$ pose in the ensemble forms the highest number of hydrophobic contacts with Spy (**Fig. 3**), the two most folded conformations form the lowest number of hydrophobic contacts (**Fig. 3**). This shift in contacts is probably the result of hydrophobic residues of Im7$_{6–45}$ preferentially forming intramolecular contacts upon folding (i.e., hydrophobic collapse), thus effectively removing themselves from the interaction sites. The diversity of conformations and binding sites observed here emphasizes the dynamic and heterogeneous nature of the chaperone–substrate ensemble. Although we do not yet have time-resolution data of these various snapshots of Im7$_{6–45}$, this ensemble illustrates how a substrate samples its folding landscape while being bound by a chaperone.

Spy changes conformation upon substrate binding

Comparison of the structure of Spy in its substrate-bound and apo[19] states revealed that the Spy dimer also undergoes marked conformational changes upon substrate binding (**Fig. 5a** and **Supplementary Movie 2**). Upon substrate binding, the Spy dimer twists 9° about its center relative to its apo form[19]. This twist yields asymmetry and results in substantially different interaction patterns in the two Spy monomers (**Fig. 4b**). This twist may increase heterogeneity in Spy by providing more binding poses. Additionally, we observed that the linker region of Spy (residues 47–57), which participates in substrate interaction, becomes mostly disordered upon binding the substrate[16,19]. This increased disorder might explain how Spy is able to recognize and bind different substrates and/or differing conformations of the same substrate. Importantly, we observed the same structural changes in Spy regardless of which of the four substrates was bound (**Fig. 5b** and **Table 1**). The r.m.s. deviations between the well-folded sections of Spy in the four chaperone–substrate complexes were very small, less than 0.3 Å. From these findings as well as competition experiments showing that the substrates compete in solution for Spy binding (**Fig. 5c** and **Supplementary Fig. 8**), we conclude that all the tested substrates share the same overall Spy-binding site.

DISCUSSION

To shed light on how chaperones interact with their substrates, we developed a new structural biology method, READ, and applied it to determine a conformational ensemble of the chaperone Spy bound to a substrate. As a substrate, we used Im7$_{6–45}$, the chaperone-interacting portion of the protein-folding model protein Im7 (refs. 20,21).

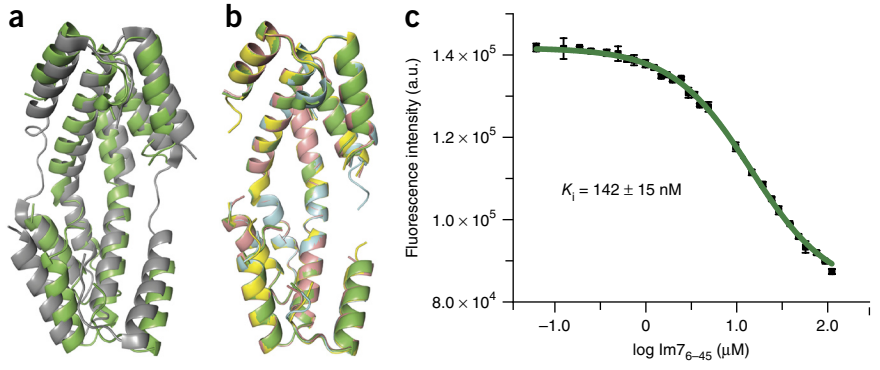

Figure 5 Spy conformation changes upon substrate binding. (**a**) Overlay of apo-Spy (PDB 3039, gray) and bound Spy (green). (**b**) Overlay of WT Spy bound to Im7$_{6–45}$ (green), H96L Spy bound to Im7 L18A L19 AL13A (blue), H96L Spy bound to WT Im7 (yellow), and WT Spy bound to casein (salmon). (**c**) Competition assay showing Im7$_{6–45}$ competing with Im7 L18A L19A L37A H40W for the same binding site on Spy (additional substrate competition assays are shown in **Supplementary Fig. 8**). Error bars, s.d. (n = 3 technical replicates). A.u., arbitrary units.

