Published in final edited form as:

J Mol Biol. 2009 April 3; 387(3): 639–652. doi:10.1016/j.jmb.2009.01.046.

Structural and motional contributions of the *Bacillus subtilis* ClpC N-domain in adaptor protein interactions

Douglas J. Kojetin^{1,*}, Patrick D. McLaughlin^{2,*}, Richele J. Thompson², David Dubnau³, Peter Prepiak³, Mark Rance^{1,†}, and John Cavanagh^{2,†}

¹Department of Molecular Genetics, Biochemistry and Microbiology, College of Medicine, University of Cincinnati, Cincinnati, OH 45267, USA

²Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA

³Public Health Research Institute, 225 Warren Street, Newark, NJ 07103, USA

Summary

The AAA+ superfamily protein ClpC is a key regulator of cell development in *Bacillus subtilis*. As part of a large oligomeric complex, ClpC controls an array of cellular processes by recognizing, unfolding, and providing misfolded and aggregated proteins as substrates for the ClpP peptidase. ClpC is unique compared to other HSP100/Clp proteins, as it requires an adaptor protein for all fundamental activities. The NMR solution structure of the N-terminal repeat domain of ClpC (N-ClpCR) comprises two structural repeats of a four-helix motif. NMR experiments used to map the MecA adaptor protein interaction surface of N-ClpCR reveal that regions involved in the interaction possess conformational flexibility, as well as conformational exchange on the μ s-ms time-scale. The electrostatic surface of N-ClpCR differs substantially compared to the N-domain of *Escherichia coli* ClpA and ClpB, suggesting that the electrostatic surface characteristics of HSP100/Clp N-domains may play a role in adaptor protein and substrate interaction specificity, and perhaps contribute to the unique adaptor protein requirement of ClpC.

Keywords

Competence; adaptor protein proteasome interactions; HSP100/Clp N-domain; NMR dynamics

Introduction

AAA+ (ATPases associated with a variety of cellular activities) superfamily proteins regulate a wide range of cellular activities in all kingdoms of life.^{1, 2} In prokaryotes, the HSP100/ Clp

†Correspondence. JC, email: E-mail: john_cavanagh@ncsu.edu, phone: (919) 513-4349, fax: (919) 515-2047; MR, email: E-mail: mark.rance@uc.edu, phone: (513) 558-0066, fax: (513) 558-8474.

*These authors contributed equally to this work

Supplemental Data: This manuscript contains additional methodological details, three supplemental figures, and three supplemental tables.

Accession Numbers: The N-ClpCR NMR structure ensemble and restraints have been deposited in the PDB under accession code 2k77, and NMR parameters have been deposited to the BMRB under accession code 15910.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

family of AAA+ proteins functions in ATP-dependent folding, assembly, chaperoning, unfolding and degradation of misfolded, aggregated and targeted proteins.³ HSP100 proteins have a conserved ATPase (AAA) core domain with either one (class II) or two (class I) copies. Other notable domains include a N-terminal domain (N-domain) and a Linker domain (also known as the I-domain or M-domain), although these domains are not conserved to the same degree as the AAA domain. The N-domain is thought to provide a means to regulate the specificity of and enlarge the substrate pool available to HSP100 chaperone or protease complexes.⁴ These roles can be assisted through the binding of an adaptor protein. Adaptor proteins bind to the N-domain, modulate the target specificity of the HSP100 complex to a particular substrate of interest, and may also regulate the activity of the complex. A recent study revealed that overexpression of the HSP100 family protein, Hsp104, cures cells of amyloid prion propagation [PSI⁺].⁵ Interestingly, overexpression of a deletion mutant Δ N-Hsp104 fails to cure [PSI⁺], suggesting an important role for the N-domain in prion disaggregation. However, some HSP100 proteins do not require the N-domain to perform fundamental activities, leading to the notion that the N-domain may be involved in substrate unfolding or substrate discrimination by modulating access of the substrate to the interior of the AAA domain.

ClpA, ClpB, ClpC, and ClpX are HSP100 proteins that function in general protein quality control, such as disaggregation, refolding, and proteolysis.^{4, 6} Most adaptor proteins, including those that target ClpA (ClpS) and ClpX (SspB, RssB, and UmuD), modify the substrate specificity of the HSP100 protein but are not functionally required. No adaptor proteins have been identified thus far for ClpB, nor is ClpB chaperone activity dependent on the N-domain.⁷ In contrast, *Bacillus subtilis* ClpC is unique in that adaptor proteins are essential for substrate binding and all general chaperone and degradation activities of the ClpCP protease, including ClpC hexamerization.^{8, 9} The two-domain adaptor protein MecA is specifically required for ClpC activation and degradation of the competence development transcription factor ComK by ClpCP.^{10, 11} The C-terminus of MecA binds to the N-domain and Linker domain of ClpC, whereas the MecA N-terminus binds ComK (substrate) and targets it to the ClpCP for degradation.^{8, 12} Other adaptor proteins bind ClpC, including McsAB and YpbH. McsAB targets the transcriptional repressor CtsR for degradation by ClpCP.¹³ The exact role of YpbH, on the other hand, remains elusive.¹⁴ Furthermore, little is known about the molecular mechanisms underlying the unique adaptor protein functional requirement of ClpC; a feature that is quite remarkable and distinct among HSP100 family proteins.

A recent study of the ClpS adaptor protein revealed that its N-terminal extension, which binds to the N-domain of ClpA, is essential for delivering some substrates to the ClpAP protease.¹⁵ One may envision that designer adaptor proteins may be engineered to target specific proteins of interest, such as those involved in a particular disease, for chaperoning or degradation. In this sense, the ClpC/MecA system provides a unique opportunity to study a system with a strict adaptor protein requirement. Furthermore, the distinct functions of the N- and C-terminal domains of MecA – the former in binding a substrate, analogous to the ClpS N-terminal extension; the latter in binding the ClpCP protease – suggest that two-domain MecA-like proteins could be engineered to bind and target specific proteins for degradation. As a first step towards understanding the unique adaptor protein requirement of *B. subtilis* ClpC, we report the NMR solution structure, backbone dynamics, and a mapping of the MecA adaptor protein interaction surface of the N-terminal repeat domain of *B. subtilis* ClpC (N-ClpCR). We compare our data to other HSP100 N-domain studies to provide insight into the structural and motional features involved in the interaction with adaptor proteins.

Results and Discussion

NMR Solution Structure of N-ClpCR

N-ClpCR comprises amino acid residues 1-145 of the 810 residue wild-type *B. subtilis* ClpC. The NMR solution structure of N-ClpCR (Figure 1a and 1b) was determined from a total of 6910 nuclear Overhauser effect (NOE)-based distances, 66 hydrogen bonds, 208 TALOS-derived dihedral angles, and 78 $^1\text{D}_{\text{NH}}$ residual dipolar couplings (RDCs). Thirty structures were determined by FINDFAM to be sufficient to represent the conformational space consistent with the experimental data and were used to represent the NMR structure ensemble. All experimental restraints were well satisfied and the high quality of the structure ensemble is evident from the NMR refinement and structure statistics (Table 1). The correlation between back-calculated and experimentally determined $^1\text{D}_{\text{NH}}$ RDCs is 0.952 ± 0.010 (Q factor, 0.302 ± 0.028) and 0.962 ± 0.008 (Q factor, 0.272 ± 0.027) before and after use during AMBER structural refinement, respectively. The percentage of residues reported by MOLPROBITY in favored and allowed Ramachandran regions are $99.54 \pm 0.05\%$. The thirty lowest-energy structures superimpose with a backbone and heavy-atom maximum likelihood (ML) RMSD of $0.475 \pm 0.098 \text{ \AA}$ and $1.197 \pm 0.113 \text{ \AA}$, respectively.

