

Unusual Arginine Formations in Protein Function and Assembly: Rings, Strings, and Stacks

Marco A. C. Neves,^{†,‡} Mark Yeager,^{*,§,||,⊥} and Ruben Abagyan^{*,†}

[†] Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, United States

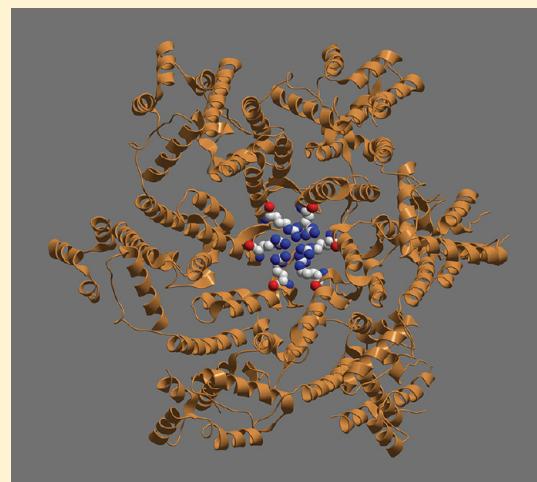
[‡] Centro de Neurociências, Lab. Química Farmacêutica, Faculdade de Farmácia, Universidade de Coimbra, Pólo das Ciências da Saúde, 3000-548 Coimbra, Portugal

[§] Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, Charlottesville, Virginia 22908, United States

^{||} Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92037, United States

[⊥] Division of Cardiovascular Medicine, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia 22908, United States

ABSTRACT: Protein–protein interfaces are often stabilized by a small number of dominant contacts, exemplified by the overrepresentation of arginine residues at oligomerization interfaces. Positively charged arginines are most commonly involved in ion pairs of opposite charge; however, previous work of Scheraga and co-workers described the stable, close range interaction between guanidinium pairs in a solvated environment. To extend this work, we searched over 70 thousand protein structures and complexes for unusual formations of arginine residues supported by the electron density. Symmetry transformations were used to generate full assemblies. Clusters of four to eight arginine residues with C_{ζ} – C_{ζ} distances <5 Å, organized as rings with four to eight members, stacks of two arginines, and strings of stacked arginines, are commonly located at the interfaces of oligomeric proteins. The positive charge is properly balanced by negatively charged counterions in about 90% of the cases. We also observed planar stacking of guanidinium groups, bridged by hydrogen bonds and interactions with water molecules. The guanidinium groups are commonly involved in five hydrogen bonds with water molecules and acceptor groups from surrounding amino acids. Water molecules have a bridging effect on the arginine pairs, but in some cases, small molecular weight chemicals in the crystallization buffer may be misinterpreted as water molecules. In summary, despite electrostatic repulsion, arginines do form various clusters that are exposed to interact with and potentially be controlled or switched by charged metabolites, membrane lipids, nucleic acids, or side chains of other proteins. Control of the stability of arginine clusters may play an important role in protein–protein oligomerization, molecular recognition, and ligand binding.



INTRODUCTION

Arginine residues are important mediators of protein–protein and protein–nucleic acid interactions in a wide range of biological processes.^{1,2} Unique physicochemical properties of arginine residues are related to the positively charged guanidinium with five hydrogen bond donors and a special type of aromatic system.³ Charged residues such as arginine are preferentially localized on the surface of proteins, whereas hydrophobic residues are predominantly buried in the protein core.⁴ The extended flexibility of the arginine side chain and the extremely high pK_a value of the guanidinium group facilitate the interaction with negatively charged groups. About 40% of the salt-bridges in proteins involve ion pairs between arginine and carboxylate groups of acidic amino acids.⁵ The positive charge

is also important for planar stacking and polar interactions with aromatic residues.^{6,7} In addition, arginine residues can establish multiple hydrogen bonds with nucleic acids and play an important role in the catalytic mechanism of a number of enzymes.⁸ Moreover, arginines make dominant energetic contributions to the stabilization of protein–protein complexes and are one of the most common residues of “hot spot” protein interfaces.^{9,10} While protein interfaces are commonly depleted of charged residues such as lysine, aspartic acid, and glutamic