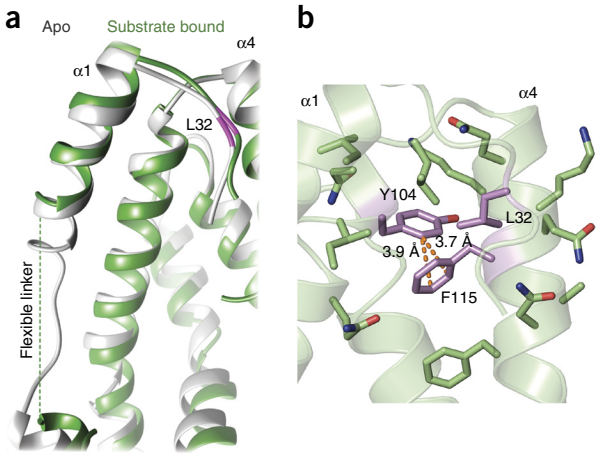

Figure 6 Flexibility of the Spy linker region and the effects of Super Spy mutants. (**a**) The Spy linker region adopts one dominant conformation in its apo state (PDB 3039, gray) but expands and adopts multiple conformations in bound states (green). (**b**) F115 and L32 tether Spy's linker region to its cradle, thereby decreasing Spy activity by limiting linker-region flexibility. The Super Spy mutants F115L, F115I, and L32P are proposed to gain activity by increasing the flexibility or size of this linker region. L32, F115, and Y104 are rendered in purple to illustrate residues that are most affected by Super Spy mutations; CH···π hydrogen bonds are depicted by orange dashes.

In the chaperone-bound ensemble, $Im7_{6–45}$ samples unfolded, partially folded, and native-like states. The ensemble provides an unprecedented description of the conformations that a substrate assumes while it explores its chaperone-associated folding landscape. This substrate–chaperone ensemble helps accomplish the longstanding goal of obtaining a detailed view of how a chaperone aids in protein folding.

We have recently shown that Im7 can fold while remaining continuously bound to Spy[22]. The high-resolution ensemble obtained here provides insight into how this occurs. The structures of our ensemble agree well with lower-resolution cross-linking data indicating that chaperone-substrate interactions primarily occur on the concave surface of Spy[24]. The ensemble suggests a model in which Spy provides an amphipathic surface that allows chaperone-bound substrate proteins to assume different conformations. This model is consistent with results from previous studies postulating that the flexible binding of chaperones allows for substrate protein folding[32]. The amphipathic concave surface of Spy probably facilitates this flexible binding and may be a crucial feature allowing Spy and potentially other chaperones[12] to bind multiple conformations of many different substrates.

In contrast to Spy's binding hotspots, $Im7_{6–45}$ displays substantially less specificity in its binding sites. Nearly all $Im7_{6–45}$ residues come in contact with Spy. Unfolded substrate conformers interact with Spy through both hydrophobic and hydrophilic interactions, whereas the binding of native-like states involves mainly hydrophilic interactions. This trend suggests that complex formation between an ATP-independent chaperone and its unfolded substrate may initially involve hydrophobic interactions, thereby effectively shielding the exposed aggregation-sensitive hydrophobic regions in the substrate. Once the substrate begins to fold within this protected environment, it progressively buries its own hydrophobic residues, and its interactions with the chaperone shift toward becoming more electrostatic. Notably, the most frequent contacts between Spy and $Im7_{6–45}$ are charge-charge interactions. The negatively charged Im7 residues Glu21, Asp32, and Asp35 reside on the surface of Im7 and form interactions with Spy's positively charged cradle in both the unfolded and native-like states. Residues Asp32 and Asp35 are close to each other in the folded state of Im7. This proximity probably causes electrostatic repulsion that destabilizes Im7's native state. Interaction with Spy's positively charged residues is likely to relieve the charge repulsion between Asp32 and Asp35, thereby promoting their compaction into a helical conformation. As intermolecular hydrophobic interactions between Spy and the substrate become progressively replaced by intramolecular interactions within the substrate, the affinity between chaperone and substrates may decrease and eventually lead to release of the folded client protein.

Recently, we have used a genetic selection system to improve the chaperone activity of Spy. This selection yielded 'Super Spy' variants that are more effective at both preventing aggregation and promoting protein folding[24]. In conjunction with our bound $Im7_{6–45}$ ensemble, these mutants allowed us to investigate structural features important to chaperone function. Previous analysis has revealed that the Super Spy variants, compared with WT Spy, bind Im7 more tightly, increase chaperone flexibility, as measured via hydrogen/deuterium exchange, or both[24]. Our ensemble revealed that two of the Super Spy mutations (H96L and Q100L) form part of the chaperone contact surface that binds $Im7_{6–45}$ (**Fig. 4a**). Moreover, our co-structure suggests that the L32P substitution, which increases Spy's flexibility[24], may unhinge the N-terminal helix and effectively expand the size of the disordered linker. This possibility is supported by the Spy–substrate structures, in which the linker region becomes more flexible than it is in the apo state[19] (**Fig. 6a**). This expansion would increase the structural plasticity of substrate binding[33]. By sampling multiple conformations, this linker region may allow diverse substrate conformations to be accommodated.