The overall fold of N-ClpCR contains two structural repeats of a four-helix motif, comprising residues E8-I68 and P82-L141. The DSSP -calculated secondary structure reveals eight α -helices (Figure 2): $\alpha 1$ {E8-L24}, $\alpha 2$ {T31-E41}, $\alpha 3$ {I45-A52}, $\alpha 4$ {S57-I68}, $\alpha 5$ {P82-K97}, $\alpha 6$ {T105-E115}, $\alpha 7$ {V119-L127}, and $\alpha 8$ {L131-L141}. Short parallel β -strand interactions are present between {F6/V103} and {N28-I29/H79-Y80}. A flexible loop (Figure 1c) between $\alpha 4$ and $\alpha 5$ {R70-I78} connects the two four-helix motifs and is characterized by a lower than average number of NOEs. Two stretches of nine residues within the repeat regions, G30-L38 ($\alpha 2$) and G104-L112 ($\alpha 6$), are 100% identical (Figure 2). Hydrophobic contacts between $\alpha 2$ {I34, L35, and L38} and $\alpha 6$ {I108, L109, and L112} contribute to the packing of the core and position the repeat regions within close proximity, giving rise to a pseudo two-fold axis of symmetry (Figure 1d). The identical regions comprising $\alpha 2$ and $\alpha 6$ align antiparallel with a disposition of $-149.7 \pm 0.9^\circ$.

Comparison of N-ClpCR to Other N-domain Structures

The solution structure of N-ClpCR is similar to other HSP100 N-domain structures. A query to the DALI database reveals significant hits to N-domains of *Escherichia coli* ClpA and ClpB (Z-scores ~ 14 -19), herein referred to as N-ClpA and N-ClpB, respectively. The structural fold of N-ClpCR corresponds to the SCOP fold “Double Clp-N motif [81922]” (multihelical; array) and superfamily “Double Clp-N motif [81923]” (duplication; contains two structural repeats of 4-helical motif). The four-helical repeat motif present in N-ClpCR corresponds to the PFAM PF02861 domain, which is noted to be “found in one or two copies at the amino terminus of ClpA and ClpB proteins from bacteria and eukaryotes”.

Two repeat sequence regions are present within N-ClpCR (Figure 2) spanning residues E19-Q51 and D93-N125 (69% identical and 100% similar). Residues flanking the identical sequence repeats contain a combination of identical and similar residues. Identical sequence repeats are not present in N-ClpA or N-ClpB; however, the four-helical structural repeats are conserved in all three proteins (Figure 3a, 3c, and 3c). ML superposition of N-ClpCR to N-ClpA and N-ClpB (Figure 4a and 4b) results in a Ca RMSD of $2.29 \pm 0.38 \text{ \AA}$. The majority of the structural displacement occurs in loop regions and, in particular, regions proximal to the flexible loop region (residues R70-I78 in N-ClpCR), as well as N-ClpA and N-ClpB residues with large B-factor values in previous crystallographic studies.

The loop region between the four-helical fold repeats in the *E. coli* N-ClpB crystal structure is not observable [PDB code: 1khy].¹⁶ Although this loop is present in the *E. coli* N-ClpA,^{17, 18} the region is characterized by large B-factor values, indicating considerable disorder in the crystal lattice. The loop region of N-ClpA contains a number of acidic and polar amino acids (Figure 2) and is noted to become more ordered upon binding the adaptor protein ClpS.¹⁸ The corresponding loop in N-ClpB is glycine rich with other negatively charged, polar, and hydrophobic amino acids. In N-ClpCR, this loop is largely polar with a few charged and hydrophobic amino acids. The variable molecular composition of this N-domain loop region may impart differences in the binding of adaptor proteins and substrates among HSP100 proteins. Previous structural studies described the interaction between N-ClpA and the ClpS adaptor protein.¹⁸⁻²⁰ Two Phe residues in N-ClpA hypothesized to provide adaptor protein specificity between ClpA homologs and ClpB/Hsp104 homologs are not conserved in N-ClpB or N-ClpCR (Figure 2, purple boxes). However, four out of eight residues that participate in N-ClpA adaptor protein “interface A” interactions are conserved in N-ClpCR (Figure 2, blue boxes). A second N-ClpA region that contributes to the ClpS interaction, “interface C”, includes a number of hydrophobic residues and two positively charged residues (Figure 2, green and red boxes, respectively). The hydrophobic residues are somewhat conserved in N-ClpB and N-ClpC, however the adjacent ClpA charged residues are hydrophobic or polar residues in N-ClpB and N-ClpCR (Figure 2, green boxes). Additional N-ClpA residues in “interface C” are not well conserved in N-ClpB and N-ClpCR (Figure 2, yellow boxes). The N-ClpA “interface A” is thought to be the predominant adaptor protein interaction surface *in vivo*, while “interface C” may provide another functional mode of interaction.¹⁸ The N-domain sequence of ClpC from *Synechococcus elongates*, which has been shown to interact with a ClpS homolog as well as *B. subtilis* MecA,²¹ suggests that the residues described above that are important for protein-protein interactions are conserved compared to *B. subtilis* ClpC.

MecA Adaptor Protein Interactions

The MecA adaptor protein binds to the ClpC N-domain and Linker domain.^{8, 12} We performed a NMR titration analysis to map the MecA binding surface of N-ClpCR. Data acquired in these experiments provide a global measure of the change in chemical environment at regularly dispersed backbone amide sites within the protein. The chemical shifts and resonance line widths of backbone amide nuclei are sensitive to changes in chemical environment. Backbone amide resonances that display a notable change in chemical environment (e.g., changes in chemical shift or resonance broadening due to intermediate exchange), when mapped on the N-ClpCR structure, can be used to infer the adaptor protein-binding surface.

[¹H, ¹⁵N]-TROSY-HSQC experiments were performed using a fixed concentration of ¹⁵N-labeled N-ClpCR (0.5 mM) and increasing concentrations of unlabeled MecA (30, 50, 100 and 200 μM). After the 50 μM MecA titration point, a large number of peaks in the [¹H, ¹⁵N]-TROSY-HSQC spectra exhibit a significant decrease in intensity and fall below the level of detection into the background noise. These types of changes are typically indicative of an intermediate exchange event at the amide resonances in question. We note two likely contributions to the resonance broadening and subsequent peak disappearance for N-ClpCR. First, the large complex formed upon binding (e.g., MecA dimer ~52 kDa, N-ClpCR monomer ~16 kDa) would cause an increase in the rotational correlation time and a concomitant broadening and signal reduction of the backbone amide resonances. The second contribution is a reduction in the binding affinity of N-ClpCR as compared to full-length ClpC, resulting in possible intermediate exchange effects due to a weak interaction between N-ClpCR and the MecA proteins studied herein. A previous study revealed that ClpC variants lacking the N-domain or Linker domain display significantly reduced binding affinity compared to full-length ClpC.⁸

