Supplemental Note
Supplemental Results

Ensemble size determination and cross-validation tests

To determine the most appropriate $Im7_{6-45}$ ensemble size, we performed the selection with sizes ranging from one to twenty. Overall, increasing the ensemble size improved the fit. The improvement declined, however, at larger ensemble sizes, particularly for the iodine data. With six members in the ensemble, the vast majority of the iodine signals were accounted for.

We assessed the validity of the ensemble and ensured that it has predictive power by using cross-validation tests. Cross validations are commonly used to test whether a model adequately reproduces the experimental data while avoiding over-interpretation[1,2]. We randomly excluded 10% of the residual electron density and iodine signals from the selection process for use as a test dataset. We then averaged over 10 independent cross-validations to generate a 10-fold cross-validation test, similar to cross-validation test strategies in NMR spectroscopy[3]. This test demonstrated that selected ensembles of at least six members back-predicts the test dataset better than ensembles randomly chosen from the MD pool (Online Methods). Passing this cross-validation test demonstrates that the obtained ensemble has predictive capability for data held back from the fit and provides a meaningful description of the substrate conformational ensemble.

Importantly, the cross-validation also examined if the underlying data points are incoherent[4,5] by testing whether iodine signals could be predicted using data arising from other pI-Phe substitutions. If the different pI-Phe substitutions in $Im7_{6-45}$ significantly impacted the binding mode or conformations of $Im7_{6-45}$, then the cross-validation test would have failed. The success of our cross-validation test, however, suggests that our labeling strategy did not induce major distortion in the binding mode or conformations of $Im7_{6-45}$. Minor differences might be introduced by the pI-Phe substitutions, but these changes remain below the resolution of the READ approach.

To further investigate the capability of our ensembles to recapitulate the electron density of the Spy:$Im7_{6-45}$ complex, we subjected the ensembles to crystallographic refinement. For ensemble sizes up to six, addition of the ensemble to the structural model of Spy produced a small improvement in R_{Free}, decreasing it from 24% to 23%. This result demonstrated that including the six-member $Im7_{6-45}$ ensemble enables a better fit to the experimental electron density. At ensemble sizes larger than six members, this decrease in R_{Free} was no longer evident. The χ^2 cost function used in the selection prioritizes fitting areas of high density over low

---

[1]Clore, G.M. & Garrett, D.S. R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. *Journal of the American Chemical Society* **121**, 9008-9012 (1999).

[2]Kleywegt, G.J. & Brunger, A.T. Checking your imagination: applications of the free R value. *Structure* **4**, 897-904 (1996).

[3]Brunger, A.T., Clore, G.M., Gronenborn, A.M., Saffrich, R. & Nilges, M. Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. *Science* **261**, 328-31 (1993).

[4]ibid.

[5]Zhang, Q., Stelzer, A.C., Fisher, C.K. & Al-Hashimi, H.M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263-7 (2007).

density. Thus, we pruned our model to only contain those atoms that were either directly selected in the READ procedure or that could be justified based on agreement with the residual electron density. Our final ensemble contains one out of every 2.5 residues in $Im7_{6-45}$.

This six-membered ensemble was the largest ensemble that could simultaneously decrease the R_{Free} and pass the 10-fold cross-validation test. We therefore used this ensemble as a model of the conformations that the substrate $Im7_{6-45}$ adopts while bound to its chaperone (**Fig. 3** and **Supplementary Movie 1**).

Additional validations of the selection and ensemble accuracy

To test whether our selection strategy was indeed successful in providing useful models to fit the READ data, we investigated its ability to reproduce important structural features of chaperone-bound $Im7_{6-45}$ using simulated data. For this set of validations, we performed our selections with noise-corrupted simulated data back-calculated from three pre-chosen target ensembles (Online Methods and **Supplementary Fig. 5a**). These target ensembles varied in the level of substrate structure and location with respect to Spy. Our goal was to test whether running selections with these back-calculated target data would recapitulate the target ensembles. In each case, we found that the selected ensembles were similar to the target ensembles (**Supplementary Fig. 5b**). These results confirmed that our selection indeed reaches convergence and is able to accurately determine ensembles based on experimental data.