Other Super Spy mutations (F115I and F115L) cause increased flexibility but not tighter substrate binding[24]. Residue 115 does not directly contact $Im7_{6–45}$ in our READ-derived ensemble. Instead, when Spy is bound to substrate, F115 engages in close CH···π hydrogen bonds with Y104 (**Fig. 6b**). This interaction presumably decreases the mobility of the C-terminal helix. The F115I and F115L substitutions would replace these hydrogen bonds with hydrophobic interactions that have little angular dependence. As a result, the C terminus, and possibly also the flexible linker, is likely to become more flexible and thus more accommodating of different conformations of substrates. Overall, comparison of our ensemble to the Super Spy variants provides specific examples corroborating the importance of conformational flexibility in chaperone-substrate interactions[33].

Although extensive studies have been performed, exactly how complex chaperone machines help proteins fold remains controversial[7,8]. Our study indicates that the chaperone Spy uses a simple surface binding approach that allows the substrate to explore various conformations and form transiently favorable interactions while being protected from aggregation. We speculate that many other chaperones may use a similar strategy. ATP and cochaperone dependencies may have emerged later through evolution to better modulate and control chaperone action.

In addition to insights into chaperone function, this work presents a new method for determining heterogeneous structural ensembles via a hybrid methodology of X-ray crystallography and computational modeling. Heterogeneous dynamic complexes or disordered regions of single proteins, once considered to be approachable solely through NMR spectroscopy, can now be visualized through X-ray crystallography. Consequently, this technique may enable structural characterization of many important dynamic and heterogeneous biomolecular systems.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Structures and datasets in this work have been deposited in the Protein Data Bank under accession codes PDB 5INA, PDB 5IOG, PDB 5IOE and PDB 5IOA.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors would like to thank J. Smith, D. Akey, U. Jakob, D. Smith, Z. Wawrzak, and F. Stull for critical comments and suggestions. Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under contract no. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). This work was funded by an NRSA National Institutes of Health (NIH) grant GM108298 (L.S.A.), a Boehringer Ingelheim Fonds fellowship (P.K.), a National Natural Science Foundation of China (NSFC) grant 31400664 (S.Q.), the Shanghai Puijiang Program (S.Q.), NIH grant GM102829 (J.C.A.B.), NIH grant GM107233 (C.L.B.), NIH grant 1P01 GM063210 (P.V.), the Phenix Industrial Consortium and the US Department of Energy Contract No. DE-AC02-05CH11231 (P.V.) and NSF grant CHE1506273 (C.L.B.). J.C.A.B. is supported as a Howard Hughes Medical Institute Investigator.