Because of the dramatic decrease in intensities for all backbone amide resonances, a straightforward quantitative analysis of N-ClpCR peak intensities as a function of adaptor protein concentration proved difficult. However, analysis of the data at the initial, lower concentrations revealed that some peaks display more pronounced changes at the initial titration points (Figure 5a and Table S1). Backbone amide resonances with the most pronounced decrease in intensity between (i) the base and 30 μ M MecA titration point include {T7, K85, K86, I88, L90, D93, V103, and L110}; and (ii) between the 30 μ M and 50 μ M MecA titration points include {E8, H26, G44, Q72, M74, S75, T77, A84, and L98}. Backbone amide resonances that display the most pronounced changes in chemical shift between the base spectrum and the 50 μ M MecA titration point include {E32, G56, Y80, R83, L110, I113, G118, S130, and N132}. Many of these residues localize to a single face of N-ClpCR (Figure 5b) comprising the flexible loop, α 5, the small parallel β -strand {T7, E8, and V103}, and surrounding loops. This surface overlaps the aforementioned N-ClpA “interface A” surface (Figure 5c and 5d) involved in the interaction with ClpS. A few backbone amide resonances reveal initial changes in the region defined as “interface C” in the N-ClpA/ClpS studies, including {L98, V103, G118, S130, and N132}. Consistent with previous studies,⁸ the binding of MecA to N-ClpCR was further verified using size-exclusion liquid chromatography (SELC; Figure S1). In summary, the results of our N-ClpCR adaptor protein NMR titrations are consistent with studies describing the interaction between N-ClpA/ClpS and suggest that the MecA adaptor protein binds to N-ClpCR primarily using the “interface A” surface utilizing a surface crevice created by the flexible loop region, with additional interactions in the “interface C” surface.

¹⁵N Backbone Relaxation Measurements and Fast Time-Scale Motions

Analysis of ¹⁵N backbone relaxation rates provides detailed information about fast ps-ns time-scale motions and identifies backbone amide resonances that may possess slower μ s-ms chemical exchange phenomena. Disorder, such as that described above for the acidic loop region in ClpA, or motions on the μ s-ms time-scale often correlate to regions involved in protein function, including protein-protein interactions.²² ¹⁵N backbone relaxation rate constants were measured for N-ClpCR at two magnetic field strengths, allowing for a robust model-free analysis of fast internal motions. Summaries of ¹⁵N relaxation parameters are found in Supplemental Data (Figure S2 and Table S2, respectively). The {¹H}-¹⁵N NOE profile reveals that residues R70-I78 within the flexible loop undergo very fast internal motions, as the values drop below 0.6. As previously mentioned, the corresponding flexible loops in the N-ClpA and N-ClpB crystal structures have high B-factors or were unobservable in the electron density. Residues that deviate from the mean values in plots of R_2/R_1 , R_1R_2 , R_2/η_{xy} , R_1R_2 vs. R_2/R_1 , and $^1D_{NH}$ vs. R_2/R_1 (Figure S2) include {M2, F3, G4, G30, V39, G42, G44, G56, I68, G71, H79, R83, G116, G118, and V119}, suggestive of conformational exchange on the μ s-ms time-scale.²³⁻²⁷ The fitting statistics for the model-free analysis as implemented in RELAX²⁸ suggest the prolate spheroid provides the best model for global tumbling with the following converged parameters: $\tau_m = 6.46$ ns, $D_a = 9.89 \times 10^{-6}$ s⁻¹, $\theta = 78.95^\circ$, and $\phi = 175.66^\circ$. Summaries of the model-free motional parameters are found in Supplemental Data (Figure S3 and Table S3, respectively).

The S^2 generalized order parameter is related to the distribution of the N-H bond vector orientations; a value of 0 indicates completely unrestricted motion, whereas a value of 1 indicates completely restricted internal motion. A plot of the S^2 general order parameters (Figure 6a and 6b) reveals several regions with notably increased mobility, or low S^2 values, compared to the average core of residues. These regions include the N- and C-termini, the flexible loop {R70-I78} between the four-helical fold repeats, as well as the loops between α 1- α 2 {N28-G30}, α 3- α 4 {L53-S57}, α 5- α 6 {G99-E106}, and α 7- α 8 {S130}. The model-free motional models (m5-m8; Table S3) selected for backbone amide resonances in the flexible

loop in particular suggest that this region is characterized by complex motions. The large S^2 error bars result from uncertainties in model selection, rather than increased uncertainty of the raw NMR data. In general, residues with lower S^2 values correlate well to regions with higher $C\alpha$ ML superposition R.M.S.D. values (Figure 6a, red line). Six residues have significant chemical exchange contributions (R_{ex}) to the ^{15}N R_2 relaxation rates, indicative of motions on the μs -ms time-scale (Figure S3, Table S3, and Figure 6b), including N28, V39, G42, I68, K85, and V119.

Principal components analysis (PCA) is a standard tool in the field of multivariate analysis for extracting from a set of inter-related variables a much smaller set that retains most of the variation contained in the full set. By setting up a correlation matrix whose elements are the ensemble average of the pairwise products of displacements from their average position of landmarks such as the $C\alpha$ positions in a protein, principal components analysis can be very helpful in identifying, from a NMR structural ensemble, correlations in conformational rearrangements within a protein.²⁹ The PCA plot first principle component of the N-ClpCR NMR structure ensemble (Figure 6c) reveals that on average within the ensemble: (i) $\alpha 1$, the $\alpha 1$ - $\alpha 2$ loop, and the flexible loop between $\alpha 4$ and $\alpha 5$ are self-correlated (i.e., move in the same direction); (ii) $\alpha 2$, $\alpha 5$, $\alpha 6$, and $\alpha 7$ are self-correlated but move in an opposite sense, or anti-correlated, to (i). The pseudo two-fold symmetry comprising the N-ClpCR core, namely $\alpha 2$ and $\alpha 6$, are self-correlated; however, this region is anti-correlated as a whole to the flexible loop region. The origin of the anti-correlated groups (i and ii) maps to the adaptor protein-binding surface (Figure 5) near the basin of the binding crevice formed at the interface between groups (i) and (ii) formed by the start of $\alpha 2$, $\alpha 5$, and $\alpha 7$, as well as the end of $\alpha 6$. In general, these residues comprise the “interface A” and “interface C” surfaces. Furthermore, residues comprising $\alpha 2$, $\alpha 5$, and $\alpha 6$ within group (i) have lower S^2 values compared to the average core of residues, suggesting that the inherent flexibility of group (i) may modulate the size of the adaptor protein-binding crevice.

In summary, backbone amide resonances with increased backbone mobility (low S^2 values) or significant R_{ex} contributions correlate well to the aforementioned N-ClpA/ClpS “interface A” and “interface C” regions, as well as the MecA-binding surface of N-ClpCR identified in this study. Our data suggest that residues with disorder or μs -ms time-scale backbone motions, in particular near the adaptor protein-binding crevice, play a direct role in the interaction between adaptor proteins and N-ClpCR, as well as HSP100 N-domains in general.