Special Issue: Harold A. Scheraga Festschrift

Received: January 30, 2012

Revised: April 11, 2012

Published: April 12, 2012

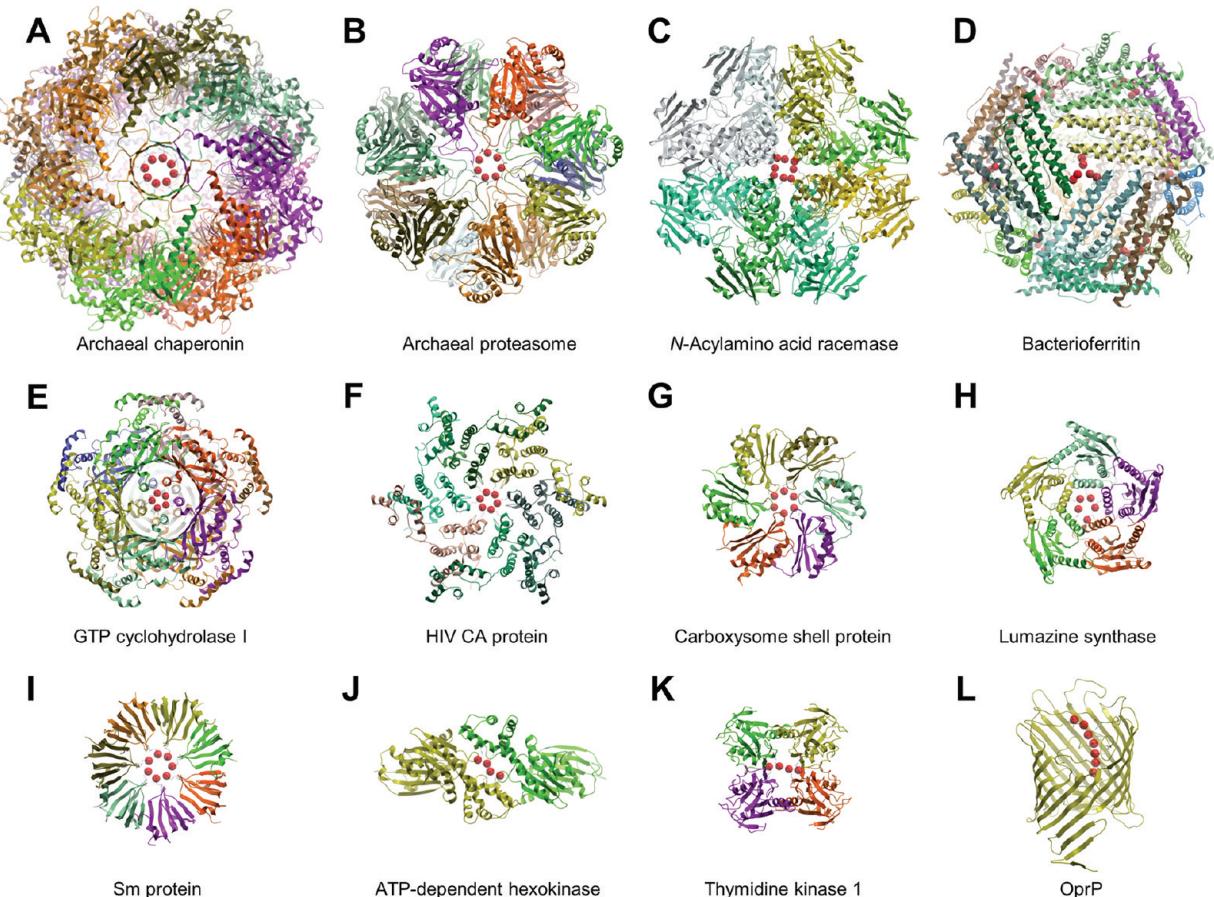


Figure 1. Representative examples of arginine clusters with $C_{\zeta}-C_{\zeta}$ distances $<5 \text{ \AA}$ obtained from the PDB database. (A) Archaeal chaperonin (PDB entry 3kfb); (B) archaeal proteasome (PDB entry 1j2q); (C) *N*-Acylamino acid racemase (PDB entry 1r0m); (D) bacterioferritin (PDB entry 1nf4); (E) GTP cyclohydrolase I (PDB entry 1wpl); (F) HIV CA protein (PDB entry 3h4e); (G) carboxysome shell protein (PDB entry 2a10); (H) lumazine synthase (PDB entry 2f59); (I) Sm protein (PDB entry 1i8f); (J) ATP-dependent hexokinase (PDB entry 2e2q); (K) thymidine kinase 1 (PDB entry 2wyj); (L) outer membrane protein (PDB entry 2o4v). The protein structures are represented with a ribbon model and colored by subunit. C_{ζ} atoms are represented in red with space filling spheres.

acid, the number of arginines at protein interfaces is larger than expected from random distribution.¹¹ In addition, guanidinium groups are significantly overrepresented at protein–protein interfaces, as opposed to other polar and charged groups, such as the terminal nitrogen in lysine, aspartate and glutamate carboxylates, and polar amide groups of asparagine and glutamine.¹²

Ion-pairs in protein structures are formed mostly between residues of opposite charge. However, previous work by Scheraga and co-workers identified an unusually large number of arginine–arginine pairs in protein structures from the PDB database.¹³ Because most of the negatively charged pairs were found at the protein surface, semiempirical calculations using the AM1 molecular orbital method were used to study the interaction between guanidinium moieties in the solvated environment. The authors concluded that water molecules stabilize the positively charged guanidinium groups and have a bridging effect on the arginine pairs. Recently, we found unusual clusters of five and six arginine residues in pentameric and hexameric oligomers of the HIV CA protein, respectively.¹⁴ During CA protein oligomerization into the basic building blocks for capsid formation, one Arg18 from each subunit comes into close contact at the center of the rings, suggesting a mechanism of electrostatic regulation.^{15,16} This prompted us to examine the PDB database for patterns of arginine clusters. Our