AUTHOR CONTRIBUTIONS

The overall concept was conceived by S.H. and J.C.A.B. Experiments were designed by S.H., S.Q., J.C.A.B., R.C.T., H.v.d.B., and P.K. Experiments were performed by S.H., S.Q., P.K., R.M., and L.W. Analyses and computational modeling were designed by C.L.B., L.S., P.V.A., L.S.A., H.v.d.B., and S.H. Computational analysis was carried out by Q.X., S.H., L.S., L.S.A., P.V.A., P.K., and R.M. The manuscript was written primarily by S.H. and J.C.A.B., with assistance from L.S., L.S.A., and all other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Keskin, O., Gursoy, A., Ma, B. & Nussinov, R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.* **108**, 1225–1244 (2008).
2. Fraser, J.S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. USA* **108**, 16247–16252 (2011).
3. Kay, L.E. New views of functionally dynamic proteins by solution NMR spectroscopy. *J. Mol. Biol.* **428** 2 Pt A, 323–331 (2016).
4. Salmon, L. & Blackledge, M. Investigating protein conformational energy landscapes and atomic resolution dynamics from NMR dipolar couplings: a review. *Rep. Prog. Phys.* **78**, 126601 (2015).
5. Blackledge, M.J. *et al.* Conformational backbone dynamics of the cyclic decapeptide antamanide: application of a new multiconformational search algorithm based on NMR data. *Biochemistry* **32**, 10960–10974 (1993).
6. Guerry, P. *et al.* Mapping the population of protein conformational energy sub-states from NMR dipolar couplings. *Angew. Chem. Int. Ed. Engl.* **52**, 3181–3185 (2013).
7. Jewett, A.I. & Shea, J.E. Reconciling theories of chaperonin accelerated folding with experimental evidence. *Cell. Mol. Life Sci.* **67**, 255–276 (2010).
8. Mashaghi, A. *et al.* Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **500**, 98–101 (2013).
9. Buckle, A.M., Zahn, R. & Fersht, A.R. A structural model for GroEL-polypeptide recognition. *Proc. Natl. Acad. Sci. USA* **94**, 3571–3575 (1997).
10. Martinez-Hackert, E. & Hendrickson, W.A. Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell* **138**, 923–934 (2009).
11. Saio, T., Guan, X., Rossi, P., Economou, A. & Kalodimos, C.G. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* **344**, 1250494 (2014).
12. Joachimiak, L.A., Walzthoeni, T., Liu, C.W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159**, 1042–1055 (2014).
13. Chen, D.H. *et al.* Visualizing GroEL/ES in the act of encapsulating a folding protein. *Cell* **153**, 1354–1365 (2013).
14. Karagöz, G.E. *et al.* Hsp90-Tau complex reveals molecular basis for specificity in chaperone action. *Cell* **156**, 963–974 (2014).
15. Dekker, C. *et al.* The crystal structure of yeast CCT reveals intrinsic asymmetry of eukaryotic cytosolic chaperonins. *EMBO J.* **30**, 3078–3090 (2011).
16. Muñoz, I.G. *et al.* Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nat. Struct. Mol. Biol.* **18**, 14–19 (2011).
17. Elad, N. *et al.* Topologies of a substrate protein bound to the chaperonin GroEL. *Mol. Cell* **26**, 415–426 (2007).
18. Albert, A. *et al.* Structure of GroEL in complex with an early folding intermediate of alanine glyoxylate aminotransferase. *J. Biol. Chem.* **285**, 6371–6376 (2010).
19. Quan, S. *et al.* Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. *Nat. Struct. Mol. Biol.* **18**, 262–269 (2011).
20. Friel, C.T., Smith, D.A., Vendruscolo, M., Gsponer, J. & Radford, S.E. The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat. Struct. Mol. Biol.* **16**, 318–324 (2009).
21. Figueiredo, A.M., Whittaker, S.B., Knowling, S.E., Radford, S.E. & Moore, G.R. Conformational dynamics is more important than helical propensity for the folding of the all α-helical protein Im7. *Protein Sci.* **22**, 1722–1738 (2013).
22. Stull, F., Koldewey, P., Humes, J.R., Radford, S.E. & Bardwell, J.C. Substrate protein folds while it is bound to the ATP-independent chaperone Spy. *Nat. Struct. Mol. Biol.* **23**, 53–58 (2016).
23. Kwon, E., Kim, D.Y., Gross, C.A., Gross, J.D. & Kim, K.K. The crystal structure *Escherichia coli* Spy. *Protein Sci.* **19**, 2252–2259 (2010).
24. Quan, S. *et al.* Super Spy variants implicate flexibility in chaperone action. *eLife* **3**, e01584 (2014).
25. Creamer, L.K., Richardson, T. & Parry, D.A. Secondary structure of bovine alpha s1- and beta-casein in solution. *Arch. Biochem. Biophys.* **211**, 689–696 (1981).
26. Chak, K.F., Safo, M.K., Ku, W.Y., Hsieh, S.Y. & Yuan, H.S. The crystal structure of the immunity protein of colicin E7 suggests a possible colicin-interacting surface. *Proc. Natl. Acad. Sci. USA* **93**, 6437–6442 (1996).
27. Pashley, C.L. *et al.* Conformational properties of the unfolded state of Im7 in nondenaturing conditions. *J. Mol. Biol.* **416**, 300–318 (2012).
28. Burling, F.T., Weis, W.I., Flaherty, K.M. & Brünger, A.T. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **271**, 72–77 (1996).
29. van den Bedem, H., Dhanik, A., Latombe, J.C. & Deacon, A.M. Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1107–1117 (2009).
30. Brennan, S. & Cowan, P.L. A suite of programs for calculating X-ray absorption, reflection, and diffraction performance for a variety of materials at arbitrary wavelengths. *Rev. Sci. Instrum.* **63**, 850–853 (1992).
31. Karanicolas, J. & Brooks, C.L. III. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **11**, 2351–2361 (2002).
32. Jewett, A.I. & Shea, J.E. Folding on the chaperone: yield enhancement through loose binding. *J. Mol. Biol.* **363**, 945–957 (2006).
33. Bardwell, J.C. & Jakob, U. Conditional disorder in chaperone action. *Trends Biochem. Sci.* **37**, 517–525 (2012).