Domain Contributions to HSP100/Clp Protein Interactions

HSP100 AAA+ proteins utilize two general mechanisms that provide the capability for broad and high substrate specificity, namely the core AAA+ domains and adaptor proteins.⁴ The core AAA+ domains implicated in providing recognition specificity include the N-domain and the Linker domain. The N-domain, although variable across the entire AAA+ family, has conserved structural features in a subset of AAA+ proteins. Despite the structural conservation of the N-domain among N-ClpA, N-ClpB, and N-ClpCR (discussed above), the specific functional mechanisms contributed by the N-domain in the recognition of target proteins are not as well understood. The N-domain is not found in all AAA+ proteins, nor is it required for fundamental activities in ClpA and ClpB, suggesting an auxiliary role for the N-domain in substrate unfolding.⁴

Differences in the electrostatic surface potential of N-ClpA, N-ClpB, and N-ClpCR may contribute to their adaptor protein and substrate target requirement or specificity. N-ClpA displays a high degree of negative charge in the flexible acidic loop region (Figure 7a, left panel) and the surface distal from the “interface A” ClpS interaction site (Figure 7a, right panel), whereas the “interface C” surface is primarily hydrophobic with a minor degree of positive charge (Figure 7a, left panel). N-ClpB displays similar characteristics to N-ClpA (Figure 7b),

although the magnitude of negative charge is less in comparison. Furthermore, the N-ClpB flexible loop, which is ~3 residues shorter than N-ClpA, was not visible in the crystal structure; however, this region contains a mixture of negatively charged, polar, and hydrophobic residues (Figure 2). The electrostatic surface of N-ClpCR, on the other hand, is significantly different compared to N-ClpA and N-ClpB. The flexible loop surface region is primarily hydrophobic, and the surface equivalent to the N-ClpA “interface C” has a high degree of positive charge contributed by residues {R83, K85, K86, R96, K97, R114, K133, and R135} (Figure 7c, left panel). The N-ClpCR surface distal from the “interface A” surface (Figure 7c, right panel) has a high degree of hydrophobicity, whereas for N-ClpA, and to a lesser degree N-ClpB, this surface is negatively charged. It is tempting to speculate that the drastic difference in electrostatic surface of N-ClpCR may play a role in the unique adaptor protein requirement for ClpC and, as detailed below, substrate and adaptor protein interaction specificity.

The Linker domain is also implicated in HSP100 adaptor protein and substrate recognition. The Linker domain is located between the D1 (AAA-1) and D2 (AAA-2) motifs of some HSP100 proteins and forms a flexible region comprised of coiled helices.⁴ As mentioned previously, the *B. subtilis* ClpC N-domain and Linker domain are both important for the interaction with adaptor proteins. Deletion of either domain results in drastically reduced binding of the MecA adaptor protein.⁸ It therefore seems plausible that the N-domain and Linker domain may be close in physical proximity and work collectively in the interaction with adaptor proteins.

Although no high-resolution structural information has been reported for full-length *B. subtilis* ClpC, a previous study revealed that the active form of the MecA/ClpCP protease complex (~750 kDa) involves two ClpC hexamers, six MecA molecules per ClpC hexamer, and a double heptameric ClpP barrel.⁸ Additionally, there are conflicting reports in the literature concerning the quaternary arrangement of the N-domain and Linker domain in other HSP100/AAA+ proteins. The first structural study describing a full-length HSP100/AAA+ protein was of ClpA bound to ADP,¹⁷ suggesting that its N-domain is oriented such that contacts are made between the N-ClpA $\alpha 4$ and the nucleotide binding domain 1 (NBD1), resulting in the acidic loop pointing towards NBD1. A similar arrangement was noted for *Thermus thermophilus* ClpB (TClpB) bound to AMPPNP in one of three structures in the crystallographic asymmetric unit.³⁰ However, TClpB was noted to possess a large degree of plasticity because, as one example, the N-domain was rotated up to 120° relative another molecule in the crystallographic asymmetric unit. This observation, as well as data from cryo-electron microscopy (cryo-EM),^{31, 32} has led to the notion that the N-domains of HSP100/AAA+ proteins are highly dynamic as a whole domain, extending from the HSP100/AAA+ core akin to a balloon on a string, and their orientation may depend on a number of factors including the active state, identity of the nucleotides bound by NBD1 and NBD2, the presence of adapter proteins, or the presence of a target (protein) molecule. Of note, dramatic differences are noted in the quaternary structures of TClpB bound to different GTP analogs, AMPPNP and ATP γ S.³²

Recently, a hexameric structural model of *Saccharomyces cerevisiae* Hsp104 determined using rigid-body fitting of individual domains into a cryo-electron microscopy (cryo-EM) map revealed a novel arrangement of individual Hsp104 molecules compared to hexameric models reported for other HSP100 proteins derived using rigid-body fitting of full-length crystal structures.³³ In the Hsp104 cryo-EM model, the flexible loop of the Hsp104 N-domain points away from the nucleotide binding domain-1 (NBD1), positioned in a ring-like arrangement with part of the Linker domain interacting with the N-domain ring. Hexameric models of other HSP100 proteins, developed from monomeric x-ray structures alone¹⁷⁻¹⁹ or x-ray structures mapped into cryo-EM data,³⁰⁻³² are inconsistent with the new Hsp104 cryo-EM model. The authors suggest that the differences may be a result of the different biological activities of these

proteins, or differences produced as a result of crystallization, as domain orientations may differ in solution as compared to crystal.³³

When compared to models developed from the aforementioned ClpA, ClpB, and TClpB studies, the Hsp104 cryo-EM model presents a better, albeit somewhat similar, spatial arrangement for describing the possible arrangement of the ClpC N-domain and Linker domain and the possible hexameric arrangement of N-ClpCR consistent with biological data describing the cooperative function of the N-domain and Linker domain. A superposition of N-ClpCR to the N-domain of the cryo-EM fitted Hsp104 model (Figure 8a) suggests that the surface comprising $\alpha 5$, $\alpha 6$, and $\alpha 8$, described as “interface C” above, may contribute to an interaction between the ClpC N-domain and Linker domain. We note, however, that the ClpC Linker domain has fewer residues compared to Hsp104, and therefore the exact residues involved in a possible interaction remain unclear. To visualize the possible electrostatic contributions of a N-ClpCR hexamer, six N-ClpCR molecules were superimposed to the hexameric cryo-EM Hsp104 N-domain ring (Figure 8b). The N-ClpCR hexamer reveals a notable degree of positive charge on the periphery, transitioning to hydrophobicity and a small degree of negative charge at the ring center. When compared to the residues most affected in the MecA NMR titration (Figure 8c) and the N-ClpCR S^2 values (Figure 8d), the N-ClpCR hexameric model suggests that the flexible loop/hydrophobic and $\alpha 5$ /positively charged regions, as well as residues with considerable disorder, contribute in the binding of adaptor proteins. A few stretches of negatively charged residues are present in the MecA amino acid sequence that might facilitate an interaction with the positively charged N-ClpCR surface. It is tempting to speculate that the drastic difference in the electrostatic surface of N-ClpCR, compared to N-ClpA and N-ClpB, contributes to protein interaction specificity of HSP100 proteins and possibly the adaptor protein requirement of ClpC specifically. Furthermore, the ClpC Linker domain contains a notable number of charged residues on average as compared to the full-length protein, suggesting that charged residues might contribute in binding adaptor proteins.