Next, we tested if the final ensemble accurately reflects the READ data as opposed to being overly influenced by the initial conformational pool. To distinguish between these two possibilities, we performed a test in which we significantly altered the MD pool used for the selection to make it energetically unrealistic (**Supplementary Fig. 6**). To do this, we first extracted ~1000 conformers from the energetically realistic pool that best recapitulated the conformational space observed in our READ selections. This extraction was accomplished by performing READ selections with ensemble sizes ranging from 1 to 20 conformers, and pooling these selected conformers. We then created an alternate energetically unrealistic MD pool that encompassed ~9,000 conformers generated from an MD simulation in which all intramolecular interactions within the $Im7_{6-45}$ substrate were turned off. The conformational features of this alternate pool were notably different and in poorer agreement with solution data reporting on $Im7_{6-45}$ than those of the original MD pool. We combined the alternate pool with the structures from our READ selections and thereby generated a mixed ensemble of ~10,000 members for a new round of selection (**Supplementary Fig. 6**). We reasoned that if the experimental data indeed guided the selection, the resulting ensembles should be enriched in structures that had previously been selected from the energetically realistic pool. In contrast, if the experimental data were not driving the selection, then the new ensembles would be heavily polluted by the alternate pool. Performing the selection using this mixed pool resulted in ensembles that were highly enriched in structures from the original pool of structures derived through READ (**Supplementary Fig. 6**). On average, two-thirds of the newly selected structures came from the original energetically realistic, READ-selected conformers, despite these conformers only forming one tenth of the combined pool. This enrichment indicated that the selection is significantly guided by the READ data and does not simply reflect the mostly populated members of the pool.

Finally, to investigate the precision of the observed Spy-$Im7_{6-45}$ interactions from the READ ensemble, we performed a bootstrapping error estimation. This classical resampling procedure (**Supplemental Methods**) was used to estimate uncertainties in the contact map of the resulting ensemble. The small standard error of contact maps from 200 bootstrapped selections

(**Supplementary Fig. 7**) suggested a high degree of precision in our selection. Moreover, our finding that selected ensembles were enriched in regions of low population in the initial MD pool demonstrated that the READ data is indeed guiding the selection. Combined, these validation tests confirmed that the selection procedure and selected six-member ensemble recapitulate the experimental data.

Supplemental Methods

Simulations of Spy-substrate interaction.

Force field. We performed MD simulations of Spy-substrate binding in CHARMM[6] by building on the Karanicolas-Brooks Gō-like model[7]. In this model, each residue was represented by a single bead located at the Cα position and with the mass of the underlying amino acid. An additive potential describes the bonded and non-bonded interactions between beads:

$$V = V_{bond} + V_{angle} + V_{dihedral} + V_{non-bond} \quad (1)$$

"Virtual" bonds and angles were specified by harmonic potentials with reference values determined by the Cα coordinates in the experimental structure. The dihedral energy term reflects the backbone dihedral angle probability distributions in the Protein Data Bank (PDB) for the 400 possible amino acid pairs and is thus independent of the experimental dihedral angles. To account for α-helical bias in this model arising from the high incidence of helical residues in the PDB, we added a correction to the dihedral potential[8]. For the set of intramolecular non-bonded interactions, residue pairs separated in sequence by three or more bonds and located in close proximity in the experimental structure ("native contacts") interact through the following potential, which leads to increased cooperativity in protein folding compared to a standard Lennard-Jones 12-6 interaction[9]:

$$V_{ij} = \varepsilon_{ij}\left[13\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{12} - 18\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{10} + 4\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{6}\right] \quad (2)$$

where r_{ij} is the distance between residues *i* and *j*, r_{ij}^{min} is the distance between the two residues at which V_{ij} is minimum, and $-\varepsilon_{ij}$ is the strength of the interaction at r_{ij}^{min}. For a given native contact, r_{ij}^{min} was set to the Cα-Cα distance between the residue pair in the experimental model,

---

[6]Brooks, B.R. et al. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **30**, 1545-1614 (2009).

[7]Karanicolas, J. & Brooks, C.L., III. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Science* **11**, 2351-2361 (2002).

[8]De Sancho, D. & Best, R.B. Modulation of an IDP binding mechanism and rates by helix propensity and non-native interactions: association of HIF1alpha with CBP. *Molecular Biosystems* **8**, 256-67 (2012).

[9]Karanicolas, J. & Brooks, C.L. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *Journal of Molecular Biology* **334**, 309-325 (2003).

and ε_{ij} was scaled in proportion to the Miyazawa-Jernigan statistical contact energies[109]. A slight repulsive interaction proportional to the first term in equation (2) is applied to residue pairs not in contact.

To describe interactions between Spy and substrate, we add a generic inter-protein potential on top of the Gō-like model. This intermolecular potential takes the form of a Lennard-Jones term:

$$V_{ij} = \varepsilon_{ij}\left[\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{6}\right] \quad (3)$$

where r_{ij} and ε_{ij} are defined as in equation (2). $r_{ij}^{min}$was obtained from the mean Cα-Cα distance for residue pairs that form intermolecular contacts in the PDB[11].