ONLINE METHODS

Construction of Spy truncation mutants and *in vitro* and *in vivo* activity measurements. To facilitate crystallization, we used $Spy_{29–124}$, a truncated Spy with the unstructured N- and C-terminal tails removed (full-length Spy is 138 amino acids). To determine whether these alterations affect Spy's chaperone activity *in vitro*, we performed *in vitro* chaperone activity assays and found that they had no significant effect; these deletions also had only a minor effect on Spy's ability to stabilize Im7 *in vivo* (**Supplementary Fig. 1**). The *in vitro* activity of $Spy_{29–124}$ was assessed with the aldolase refolding assay, as previously described[24]. Briefly, in the denaturing step, 100 μM aldolase was denatured in buffer containing 6.6 M GdmCl, 40 mM HEPES, pH 7.5, and 50 mM NaCl overnight at 22 °C (room temperature). In the refolding step, denatured aldolase was diluted to 3 μM in refolding buffer (40 mM HEPES, 150 mM NaCl, and 5 mM DTT, pH 7.5) in the presence of 6 μM WT Spy or $Spy_{29–124}$ (2:1 Spy/aldolase). As a control, an identical experiment without Spy added was also performed. The refolding temperature was 37 °C, and samples were subjected to continuous shaking. The refolding status was monitored at different time points (1 min, 4 min, 10 min, and 20 min) and tested by dilution of the refolding sample by 15-fold into the reaction buffer (0.15 mM NADH, 2 mM F1,6-DP, 1.8 U/ml GDH/TPI, 40 mM HEPES, and 150 mM NaCl, pH 7.5) at 28 °C. The absorbance was monitored for 1.5 min at 340 nm. The percentage refolding was calculated and averaged over three repeats.

To determine the *in vivo* activity of the Spy mutants, the quantity of the unstable Im7 variant L53A I54A expressed in the periplasm was compared during Spy-variant coexpression as previously described[19]. Plasmid Spy (pTrc-spy)[19] was used as the template for the construction of the variant plasmids of Spy for *in vivo* chaperone-activity measurement (**Supplementary Table 3**). To use the native signal sequence of *spy* for the periplasmic export of the Spy variants, an NheI site was first introduced between the signal sequence and the mature-protein-coding region of Spy. The vector was then digested with NheI and BamHI, purified, and ligated with the linear fragments corresponding to truncated sequences (21–130, 24–130, 27–130, 30–130, and 33–130) of Spy.

Cells containing a strain that expressed the unstable Im7 mutant IL53A I54A (pCDFTrc-ssIm7L53A I54A)[19] were transformed with expression plasmids for either WT or one of the five truncated Spy mutants and grown to mid-log phase in LB medium at 37 °C. Im7 L53A I54A and Spy expression were induced with various concentrations of IPTG for 2 h to compare the *in vivo* chaperone activity of WT Spy and the truncated Spy mutants at similar expression levels. Periplasmic fractions were prepared as previously described[34] and were separated on 16% tricine gel (Life Technologies). The bands corresponding to Spy and the C-terminal histidine-tagged Im7 were either directly visualized on Coomassie-stained gels or determined by western blotting with anti-histidine antibody (Abcam ab1187; validation provided on manufacturer's website).