Differences in the primary sequences of ClpS and MecA, in correlation with the respective N-ClpA and N-ClpCR electrostatic surfaces, further support a model for N-domain electrostatic discrimination of adaptor proteins. For example, ClpS residues P24, P25, and Y28 bind to the hydrophobic surface of N-ClpA as described above (Figure 7a, left panel). The homologous residues in MecA that would bind to the positively charged surface of N-ClpCR, inferred via ^{TCOFFEE} sequence alignment of ClpS and MecA, are E24, P25, and E28. Additionally, ClpS $\alpha 3$ makes contacts to the N-ClpA hydrophobic-to-negative charge region; the homologous region in MecA would make contacts to the N-ClpCR positive charge-to-hydrophobic region. In concordance, ClpS $\alpha 3$ is more positively charged at the C-terminal end compared to MecA.

The N-domain of ClpA as a whole, as compared to the core body of ClpA, has been described as highly mobile – akin to a balloon (N-ClpA) on a string (ClpA core) – and this domain mobility is thought to aid in delivering substrate to the distant ClpA pore.³¹ The N-domain of ClpC may possess a similar “domain mobility” function, where complex motions – e.g., N-ClpCR internal flexibility (the N-domain flexible loop region), as well as overall N-domain mobility (tethered to the core body of ClpC as a whole) – are involved in not only the recognition of adaptor proteins (e.g., MecA), but also in the hand-off of substrates (e.g., ComK) to the ClpC pore. Future studies are warranted to delineate the precise role of the ClpC N-domain and Linker domain in the recognition of adaptor proteins, in particular in relation to the unique adaptor protein requirement of ClpC, as well as the possibility that differences in electrostatic surfaces contribute to HSP100 adaptor protein and substrate recognition specificity.

Materials and Methods

NMR Spectroscopy and Structure Calculation

N-ClpCR (**P37571**) protein preparation and resonance assignment was described previously.³⁴ SELC and micro-electrospray ionization mass spectrometry revealed a monomeric state under the conditions studied (data not shown). NMR samples contained 0.5-1.0 mM N-ClpCR, 20 mM MES (pH 5.5), 50 mM NaCl, and 2 mM EDTA. Pulse sequences for all NMR experiments were written in-house based on previously described methods.³⁵ NMR spectra were processed with NMRPIPE/NMRDRAW³⁶ and analyzed with NMRVIEW.³⁷

NOE restraints were obtained from experiments listed in Supplemental Data. One-bond ^1H - ^{15}N ($^1D_{\text{NH}}$) residual dipolar couplings (RDCs) were measured from 2D IPAP [^1H , ^{15}N]-HSQC spectra³⁸ using a stretched 7% polyacrylamide gel.³⁹ Backbone ϕ and ψ angle restraints were obtained from analysis of backbone chemical shifts using TALOS⁴⁰ and $^3J_{\text{HNH}\alpha}$ coupling values.⁴¹ Hydrogen bond restraints were obtained experimentally from hydrogen exchange and qualitatively via α -helical and β -strand NOE patterns.⁴²

Initial structure calculations were performed using ARIA v.1.2⁴³ and CNS v.1.1.⁴⁴ All NOE peak lists were calibrated independently using spin diffusion correction.⁴⁵ $^1D_{\text{NH}}$ RDCs were excluded for residues with $\{^1\text{H}\}$ - ^{15}N NOE values < 0.5 at 600 MHz. Twenty structures were calculated per iteration from it0 to it7, whereas two hundred structures were calculated in it8. One-hundred lowest total energy ARIA it8 structures were subsequently refined using the generalized Born (GB) potential⁴⁶ implemented in AMBER⁹⁴⁷ to account for solvent effects.

Structural Analyses

Final structures were analyzed using FINDAM,⁴⁸ PROCHECK,⁴⁹ DSSP,⁵⁰ and MOLPROBITY,⁵¹ as well as the DALI,⁵² SCOP,⁵³ and PFAM⁵⁴ databases. The agreement between experimental and back-calculated $^1D_{\text{NH}}$ RDCs was determined using PALES.⁵⁵ Helical angles were measured using INTERHLX; the interhelical angle reported is 180° minus angle θ reported by VGM.⁵⁶ Structure-based figures were prepared with PYMOL (DeLano Scientific; <http://www.pymol.org>), and electrostatic surfaces were calculated using APBS⁵⁷ and the PYMOL-APBS plug-in. ML structural RMSD analysis of the N-ClpCR NMR ensemble, comparisons to other N-domain structures, and PCA analysis were performed using THESEUS.⁵⁸ Sequence alignments were performed using TCOFFEE.⁵⁹ The hexameric N-ClpCR model was created by independently superimposing the structure most similar to the average structure (well-ordered backbone atoms) in the N-ClpCR NMR ensemble as calculated by THESEUS to the N-domain of each of the six chains in the cryo-EM fitted Hsp104 model.³³

Adaptor Protein Titrations

B. subtilis MecA (**P37958**) was overexpressed using a pET-28a+ vector in *E. coli* BL21(DE3) cells as thrombin-cleavable hexa-His-tag fusion proteins. Proteins were purified using Ni-NTA affinity resin, cleaved, and dialyzed for NMR analysis. The interaction between ^{15}N -labeled ClpCR (0.5 mM) and MecA adaptor protein constructs were monitored with [^1H , ^{15}N]-TROSY-HSQC spectra. Samples were prepared by combining and concentrating the proper amounts of N-ClpCR and MecA from diluted protein stocks using a Millipore Centriprep centrifugal filter device to the appropriate volume. Titrations were performed with MecA (dimeric) concentrations of 30, 50, 100, and 200 μM . Differences in peak intensity were quantified using

an equation akin to $\Delta = \left(\frac{I_{\text{pt}}}{\sigma_{\text{pt}}} \right) / \left(\frac{I_{\text{base}}}{\sigma_{\text{base}}} \right)$, where I is the peak intensity and σ is the spectrum

baseline noise; $\Delta_{30 \mu\text{M vs. base}}$ OR $(\Delta_{30 \mu\text{M vs. base}} - \Delta_{50 \mu\text{M vs. base}})$. Chemical shift changes were quantified using an equation akin to $\sqrt{(\Delta\delta_{(x-y)}^{1\text{H}})^2 + (0.101 \times \Delta\delta_{(x-y)}^{15\text{N}})^2}$.

¹⁵N Backbone Relaxation Measurements and Model-Free Analysis

¹⁵N backbone relaxation rate constants R_1 , R_2 , and $\{^1\text{H}\}$ -¹⁵N NOE values were measured at 500 and 600 MHz for 115 and 123 residues, respectively, out of 145 total N-ClpCR residues. Pulse sequences were written in-house from methods previously described.^{60,61} Transverse cross-correlation η_{xy} relaxation rate constants²⁴ were measured at 600 MHz. R_1 and R_2 constants were obtained by fitting the peak intensities as a function of relaxation delay time to a single exponential decay function using CURVEFIT v1.4 (A. G. Palmer III, Columbia University) from data collected with fully interleaved planes. Monte Carlo simulations were performed to estimate the uncertainty in the fitted parameters. Experimental errors associated with fitting the R_1 and R_2 relaxation data were estimated from the baseline noise and the calculated uncertainty in peak intensities for duplicate data sets collected with the same relaxation delays. Steady-state $\{^1\text{H}\}$ -¹⁵N NOE values were calculated as the ratio between peak intensities with (I_{sat}) and without (I_{ref}) ¹H saturation, $\text{NOE} = I_{\text{sat}}/I_{\text{ref}}$, averaged over three replicate $\{^1\text{H}\}$ -¹⁵N NOE spectra recorded with proton saturation using a 4 s period of saturation and interleaved with the reference spectrum, which was recorded with a 5 s recycle delay and no saturation;

errors (σ) were obtained by $\sigma_{\text{noe}} = \sqrt{\frac{(\sigma_{\text{sat}} I_{\text{ref}})^2 + (\sigma_{\text{ref}} I_{\text{sat}})^2}{I_{\text{ref}}^2}}$. Experiments were collected from a ~ 1.0 mM ¹⁵N-labeled sample; measurements collected on a one-third-diluted sample indicated that there were no significant concentration-dependent effects on the relaxation rates.