analysis showed that short-range interactions between arginine pairs and higher-order formations are commonly found at the interface of protein–protein oligomerization. Stabilizing interactions are discussed using a test set of high quality structures and representative examples from the PDB database.

EXPERIMENTAL METHODS

Clusters of positively charged arginines were identified using Molsoft Internal Coordinate Mechanics (ICM) software^{17,18} among 67 520 X-ray crystal structures with resolution $<3 \text{ \AA}$ deposited in the PDB database.¹⁹ The initial analysis was focused on arginine clusters with at least four positively charged residues. A script (available upon request) was written in the ICM language and was used to download the PDB entries and generate symmetry-related neighbors within a cutoff distance of 20 \AA . Symmetry-related residues were generated by replication of the unit cell according to the crystallographic space group and the unit cell parameters provided in the PDB files. For each arginine within the unit cell, close proximity arginines with $C_{\zeta}-C_{\zeta}$ distances $<5 \text{ \AA}$ were collected. This procedure was repeated until all arginines in close proximity were clustered together. The identified clusters at protein–protein interfaces were further confirmed to be biologically relevant using the available literature.

The arginine set was further filtered using electron density. A subset of high-quality clusters were identified by unambiguous electron density around arginine side chains to ensure their direct interaction. The maps were downloaded from the Uppsala Electron Density server²⁰ and fitted to the arginine side chains using a previously described algorithm.²¹ Arginine residues with low occupancy atoms or with side chain conformations not supported by the electron density were excluded from the analysis. Acidic residues, hydrogen bond acceptors, and water molecules were counted, and the electron density for potentially misassigned or unassigned anions was analyzed (see below). Negatively charged hetero groups were counted after assigning MMFF atom types.²²

RESULTS AND DISCUSSION

As anticipated by Scheraga and co-workers, we found a large number of short-range interactions between positively charged arginine side chains in the PDB database. While most of the clusters were arginine pairs, a number of unusual geometrical formations with up to eight residues was found. Representative examples are shown in Figure 1 and listed in Table 1. Clusters