Protein expression and purification. The gene encoding $Spy_{29–124}$ was amplified from plasmid pET28sumo-spy[19] with primer 1 (5′-CGC GGG ATC CTT CAA AGA CCT GAA CCT GAC CG-3′) and primer 2 (5′-CGC GCT CGA GTT ATG TCA GAC GCT TCT CAA AAT TAG C-3′) and was cloned into pET28sumo via BamHI and XhoI sites. The H96L variant was generated by Phusion site-directed mutagenesis (New England BioLabs). WT and H96L $Spy_{29–124}$ were expressed and purified as previously described[19] with the exception that Ni–HisTrap columns (GE Healthcare) were used instead of the Ni–NTA beads and mini chromatography column. ULP1 cleavage occurred after elution from the Ni–HisTrap column overnight at 4 °C during dialysis against 40 mM Tris and 300 mM NaCl, pH 8.0. After dialysis, Spy was passed over the HisTrap column to remove the cleaved SUMO tag (20 mM imidazole was left over from the dialysis). Cleavage of the SUMO tag leaves a single serine at position 28 of Spy. The flow through was then concentrated and diluted five times with 20 mM Tris, pH 8, for further purification on a HiTrap Q column. Spy has an isoelectric point of 9.5 and therefore was collected in the flow through. The flow through containing Spy was concentrated and diluted five-fold with 50 mM sodium phosphate, pH 6.5, and passed over a HiTrap SP column. Spy was then eluted with a gradient from 0 M to 1 M NaCl. Rebuffering to the final reaction buffer was accomplished by gel filtration, in which the pooled and concentrated fractions containing Spy were passed over a HiLoad 75 column in 40 mM HEPES and 100 mM NaCl, pH 7.5. Fractions containing Spy were then concentrated, frozen in liquid nitrogen, and stored at −80 °C. WT Im7, Im7 L18A L19A L37A H40W, and Im7 L18A L19A L37A were purified by the same protocol as that for Spy, but without the SP-column step. In addition to using WT Im7 and these various Im7 mutants, cocrystallization experiments extensively used $Im7_{6–45}$, a minimal Spy-binding segment that encompasses the first two helices of Im7 and contains 46% of the total Im7 sequence. It displays partial helicity when free in solution[24] (**Supplementary Fig. 3**). The 6–45 portion of Im7 (H_2N-SISDYTEAEFVQLL KEIEKENVAATDDVLD VLLEHFVKIT-OH), 4-iodophenlyalanine variants, and a peptide corresponding to a portion of bovine alpha casein $S1_{148–177}$ (Ac-ELFRQFYQLDAYPSGAWYYVPLGTQYTDAP-amide) were obtained from New England Peptide at ≥95% purity. Anomalous signals for residues E12, E14, L19, and E21 substitutions were determined with a peptide containing $Im7_{6–26}$, which was also obtained from New England Peptide at ≥95% purity.

Protein crystallization. Cocrystals of WT $Spy_{29–124}$ and Spy $H96L_{29–124}$ in complex with Im7 variants and casein were grown by vapor diffusion. 25–130 mg/ml dimeric Spy was incubated with various Im7 or casein substrates at concentrations ranging from equimolar to three-fold-excess substrate in 22–33% PEG 3000, 0.88–1.0 M imidazole, pH 8.0, and 40–310 mM zinc acetate at 20 °C. Crystals were flash frozen in liquid nitrogen with 35% PEG 3000 as a cryoprotectant. Notably, flash freezing can somewhat bias the conformations observed in crystal structures[35]. However, we chose to freeze the crystal to provide us with the maximum capability to identify and interpret the iodine anomalous signals.

Assessing the presence of substrate in crystals. Crystals were washed by sequential transfer between three to six 2-μl drops of mother liquor, with incubation in each wash solution for 2–10 s, in an effort to remove all surface-bound and precipitated substrate protein. The crystals were then dissolved for visualization by SDS–PAGE. Samples were boiled for 10 min in reducing loading buffer and then loaded onto 16% tricine gels. Wash samples and dissolved crystal samples were analyzed by Lumetein staining (Biotium) and Flamingo staining (Bio-Rad), per the manufacturer's instructions, and imaged with a FluorChem M Imager (ProteinSimple).

X-ray crystallography. Data were collected at the LS-CAT beamlines at the Advanced Photon Source at 100 K. SeMet and native Spy–$Im7_{6–45}$ crystals were collected at 12.7 keV and 9.7 keV, respectively. Spy–$casein_{148–177}$, and Spy H96L–WT Im7 crystals were also collected at 12.7 keV. Data integration and scaling were performed with iMosflm[36] and AIMLESS[37], respectively. Because molecular-replacement attempts with the previously published apo Spy structures (PDB 3O39 and 3OEO) were unsuccessful, the Spy–$Im7_{6–45}$ complex was solved with Se-SAD phasing with SeMet-Spy, and this was followed by density modification and initial model building by AutoSol in Phenix[38]. The initial model was completed and refined with the data for the native Spy–$Im7_{6–45}$ complex. The rest of the structures were built with the native Spy–$Im7_{6–45}$ structure as a molecular-replacement search model. Refinements, including TLS refinement, were performed with COOT[39] and Phenix[38]. All refined structures were validated with the MolProbity server[40] and had clashscores ranking better than the 90th percentile for all structures. Structural figures were rendered with PyMOL (http://www.pymol.org/) and UCSF Chimera[41], and movies were generated with UCSF Chimera[41]. Several partially occupied zinc atoms were observed in the crystal structure. Although some of these zinc atoms could also potentially be modeled as water molecules, doing so resulted in an increase in the R_{free}. Additionally, a section of density near His96 that is potentially partially occupied by a combination of water, the Spy linker region, and possibly zinc, was modeled as containing water molecules. Spy H96L–$Im7_{6–45}$ was used for iodine anomalous scattering experiments, owing to increased robustness and reproducibility of the crystals.