RELAX²⁸ was used to fit the experimental ¹⁵N relaxation data to ten motional models of the Lipari-Szabo formalism (see Supplemental Data for details). Model-free parameters were calculated individually for five diffusion tensors (local τ_m , sphere, prolate, oblate, and ellipsoid) to determine the universal solution⁶² describing the internal dynamics. For each diffusion tensor, parameters for the ten motional models were optimized in the absence of tensor estimates⁶³ using Newton minimization for each spin system, followed by model elimination⁶⁴, and Akaike's Information Criteria (AIC) model selection^{65,66} in the form of $\text{AIC} = \chi^2 + 2k$, where k is the number of model-free parameters in the model and χ^2 describes fit of the relaxation data to the model. Finally, AIC was used to select the diffusion tensor providing the best fit and model parameter errors were calculated from two hundred Monte Carlo simulations. Values for the ¹⁵N CSA and N-H bond length (r_{NH}) were set to -172 ppm and 1.02 Å, respectively. RELAX analysis was performed on the N-ClpCR structure most similar to the average structure (well-ordered backbone atoms) in the NMR ensemble as calculated by THESEUS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Dr. Benjamin Bobay (NC State) for preparation of stretched gel samples and helpful discussions; Dr. Edward d'Auvergne for assistance with RELAX; Dr. Ron Venter (Duke) for helpful discussions; Dr. Geoff Mueller (NIEHS) for analysis scripts; Drs. Walter Chazin and Jarrod Smith (Vanderbilt) for providing FINDFAM; Dr. Michael Lerner for providing the PYMOL-APBS plug-in; Dr. Peter Wright (Scripps) for AMBER refinement scripts; Drs. Petra Wendler and Helen Saibil (Birkbeck) for the PDB file of the cryo-EM model of hexameric Hsp104; and the Cincinnati Children's Hospital Medical Center, Division of Biomedical Informatics for computational resources. Funding was provided by NIH GM065156 (JC), NIH GM057720 (DD), NIEHS T32-ES07250 (DK), and NIAID T32-AI055406 (DK).

References

1. Ogura T, Wilkinson AJ. AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* 2001;6:575–97. [PubMed: 11473577]
2. Beyer A. Sequence analysis of the AAA protein family. *Protein Sci* 1997;6:2043–58. [PubMed: 9336829]
3. Schirmer EC, Glover JR, Singer MA, Lindquist S. HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci* 1996;21:289–96. [PubMed: 8772382]
4. Mogk A, Dougan D, Weibezahn J, Schlieker C, Turgay K, Bukau B. Broad yet high substrate specificity: the challenge of AAA+ proteins. *J Struct Biol* 2004;146:90–8. [PubMed: 15037240]
5. Hung GC, Masison DC. N-terminal domain of yeast Hsp104 chaperone is dispensable for thermotolerance and prion propagation but necessary for curing prions by Hsp104 overexpression. *Genetics* 2006;173:611–20. [PubMed: 16582428]
6. Dougan DA, Mogk A, Zeth K, Turgay K, Bukau B. AAA+ proteins and substrate recognition, it all depends on their partner in crime. *FEBS Lett* 2002;529:6–10. [PubMed: 12354604]
7. Beinker P, Schlee S, Groemping Y, Seidel R, Reinstein J. The N terminus of ClpB from *Thermus thermophilus* is not essential for the chaperone activity. *J Biol Chem* 2002;277:47160–6. [PubMed: 12351638]
8. Kirstein J, Schlothauer T, Dougan DA, Lilie H, Tischendorf G, Mogk A, Bukau B, Turgay K. Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *Embo J* 2006;25:1481–91. [PubMed: 16525504]
9. Schlothauer T, Mogk A, Dougan DA, Bukau B, Turgay K. MecA, an adaptor protein necessary for ClpC chaperone activity. *Proc Natl Acad Sci U S A* 2003;100:2306–11. [PubMed: 12598648]
10. Turgay K, Hahn J, Burghoorn J, Dubnau D. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *Embo J* 1998;17:6730–8. [PubMed: 9890793]
11. Turgay K, Hamoen LW, Venema G, Dubnau D. Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev* 1997;11:119–28. [PubMed: 9000055]
12. Persuh M, Turgay K, Mandic-Mulec I, Dubnau D. The N- and C-terminal domains of MecA recognize different partners in the competence molecular switch. *Mol Microbiol* 1999;33:886–94. [PubMed: 10447896]
13. Kirstein J, Dougan DA, Gerth U, Hecker M, Turgay K. The tyrosine kinase McsB is a regulated adaptor protein for ClpCP. *Embo J* 2007;26:2061–70. [PubMed: 17380125]
14. Persuh M, Mandic-Mulec I, Dubnau D. A MecA paralog, YpbH, binds ClpC, affecting both competence and sporulation. *J Bacteriol* 2002;184:2310–3. [PubMed: 11914365]
15. Hou JY, Sauer RT, Baker TA. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nat Struct Mol Biol* 2008;15:288–94. [PubMed: 18297088]
16. Li J, Sha B. Crystal structure of the *E. coli* Hsp100 ClpB N-terminal domain. *Structure* 2003;11:323–8. [PubMed: 12623019]
17. Guo F, Maurizi MR, Esser L, Xia D. Crystal structure of ClpA, an Hsp100 chaperone and regulator of ClpAP protease. *J Biol Chem* 2002;277:46743–52. [PubMed: 12205096]
18. Xia D, Esser L, Singh SK, Guo F, Maurizi MR. Crystallographic investigation of peptide binding sites in the N-domain of the ClpA chaperone. *J Struct Biol* 2004;146:166–79. [PubMed: 15037248]
19. Guo F, Esser L, Singh SK, Maurizi MR, Xia D. Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. *J Biol Chem* 2002;277:46753–62. [PubMed: 12235156]
20. Zeth K, Ravelli RB, Paal K, Cusack S, Bukau B, Dougan DA. Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. *Nat Struct Biol* 2002;9:906–11. [PubMed: 12426582]
21. Andersson FI, Blakytyn R, Kirstein J, Turgay K, Bukau B, Mogk A, Clarke AK. Cyanobacterial ClpC/HSP100 protein displays intrinsic chaperone activity. *J Biol Chem* 2006;281:5468–75. [PubMed: 16361263]
22. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature* 2007;450:964–72. [PubMed: 18075575]