Table 1. Clusters of Four to Eight Arginines Collected from the PDB Database^a

protein	no. of arginines and type	PDB entry	description
archaeal chaperonin	8 – ring	3kfb	cylindrical protein complex with two stacking rings of 8 subunits each ³⁴
N-acylamino acid racemase	8 – stack	1r0m	homooctameric protein with a 4-fold symmetry axis ^{25,26}
Sm protein	7 – ring	1i8f	heptameric protein in a ring ^{41,53}
archaeal proteasome	7 – ring	1j2q	four heptameric proteins in a barrel-shaped structure ³⁵
outer membrane protein	7 – string	2o4v	transporter protein ⁴²
HIV CA protein	6 – ring	3h4e	protein hexamer; assembles into hexamers and pentamers to form closed shell capsids ¹⁶
bacterioferritin	6 – ring	1nf4	spherical shell protein with 24 subunits ⁴⁰
carboxysome shell protein	6 – ring	2a10	protein hexamer; assembles into hexamers and pentamers to form polyhedral compartments ³⁷
GTP cyclohydrolase I	5 – ring	1wpl	two rings of 5 subunits each ²⁷
lumazine synthase	5 – ring	2f59	pentameric protein in a ring ⁵⁴
ATP-dependent hexokinase	4 – string	2e2q	protein dimer ³⁰
thymidine kinase 1	4 – string	2wvj	protein tetramer ³¹

^aCommon types of arginine clusters include rings, stacks, and strings of stacked residues.

of four to eight arginines with $C_{\zeta}-C_{\zeta}$ distances $<5 \text{ \AA}$, organized as rings with four to eight members, stacks of two arginines, and strings of stacked arginines, are commonly located at the interface of oligomeric proteins. Furthermore, these arginine formations are usually exposed to interact with other external partners not necessarily present in the analyzed structures. These potential interaction partners may include charged metabolites, membranes, nucleic acids, and other proteins (possibly depending on their phosphorylation state).

Protein–Protein Oligomerization. The hydrophobic and aromatic interaction between nonpolar residues, a major driving force for protein folding and protein–protein assembly, plays an important role in defining the architecture of oligomeric interfaces. However, subunit interface areas are more hydrophilic than a typical protein core, and have a larger number of polar and aromatic residues.²³ Early studies showed that arginine residues are commonly involved in hydrogen bonding contacts at the interface of oligomeric proteins.²⁴ Moreover, the protein interface recognition method (PIER) demonstrated that guanidinium groups are significantly overrepresented at protein interfaces.¹² Our results show that quaternary interactions between clusters of arginines, negatively charged counterions, hydrogen bonding residues, and water molecules are involved in protein oligomer stabilization.

Figure 1 displays representative protein complexes with arginine clusters. *N*-Acylamino acid racemase catalyzes the interconversion of *N*-acylamino acid enantiomers. The 1.3 Å X-ray crystal structure of the enzyme reveals a homo-octameric architecture with individual subunits packed around a 4-fold symmetry axis.^{25,26} As shown in Figure 1C, two arginines from each of the four centrally located subunits form a large cluster of eight positive charges. This unusually large cluster of arginines is exposed to the solvent and stabilized by a network of hydrogen bonds between the guanidinium groups, water molecules, and the carbonyl oxygen of Arg237. Figure 2A illustrates the hydrogen bonding network between a pair of adjacent arginines with $C_{\zeta}-C_{\zeta}$ separated by 3.5 Å. The guanidinium groups adopt a planar stacking conformation bridged by hydrogen bonds, involving either a pair of water molecules or one water molecule and the carbonyl oxygen of Arg237. Arginine pairs in the same subunit are stabilized by a salt bridge with Asp212 and separated by a $C_{\zeta}-C_{\zeta}$ distance of 4.6 Å.