The expected anomalous scatterers in the structures were sulfur in methionine residues of Spy, zinc from the crystallization buffer, and iodine in the single pI-Phe residue of each synthetic $Im7_{6–45}$ peptide. Each iodine site is expected to be partially occupied, because $Im7_{6–45}$ had diffuse density corresponding to multiple partially occupied conformers; the zinc sites may also be partially occupied. To identify iodine, sulfur, and zinc atomic positions with anomalous scattering, data sets were collected at 6.5 keV and 14.0 keV at 100 K at the ID-D beamline at LS-CAT[30]. Anomalous difference maps for initial anomalous signal screening were calculated with phases from a molecular-replacement search with the native Spy–$Im7_{6–45}$ (with no $Im7_{6–45}$ built in) complex as the search model.

Anomalous difference maps calculated with the 14.0-keV data were used as controls to distinguish iodine from zinc atoms, because the iodine and zinc

doi:10.1038/nsmb.3237

anomalous scattering factors are comparable at 14.0 keV, whereas at 6.5 keV, f″ is approximately nine-fold greater for iodine than for zinc[30]. Anomalous differences were also collected and analyzed for a crystal of WT $Spy_{29–124}$–$Im7_{6–45}$ containing no iodine. The resulting anomalous difference map was inspected for peaks corresponding to sulfur, which were then excluded when iodine peaks were selected. In addition, peaks that overlapped with Spy in the crystal lattice were excluded from analysis.

As an initial screen for placing iodine atoms in the 6.5-keV anomalous difference maps, the median methionine sulfur signal was used as a cutoff for each individual map to control for varying data quality among crystals. All anomalous atoms were then refined in Phenix[42] with anomalous group refinement. Refined B factors of placed iodine ions were then used to estimate the positional fluctuation of the anomalous signals. This positional fluctuation was used as estimated error in the ensuing selection. A summary of all the anomalous signal heights (**Supplementary Table 1**) and anomalous difference maps (**Supplementary Data set 1**) is displayed at varying contour levels for maximum clarity of iodine and methionine peak heights. Additional raw data and maps are available upon request.

Substrate binding to Spy. The dissociation constant of $Im7_{6–45}$ was determined via a fluorescence-based competition experiment with Im7 L18A L19A L37A H40W, and its ability to compete with $casein_{148–177}$ for Spy binding was tested. Im7 L18A L19A L37A H40W was chosen for competition experiments on the basis of its tight binding (**Supplementary Fig. 8**) and substantial fluorescence change upon binding. This mutant binds to Spy more tightly than does Im7 L18A L19A L37A. 10 μM $Spy_{29–124}$ dimer was mixed with 10 μM Im7 L18A L19A L37A H40W or $casein_{148–177}$ to form a 1:1 complex, in a buffer containing 40 mM HEPES, pH 7.5, and 100 mM NaCl at 22 °C. Complex formation was monitored with a QuantaMaster 400 (Photon Technology International) on the basis of the tryptophan fluorescence of Im7 L18A L19A L37A H40W. Naturally tryptophan-free $Im7_{6–45}$ was then titrated into the complex to compete with Im7 L18A L19A L37A H40W for Spy binding. The observed fluorescence intensity at 350 nm was plotted as a function of the logarithm of the $Im7_{6–45}$ or $casein_{148–177}$ concentration. The data were fit for a one-site-binding competition model (OriginLab 9.1):

$$y = A2 + \frac{A1 - A2}{1 + 10^{x - \log x0}}$$

where $A1$ and $A2$ are the maximum and minimum asymptotes, respectively, and x is the concentration of $Im7_{6–45}$. $x0$ is the apparent K_d for $Im7_{6–45}$, on the basis of its ability to compete with Im7 L18A L19A L37A H40W. With the K_d of Im7 L18A L19A L37A H40W binding to $Spy_{29–124}$, we then calculated the K_d for $Im7_{6–45}$ binding to $Spy_{29–124}$ with the Cheng–Prusoff equation:

$$K_i = \frac{x_0}{1 + L / K_d}$$

where L is the concentration of Im7 L18A L19A L37A H40W, and K_d is the dissociation constant for Im7 L18A L19A L37A H40W binding to Spy. Because of interaction between higher-oligomeric states of $Im7_{6–45}$ and $casein_{148–177}$ (**Supplementary Fig. 8**), the competition curve was unable to be fit for $casein_{148–177}$ competing with $Im7_{6–45}$.