23. Kneller JM, Lu M, Bracken C. An effective method for the discrimination of motional anisotropy and chemical exchange. *J Am Chem Soc* 2002;124:1852–3. [PubMed: 11866588]
24. Kroenke CD, Loria JP, Lee LK, Rance M, Palmer AGI. Longitudinal and transverse ^1H - ^{15}N dipolar/ ^{15}N chemical shift anisotropy relaxation interference: unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules. *J Am Chem Soc* 1998;120:7905–15.
25. Seewald MJ, Pichumani K, Stowell C, Tibbals BV, Regan L, Stone MJ. The role of backbone conformational heat capacity in protein stability: temperature dependent dynamics of the B1 domain of Streptococcal protein G. *Protein Sci* 2000;9:1177–93. [PubMed: 10892810]
26. Jarymowycz VA, Stone MJ. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev* 2006;106:1624–71. [PubMed: 16683748]
27. de Alba E, Baber JL, Tjandra N. The use of residual dipolar coupling in concert with backbone relaxation rates to identify conformational exchange by NMR. *J Am Chem Soc* 1999;121:4282–3.
28. d'Auvergne EJ, Gooley PR. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J Biomol NMR* 2008;40:107–19. [PubMed: 18085410]
29. Theobald DL, Wuttke DS. Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput Biol* 2008;4:e43. [PubMed: 18282091]
30. Lee S, Sowa ME, Watanabe YH, Sigler PB, Chiu W, Yoshida M, Tsai FT. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell* 2003;115:229–40. [PubMed: 14567920]
31. Ishikawa T, Maurizi MR, Steven AC. The N-terminal substrate-binding domain of ClpA unfoldase is highly mobile and extends axially from the distal surface of ClpAP protease. *J Struct Biol* 2004;146:180–8. [PubMed: 15037249]
32. Lee S, Choi JM, Tsai FT. Visualizing the ATPase cycle in a protein disaggregating machine: structural basis for substrate binding by ClpB. *Mol Cell* 2007;25:261–71. [PubMed: 17244533]
33. Wendler P, Shorter J, Plisson C, Cashikar AG, Lindquist S, Saibil HR. Atypical AAA+ subunit packing creates an expanded cavity for disaggregation by the protein-remodeling factor Hsp104. *Cell* 2007;131:1366–77. [PubMed: 18160044]
34. Kojetin DJ, McLaughlin PM, Venters RA, Thompson RJ, Rance M, Cavanagh J. NMR assignment of the N-terminal repeat domain of *Bacillus subtilis* ClpC. *Biomolecular NMR Assignments* 2007;1:163–5.
35. Cavanagh, J.; Fairbrother, WJ.; Palmer, AG., III; Rance, M.; Skelton, NJ. *Protein NMR spectroscopy : principles and practice*. Vol. 2. Academic Press; San Diego: 2007.
36. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–93. [PubMed: 8520220]
37. Johnson BA, Blevins RA. *NMR View: A computer program for the visualization and analysis of NMR data*. *J Biomol NMR* 1994;4:603–14.
38. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 1998;131:373–8. [PubMed: 9571116]
39. Chou JJ, Gaemers S, Howder B, Louis JM, Bax A. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR* 2001;21:377–82. [PubMed: 11824758]
40. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 1999;13:289–302. [PubMed: 10212987]
41. Kuboniwa H, Grzesiek S, Delaglio F, Bax A. Measurement of H^{N} - H^{α} J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J Biomol NMR* 1994;4:871–8. [PubMed: 7812158]
42. Wüthrich, K. *NMR of proteins and nucleic acids*. Wiley; New York: 1986.
43. Linge JP, O'Donoghue SI, Nilges M. Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol* 2001;339:71–90. [PubMed: 11462826]
44. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. *Crystallography & NMR*

- system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54:905–21. [PubMed: 9757107]
45. Linge JP, Habeck M, Rieping W, Nilges M. Correction of spin diffusion during iterative automated NOE assignment. *J Magn Reson* 2004;167:334–42. [PubMed: 15040991]
 46. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 2000;56:275–91. [PubMed: 11754341]
 47. Case, DA.; Darden, TA.; Cheatham, TEI.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Merz, KM.; Pearlman, DA.; Crowley, M.; Walker, RC.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, KF.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, DH.; Schafmeister, C.; Ross, W.; Kollman, PA. AMBER 9. University of California; San Francisco: 2006.
 48. Smith JA, Gomez-Paloma L, Case DA, Chazin WJ. Molecular dynamics docking driven by NMR-derived restraints to determine the structure of the calicheamicin γ_1^I oligosaccharide domain complexed to duplex DNA. *Magn Reson Chem* 1996;34:S147–55.
 49. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996;8:477–86. [PubMed: 9008363]
 50. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–637. [PubMed: 6667333]
 51. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Bryan Arendall W 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 2007;35:W375–83. [PubMed: 17452350]
 52. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993;233:123–38. [PubMed: 8377180]
 53. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;247:536–40. [PubMed: 7723011]
 54. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. Pfam: clans, web tools and services. *Nucleic Acids Res* 2006;34:D247–51. [PubMed: 16381856]
 55. Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J Am Chem Soc* 2000;122:3791–3792.
 56. Yap KL, Ames JB, Swindells MB, Ikura M. Vector geometry mapping. A method to characterize the conformation of helix-loop-helix calcium-binding proteins. *Methods Mol Biol* 2002;173:317–24. [PubMed: 11859772]
 57. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 2001;98:10037–41. [PubMed: 11517324]
 58. Theobald DL, Wuttke DS. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* 2006;22:2171–2. [PubMed: 16777907]
 59. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205–17. [PubMed: 10964570]
 60. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 1994;33:5984–6003. [PubMed: 7514039]
 61. Skelton JJ, Palmer AGI, Akke M, Kördel J, Rance M, Chazin WJ. Practical aspects of two-dimensional proton-detected ^{15}N spin relaxation measurements. *J Magn Reson ser B* 1993;102:253–64.
 62. d'Auvergne EJ, Gooley PR. Set theory formulation of the model-free problem and the diffusion seeded model-free paradigm. *Mol Biosyst* 2007;3:483–94. [PubMed: 17579774]
 63. d'Auvergne EJ, Gooley PR. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J Biomol NMR* 2008;40:121–33. [PubMed: 18085411]

64. d'Auvergne EJ, Gooley PR. Model-free model elimination: a new step in the model-free dynamic analysis of NMR relaxation data. *J Biomol NMR* 2006;35:117–35. [PubMed: 16791734]
65. d'Auvergne EJ, Gooley PR. The use of model selection in the model-free analysis of protein dynamics. *J Biomol NMR* 2003;25:25–39. [PubMed: 12566997]
66. d'Auvergne, EJ. Protein dynamics: a study of the model-free analysis of NMR relaxation data. University of Melbourne; 2006.

Figure 1. NMR solution structure of N-ClpCR

(a) Stereo backbone trace of the thirty lowest-energy NMR solution structures colored using the chainbow feature of *PYMO*L (N-terminus, blue; C-terminus, red).

(b) Cartoon diagram of the thirty lowest-energy NMR solution structures colored using the chainbow feature of *PYMO*L (N-terminus, blue; C-terminus, red).

(c) Putty/sausage diagram illustrating regions of structural variation within the NMR ensemble colored using the chainbow feature of *PYMO*L from small (blue) to large (red).

(d) Cartoon diagram illustrating the sequentially similar/identical regions and hydrophobic contacts between the $\alpha 2$ (red) and $\alpha 6$ (blue) giving rise to a pseudo two-fold axis of symmetry. Color scheme according to the N-ClpCR repeat sequence in Figure 2A.