GTP cyclohydrolase I is involved in the synthesis of the tetrahydrobiopterin cofactor. The protein assembles as two rings of homopentamers with a cluster of five arginines located at the central symmetry axis (Figure 1E). Individual arginines establish a hydrogen bond with the hydroxyl group of Ser241 located in the adjacent subunit.²⁷ The guanidinium groups form a solvent accessible ring.

Figure 1K shows thymidine kinase, a key enzyme in the synthesis of thymidine monophosphate, and Figure 1J shows hexokinase, an enzyme that catalyzes the phosphorylation of glucose into glucose-6-phosphate. Thymidine kinase has been reported to assemble as a low activity dimer and as a highly active tetramer.²⁸ Hexokinase forms dimers in solution.²⁹ Interestingly, both X-ray crystal structures show a typical string of stacked arginines with $C_{\zeta}-C_{\zeta}$ distances $<5 \text{ \AA}$.^{30,31} As shown in Figure 2B, the arginine cluster in a 2.0 Å structure of hexokinase is stabilized by salt bridges with Asp141 and by a network of hydrogen bonds. Arginines establish the maximum possible number of hydrogen bonding contacts, using either the solvent or hydrogen bond acceptors from surrounding residues. Bridges of two to four water molecules can be found at the geometrical center of the cluster (Arg149–Arg149 pair).

Highly Dynamic Oligomeric Protein Domains. We found a number of arginine clusters at highly dynamic domains of large protein complexes, such as chaperonins and the proteasome (Figure 1A and B). Molecular chaperones mediate protein folding, whereas the proteasome is involved in protein degradation.^{32,33} In both cases, the interacting proteins are processed in an inner chamber, accessible through an axial