The stoichiometry of binding of $casein_{148–177}$ and Spy was determined on the basis of tryptophan fluorescence of the casein after addition of $Spy_{29–124}$. Increasing concentrations of $Spy_{29–124}$ were titrated against 20 μM of $casein_{148–177}$ in 40 mM HEPES, pH 7.5, 100 mM NaCl, at 22 °C. Complex formation was monitored with a QuantaMaster 400 (Photon Technology International) with the tryptophan fluorescence of $casein_{148–177}$. The observed fluorescence intensity at 339 nm was plotted as a function of the $Spy_{29–124}$ dimer concentration and fit with a quadratic equation with Origin 9.1 (OriginLab).

To determine the dissociation constant for the casein–Spy complex, increasing concentrations of $Spy_{29–124}$ were titrated to 0.25 μM of casein in 40 mM HEPES, pH 7.5, and 100 mM NaCl, at 22 °C. Complex formation was monitored with a QuantaMaster 400 (Photon Technology International) on the basis of the tryptophan fluorescence of $casein_{148–177}$. The observed fluorescence intensity at 339 nm was corrected for dilution due to the titration and then plotted as a function of the $Spy_{29–124}$ dimer concentration. The data were fit with a square hyperbola function in Origin 9.1 (OriginLab):

$$F = \frac{F_{max} \times L}{K_d + L} + C$$

where F is the recorded fluorescence signal, F_{max} is the maximum fluorescence reached upon saturation of the complex, L is the concentration of free Spy in solution, K_d is the dissociation constant, and C is a parameter for the offset. The calculated K_d is an average of three independent repetitions. The measured dissociation constants for the different substrates ranged from 0.1 to 1 μM.

Isothermal titration calorimetry (ITC). $Spy_{29–124}$ and Im7 L18A L19A L37A H40W were dialyzed overnight against 40 mM HEPES, 100 mM NaCl, pH 7.5. 165 μM Spy dimer was loaded into a syringe and titrated into a cell containing 15 μM Im7 L18A L19 AL37A H40W at 25 °C in an iTC200 (Malvern Instruments) with an injection interval of 120 s and an initial delay time of 60 s. The solution was stirred at 1,000 r.p.m., and the reference power was set to 6 μcal s^{-1} in high-feedback mode. Data analysis was conducted with a plugin for Origin 7 (OriginLab), the software provided by the manufacturer.

Analytical ultracentrifugation. Sedimentation-velocity experiments for the $Im7_{6–45}$ and the casein peptide were performed with a Beckman Proteome Lab XL-I analytical ultracentrifuge (Beckman Coulter). Both peptides were first dialyzed against 40 mM HEPES, 100 mM NaCl, pH 7.5, then diluted to a concentration of 10 μM with the dialysis buffer. Samples were loaded into cells containing standard sector-shaped two-channel Epon centerpieces with 1.2-cm path length (Beckman Coulter) and equilibrated to 22 °C for at least 1 h before sedimentation. All samples were spun at 48,000 r.p.m. in a Beckman AN-50 Ti rotor, and the sedimentation of the protein was monitored continuously with interference optics, because the $Im7_{6–45}$ does not absorb strongly at 280 nm. Data analysis was conducted with SEDFIT (version 14.1)[43], with the continuous c(s) distribution model. The confidence level for the maximum entropy (ME) regularization was set to 0.95. Buffer density and viscosity were calculated with SEDNTERP (http://sednterp.unh.edu/).

Additional methodology. Computational methods, including simulations of Spy-substrate interactions, binning of the residual Im7 electron density, ensemble selection, validation tests, and contact-map generation are described in **Supplementary Note 1**. Additional models and source code are available upon request.

34. Quan, S., Hiniker, A., Collet, J.F. & Bardwell, J.C. Isolation of bacteria envelope proteins. *Methods Mol. Biol.* **966**, 359–366 (2013).
35. Fischer, M., Shoichet, B.K. & Fraser, J.S. One crystal, two temperatures: cryocooling penalties alter ligand binding to transient protein sites. *ChemBioChem* **16**, 1560–1564 (2015).
36. Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R. & Leslie, A.G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
37. Winn, M.D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
38. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
39. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
40. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
41. Pettersen, E.F. *et al.* UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
42. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
43. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).