Figure 2. Secondary structure and sequence alignment of N-ClpCR to N-ClpA and N-ClpB

DSSP calculated secondary structure of N-ClpCR and *TCOFFEE* sequence alignment to *E. coli* N-ClpA and N-ClpB. Secondary structure elements comprising the N-ClpCR NMR structure are noted in *DSSP* code ((H, α -helix; B or E, β -strand; T, hydrogen-bonded turn) and diagram form (top). Conservation notation is as follows: identical (*), highly conserved (:) and somewhat similar (.). Repeat regions (bold font and underlined) are colored as follows: similar (green or orange) and identical (blue or red). See text for a description of colored box regions. Organism notations is as follows: B, *B. subtilis*; E, *E. coli*; S, *S. elongates*.

Figure 3. Structures of HSP100 N-domain proteins

(a) Cartoon diagram of the N-ClpA crystal structure [PDB code: 1r6c].

(b) Cartoon diagram of the N-ClpB crystal structure [PDB code: 1khy; chain A].

(c) Cartoon diagram of the N-ClpCR NMR solution structure closest to the mean calculated ML superposition structure reported by *THESEUS*.

Figure 4. Structural superposition of N-ClpA, N-ClpB, and N-ClpCR

(a) Ribbon diagram of the ML superposition of N-ClpA, N-ClpB, and N-ClpCR.

(b) Putty/sausage diagram illustrating regions of structural variation in the ^{THESEUS} ML superposition of N-ClpA, N-ClpB, and N-ClpCR.

Figure 5. Mapping of the MecA-binding surface of N-ClpCR

(a) Examples of MecA-induced peak intensity decreases in the $[^1\text{H}, ^{15}\text{N}]$ -TROSY-HSQC spectra spectra of N-ClpCR at 30 μM (black) and 50 μM (blue) MecA, as compared to the N-ClpCR base spectrum (red).

(b) N-ClpCR residues that display larger initial changes in chemical shift perturbation (green), line broadening (peak intensity decrease; red), or both (yellow). Residues highlighted correspond to values that differ from the mean value by $(1.2 \times \sigma)$.

(c) N-ClpA (white) and ClpS (orange) co-crystal structure [PDB code: 1r6q].

(d) N-ClpA residues involved in the interaction with ClpS via interface A/acid loop surface (orange) or the non-specific binding interface C surface (blue).

Figure 6. N-ClpCR backbone dynamics and PCA correlations

(a) Backbone order parameters (S^2 ; black circles) and ML $C\alpha$ R.M.S.D. (red line) vs. residue number.

(b) S^2 and R_{ex} values plotted on the N-ClpCR structure. The radius of the backbone is directly proportional to $1-S^2$ (green regions); a larger radius indicates increased flexibility. White regions indicate the lack of experimental data, and the backbone radii of these residues were interpolated between neighboring residues. Residues with R_{ex} contributions to the ^{15}N R_2 relaxation rates (red spheres) are indicative of motions on the μs -ms timescale.

(c) The first principal component from the maximum likelihood correlation matrix for the ML superposition of the N-ClpCR structure ensemble. Regions colored similarly (cyan or magenta) are self-correlated, whereas regions colored differently (cyan vs. magenta) are anti-correlated.

Figure 7. N-domain electrostatic surface characteristics. $_{APBS}$ -calculated electrostatic potential displayed on the solvent accessible surfaces of (a) N-ClpA, (b) N-ClpB, and (c) N-ClpCR. The structures displayed in the left panels are oriented as shown in Figure 3A, 3B, and 3C, respectively.

Figure 8. Model of ClpC N-domain and Linker domain interactions and N-ClpCR hexameric ring structure based on the cryo-EM Hsp104 model

(a) Structural superposition of N-ClpCR to the cryo-EM model of Hsp104.³³ The Hsp104 model is shown as a α /sphere trace colored according to the domain identity (yellow, N-domain; red/ orange, NBD1; green, linker domain; blue/light blue NBD2), whereas the N-ClpCR structure is shown with cylindrical helices colored as in Figure 1A.

(b) $_{APBS}$ -calculated electrostatic potential of the hexameric model of N-ClpCR viewed looking down the C6 axis of symmetry. Six N-ClpCR molecules were super-positioned to a monomer within the Hsp104 cryo-EM model.

(c) N-ClpCR residues that display a large initial line broadening and/or chemical shift changes in the MecA NMR titration mapped onto the hexameric model of N-ClpCR. Color coding is the same as that shown in Figure 5B.

(d) S^2 values mapped onto the hexameric model of N-ClpCR. Amino acid side-chains are colored white-to-brown as represented in Figure 6A and 6B and proportional to $1-S^2$. One monomer is removed from the hexamer to illustrate the monomeric contribution to the hexameric assembly.

Table 1

Structural statistics for the N-ClpCR structure ensemble (n=30)

NMR constraints	
Distance total	6910
Ambiguous	2346
Unambiguous	4564
Intra-residue ($i = j$)	1288
Sequential ($ i - j = 1$)	992
Medium-range ($1 < i - j \leq 4$)	1089
Long-range ($ i - j \geq 5$)	1195
Hydrogen bonds	66
Dihedral angles (ϕ and ψ)	208
$^1D_{HN}$ RDCs	78
AMBER Constraint Violations	
Average number of distance constraint violations per structure	
$0.1\text{\AA} < d < 0.2\text{\AA}$	1.833±1.098
$0.2\text{\AA} < d$	0
Maximum distance violation ($\text{\AA}$)	0.172
Average number of dihedral angle violations per structure, $\theta > 0^\circ$	0 *
Maximum dihedral angle violation ($\text{\AA}$)	0 *
Ensemble superposition statistics ($\text{\AA}$) [†]	
All (residues 1-145)	
Backbone atoms	0.824±0.104 (0.331)
Heavy atoms	1.501±0.119 (0.602)
Well-ordered regions (residues 5-70 and 78-142)	
Backbone atoms	0.476±0.099 (0.191)
Heavy atoms	1.197±0.113 (0.481)
AMBER energies (kcal mol ⁻¹)	
Constraint	9.641±1.294
Bond	4.067±0.639
Angle	0.234±0.010
Torsion	0.555±0.165
Align	4.785±1.282
Lennard-Jones energy	-1200.656±7.502
Total	-6032.179±6.066
PROCHECK statistics [‡]	
Ramachandran space (% residues)	
Most favored	95.3
Additionally allowed	4.3
Generously allowed	0.3
Disallowed	0.1
MOLPROBITY statistics **	
Clashscore	5.68±0.62
Clashscore percentile	91.60±2.14

Ramachandran space (%)

Favored	96.81±1.21
Allowed	2.73±1.16
Outliers	0.46±0.51

* TALOS-predicted dihedral angles only; when AMBER chirality restraints are included, this value is 1.433±1.086 with a maximum violation of 4.538°.

† Ensemble superposition statistics were calculated using THESEUS, reported as classical least-squares (LS) pairwise <r.m.s.d> ± ML <σ> (LS <σ>).

‡ PROCHECK statistics calculated over the entire structure ensemble.

** The MOLPROBITY clashscore is the number of serious steric overlaps (> 0.4Å) per 1000 atoms; Clashscore percentile: 100th is the best among structures of comparable resolution; 0th is the worst.