System setup. An initial model of $Im7_{6\text{-}45}$ was obtained from residues 6-45 of the full-length Im7 crystal structure (PDB ID: 1CEI)[12]. We renormalized the strength of intramolecular $Im7_{6\text{-}45}$ interactions such that the per-residue helical propensities during simulation matched those obtained from NMR chemical shifts (**Supplementary Fig. 4**). Initial coordinates for Spy were taken from the co-crystal structure in the current study, while rebuilding the disordered linker and terminal residues by choosing the lowest energy configurations generated by MODELLER 9.11[1312]. To account for the crystal environment in generating Spy:$Im7_{6\text{-}45}$ peptide poses for the subsequent selection procedure, we performed simulations of two different "pseudo crystals" that together encompass the possible contacts of $Im7_{6\text{-}45}$ with Spy within the crystal (**Supplementary Fig. 4** and **Fig. 2**).

Dynamics. Six independent pseudo-crystal MD simulations (three in each crystal environment) were seeded from different unfolded $Im7_{6\text{-}45}$ conformations randomly oriented about the Spy molecules chosen from the $Im7_{6\text{-}45}$ alone simulations (see above). In each simulation, Langevin dynamics were propagated at 300 K with a friction coefficient of 0.1 ps^{-1} for 10^8 steps with a 15 fs time step and different random seeds for assigning velocities. To preserve the pseudo-crystal arrangement, we maintained the non-loop and non-terminal atoms of the Spy molecules about their initial coordinates using harmonic restraints with a 10 kcal mol^{-1} $Å^{-2}$ force constant. These simulations used the same NMR-parameterized intramolecular $Im7_{6\text{-}45}$ interaction strengths as described in step *II* (above). In total, we obtained a diverse pool of 600,000 configurations for analysis.

Selection Pool Preparation. For each configuration from the pseudo-crystal MD simulations, we determined the Spy-$Im7_{6\text{-}45}$ pair with the greatest number of intermolecular residue-residue contacts. We then analyzed these binary complexes for steric clashes within the full crystal

---

[10]Miyazawa, S. & Jernigan, R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. Ibid.**256**, 623-644 (1996).

[11]Cheng, S., Zhang, Y. & Brooks, C.L., III. PCalign: a method to quantify physicochemical similarity of protein-protein interfaces. *BMC Bioinformatics* **16**, 33 (2015).

[12]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. *Proceedings of the National Academy of Sciences* **93**, 6437-6442 (1996).

[13]Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234**, 779-815 (1993).

environment. We selected a total of ~10,000 "non-clashing" binding poses at an even extraction rate using a Cα-Cα cutoff of 4.0 Å, 0.5 Å less than the shortest r_{ij}^{min} in our course-grained force field. For each pose, we "recombined" the Spy loop and termini and $Im7_{6-45}$ Cα coordinates with the Spy co-crystal structure and refined XYZ coordinates, B-factors, and occupancies using Phenix, with harmonic restraints set to limit the motion of $Im7_{6-45}$[14]. This refinement of each individual conformer was necessary to establish $2mF_o$-DF_c map values for each conformer using refined B-factors that would be directly used in the selection. By performing this refinement for each conformer individually, the bulk-solvent scaling would be as accurate as possible at this stage. The sets of Spy:$Im7_{6-45}$ complexes and resulting electron density maps served as the starting point for selection. The final binding poses from the MD pool were checked against both the residual electron density and iodine anomalous signals to ensure that the pool sufficiently sampled regions proximal to the data. Small changes in the Spy model (e.g. changes that caused R_{Free} to increase up to 26%) were found to not make significant changes in the residual electron density.

Binning the residual Im7 electron density. As the selection algorithm (see Ensemble selection, below) requires many cycled repeating comparisons between the residual electron density and test ensembles, the following procedure was applied to provide fast model to map comparisons. The output $2mF_o$-DF_c maps from refining the MD simulation were averaged to create a $2mF_o$-DF_c map averaged over the entire simulation. Then, we extracted map values from the averaged $2mF_o$-DF_c map at the position of each Cα of each frame of the ~10,0000 refined MD configurations. A 4 Å 3D-grid was then used to average those values within bins for the selection. All bins with a density lower than a defined bottom threshold ranging from 0.025 e/Å^3 to 0.050 e/Å^3 were discarded, with 0.050 e/Å^3 providing the best results as determined by hold-out cross-validation tests (see below). All values above 0.2 e/Å^3 were truncated to 0.2 e/Å^3 to avoid any contaminating electron density arising from well-ordered solvent or Spy that would bias the selection. Finally, each of the 4 Å bins sampled by less than 100 Cα over the course of the MD trajectory were discarded due to insufficient sampling that could lead to incomplete electron density averaging. This procedure led to an "experimental" histogram of 90 points, a computationally feasible number for incorporation into the selection procedure described below. Then, we applied these 90 bins to each of the individual MD frames, leading to "individual" histograms for each of the separate test conformers.

Ensemble selection. A genetic algorithm was used to select a sub-ensemble of N conformers from the MD pool of ~10,000 configurations[15] (**Fig. 2**). The algorithm is initialized by generating a set of X=100 ensembles of N conformers generated randomly. Then, 50,000 steps of evolution are iteratively performed in which 4X new ensembles are generated using internal and external mutations, reproduction, or by generating new random ensembles. Mutations consist of modifying a given ensemble by substituting one of its conformers by another conformer, either

---

[14]Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D-Biological Crystallography* **D68**, 352-367 (2012).

[15]Nodet, G. et al. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *Journal of the American Chemical Society* **131**, 17908-17918 (2009).

from the conformations not yet selected within the ensembles (external mutations) or from the conformations already selected (internal mutations). Reproduction consists in paring existing ensembles and mixing the conformers of these existing ensembles. The obtained 5X ensembles are pooled in T=1–100 different tournaments. The number of tournaments starts high (100) to provide a low selection pressure and favor conformational diversity. Reducing the number of tournaments through the iterations progressively increases the selection pressure as the χ^2 decreases to focus the selection towards an optimal ensemble. The ensembles providing the best data reproduction in each tournament are kept to obtain a new set of X ensembles. The data reproduction, at each step, is assessed by using a fitness function χ^2_{tot} reporting on both the iodine anomalous signal and the binned residual density of $Im7_{6\text{-}45}$ and the flexible linker regions of Spy. No helicity constraints were used in the selection procedure.

For the residual electron density selection, the reproduction of the "experimental" histogram (see above) was assessed using a classical χ^2 function:

$$\chi^2_{elec} = \sum_i \left(\frac{\rho_i^{calc} - \rho_i^{exp}}{\delta^{elec}} \right)^2$$

where *i* runs over the different bins of the experimental histogram, ρ_i^{exp} is the value of the "experimental" histogram in that bin, and ρ_i^{calc} is the averaged value of the "individual" histograms for the conformations in the test ensemble (see above) and δ^{elec}, a constant weight fixed to 0.05 to estimate of the "noise" in the residual electron density map. The choice of χ^2 biases the selection towards fitting areas of high electron density preferentially over areas of low electron density. As such, analysis of the χ^2 selections showed that bins in the top 50% as ranked by electron density level were twice as likely to be fit by protein than bins in the lower 50% as ranked by electron density level. For comparison, a correlation coefficient selection was also performed, in which the χ^2 method was tested against the scale-independent correlation coefficient. The χ^2 selection performed better in the hold-out cross-validation tests (see below) and was therefore used in further selections. Minimizing χ^2 has the advantage of preferentially fitting regions of high electron density while de-prioritizing areas of low electron density.

For the iodine anomalous signal, a χ^2 function modified for better handling of outliers was used[16,17]:

$$\chi^2_{iodo} = c^2 \sum_i 1 - exp\left[-\left(\frac{D_i^{calc} - D_i^{exp}}{c\delta^{iodo}} \right) \right]^2$$

where *i* runs over the different iodine signal, D_i^{calc} is the closest distance to D_i^{exp} = 6.5 Å between the iodine signal and the corresponding Cα of any of the conformers in the ensemble, δ^{iodo} is a weight estimated using the atomic fluctuation of the particular iodine as defined above, and *c* is fixed at its canonical value 2.9846[14]. This target function was selected to handle the possibility of iodine substitutions altering the conformational sampling of $Im7_{6\text{-}45}$.

---

[16]Bouvignies, G., Meier, S., Grzesiek, S. & Blackledge, M. Ultrahigh-resolution backbone structure of perdeuterated protein GB1 using residual dipolar couplings from two alignment media. *Angewandte Chemie-International Edition* **45**, 8166-8169 (2006).

[17]Press, W.H. *Numerical recipes : the art of scientific computing*, xxi, 1235 p. (Cambridge University Press, Cambridge, UK ; New York, 2007).

The total fitness function, χ^2_{tot}, used to select the best ensembles in each tournament is then computed as:

$$\chi^2_{tot} = \chi^2_{elec} + \lambda\chi^2_{iodo}$$

where $\lambda = 2.9$ is a scaling factor introduced to ensure that the selection receives equal guidance from residual electron density and anomalous signals. The value of λ was computed as the ratio of the density and the iodine data points.

Validation tests.

10-Fold cross-validation tests. To test the significance of the selection, we removed 10% of both electron density and iodine anomalous signal positions from the selection. The selection was then performed using the 90% remaining, and the data reproduction was assessed by computing the final χ^2_{tot} at the end of the selection using only the 10% of data unused in the fit. The procedure was repeated using 10 different sets of randomly removed data, and averaged (**Fig. 1c,d**). Alternatively, each iodine probe was removed one after the other and the active selection was carried out using the remaining iodine probes and the electron density. Again, the data reproduction was assessed using only the iodine signals that were not included in the selection procedure. Data reproduction using this alternative iodine removal procedure was roughly equivalent to the former case (data not shown).

Boostrapping. Boostrapping was performed by resampling the experimental data distribution and randomly replacing a fraction of the measured points with repeat data. In the procedure, 200 new datasets were generated in which the canonical value of 37% of the data points[18] (both iodine anomalous signals and residual electron density bins) were removed and replaced with randomly chosen repeat data. For each of these datasets, the selection procedure was as described above (Ensemble selection). The 200 final ensembles were used to estimate the precision of the Spy:Im7$_{6\text{-}45}$ contact map.

Simulated data. Three separate six-member target ensembles were chosen from the MD pool to provide target ensembles with varying levels of structure and placement relative to Spy. Simulated iodine signals and electron density were back calculated from these structures and noise corrupted. For iodine back-prediction, the same number of anomalous signals was used as in experiment. Simulated iodine anomalous signals were inserted by choosing random points 6.5 Å away from the Cα of the pI-Phe-substituted residue. The position of this simulated iodine signal was then noise corrupted using a Gaussian function, with a standard deviation of 0.5 Å. To generate a target electron density map, simulated $2mF_o$-DF_c maps were back-calculated from the target ensembles using Phenix. Map values were extracted for each Cα of these ensembles. To noise-corrupt the map, we took the positional coordinates of each Cα in the MD pool and replaced the $2mF_o$-DF_c map values for each Cα in the MD pool with noise. Gaussian electron density noise for the map values was set to an average amplitude of 0.025 e/Å^3 and standard deviation of 0.100 e/Å^3. The map values and Cα positions from the target data were then combined with the noise map values and positions. Binning the electron density then proceeded as described above (see Binning the residual Im7 electron density**)**. This noise-corrupted data was then used to perform READ selections. The difference between the target and initial ensembles were analyzed by pairwise RMSD. In this analysis, the pairwise RMSDs between the

[18]Press, W.H. *Numerical recipes : the art of scientific computing*, xxi, 1235 p. (Cambridge University Press, Cambridge, UK ; New York, 2007).

target and selected ensembles were rank-ordered by lowest to highest RMSD. Then, the pairs between the target and selected ensembles were chosen to obtain a 1:1 mapping of the members from the selected pool with their lowest-RMSD analogs in the target ensemble. Then, the average RMSD of these pairs was calculated.

Pool corruption. READ selections of ensemble of sizes from 1 to 20 were pooled together and supplemented with structures in their direct vicinity to provide a ~1000 conformer pool, sampling the conformational space in agreement with experimental data. A second pool of ~9,000 members were generated by performing analogous pseudo-crystalline simulations with all intra-molecular contacts within $Im7_{6\text{-}45}$ turned off. Thus, the substrate can still interact with the Spy lattice, but will adopt conformations that are unstructured.

These two pools were combined into a single pool that was used to carry out selections as described above (**Ensemble Selection**). The similarity between the ensemble obtained using this pool and the ensemble obtained from the initial MD pool was estimated by counting the number of members from the initial MD pool in the final ensembles.

R_{Free}. All $Im7_{6\text{-}45}$ atoms that were actively selected were included in the final six-member model. The remaining atoms from the selected conformers were included based on refinement in Phenix[13] in which each of the remaining atoms was added back into the structure one at a time.

Contact map generation. For the selected ensemble, we computed intermolecular contact maps between Spy and $Im7_{6\text{-}45}$. We assigned a contact to a given residue pair if the Cα-Cα distance between the two residues was less than or equal to $\lambda \cdot r_{ij}^{min}$ from the intermolecular potential presented in equation (3), where $\lambda = 1.2$. For each intermolecular residue pair, we reported the contact probability averaged over all possible peptide and Spy loop conformations. To project the contact maps onto the structures of Spy and $Im7_{6\text{-}45}$ on the same scale, the contact frequency for each residue pair was averaged over all residues.