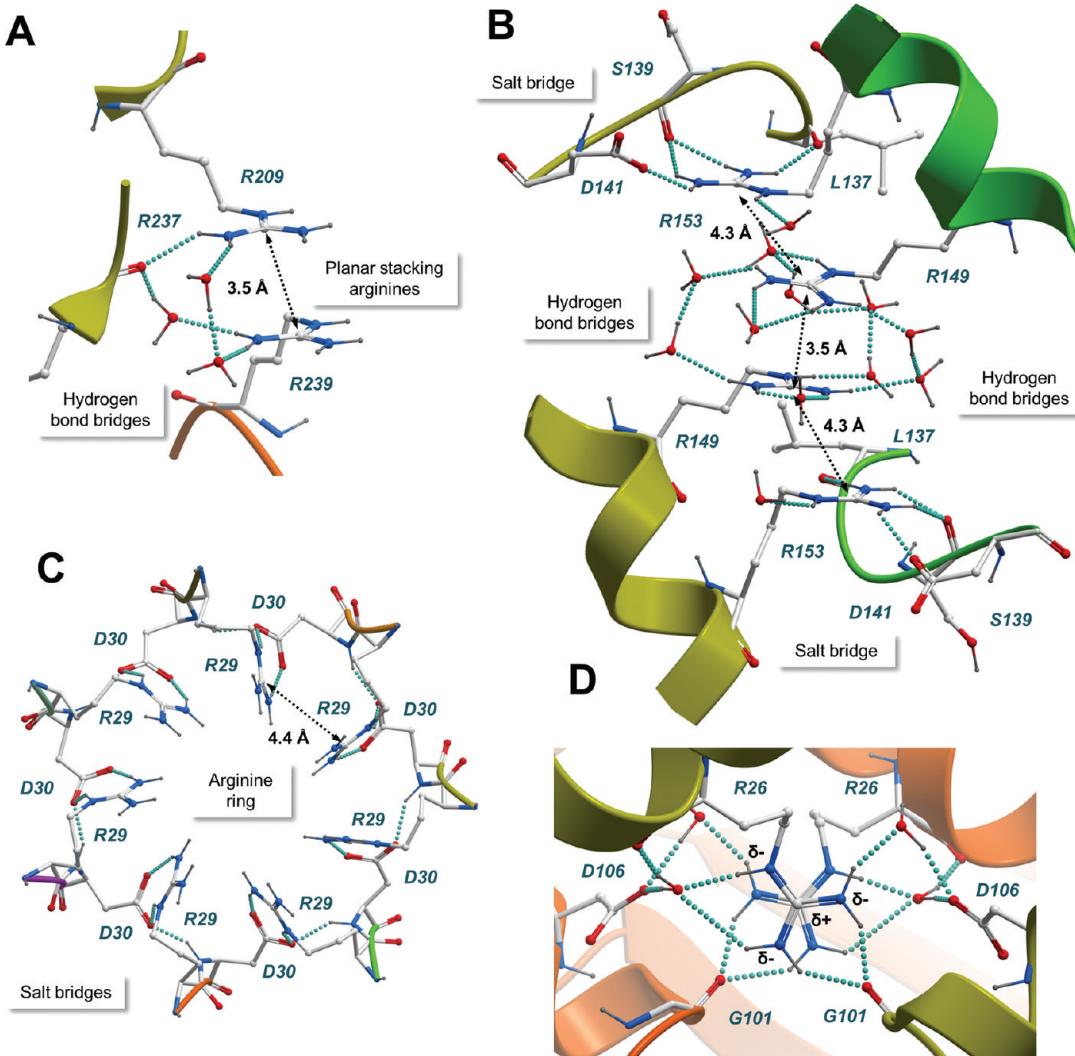


Figure 2. Atomic details of representative arginine clusters. (A) N-Acylamino acid racemase (PDB entry 1r0m); (B) ATP-dependent hexokinase (PDB entry 2e2q); (C) Sm protein (PDB entry 1i8f); (D) nucleoside diphosphate kinase (PDB entry 2hur). The protein structures are represented with a ribbon model and colored by subunit. Interacting residues and water molecules are shown with a “balls and sticks” model. Hydrogen bonds are represented with cyan spheres, and the C_z-C_z distances between arginine pairs are shown. Common stabilizing interactions such as planar stacking of arginines, counterions, and bridges of hydrogen bonds are shown.

channel that adopts distinct conformations in the open and closed state. Figure 1A illustrates a cluster of eight arginines stabilizing the closed state of a built-in lid of the *Methanococcus maripaludis* archaeal chaperonin.³⁴ The X-ray crystal structure shows a ring conformation with adjacent arginine pairs separated by an average C_z-C_z distance of 4.1 Å. This unusually large number of positive charges is stabilized by salt bridge interactions with an adjacent ring of negatively charged residues. Lid residues are highly conserved and play an important role in the chaperonin mechanism of protein folding. The arginine ring is disrupted in the open conformation in order to facilitate access to the folding chamber.

Archaeal proteasomes are another example of arginine rings in flexible protein domains. A highly conserved and essential YDR motif is involved in the gating mechanism that allows protein substrates to reach the inner chamber of the *Archaeoglobus fulgidus* 20 S proteasome.³⁵ As shown in Figure 1B, seven arginines with an average C_z-C_z distance of 4.5 Å interact at the top of the channel entrance. The arginine side chains point toward the outside of the proteasome chamber,

suggesting that hydrogen-bonding contacts with the solvent, as well as negatively charged counterions, might be the major stabilizing interactions.

Closed Shell Assembly. The building blocks for assembly of the HIV-1 capsid, a large closed shell that encapsidates the viral genome, are quasi-equivalent pentamers and hexamers of the viral CA protein.¹⁴ The capsid can be modeled as a fullerene-like cone assembled from ~250 hexamers and exactly 12 pentamers that are required to close the capsid. As shown in Figure 1F, six arginines with C_z-C_z distances <5 Å are found in the X-ray crystal structure of the hexamer.¹⁶ The arginine side chains are exposed to the solvent and could be stabilized by water molecules and/or charge-compensated by negatively charged counterions from either a permanent charge of Glu or Asp or a phosphorylated Ser, Thr, or Tyr. Other compensatory anions are also possible. An analogous ring of arginines resides at the center of the CA protein pentamers; however, in this case, the residues pack in a very small area, suggesting at least some electrostatic repulsion between individual subunits compared with the hexamer.¹⁵ Indeed, biochemical studies

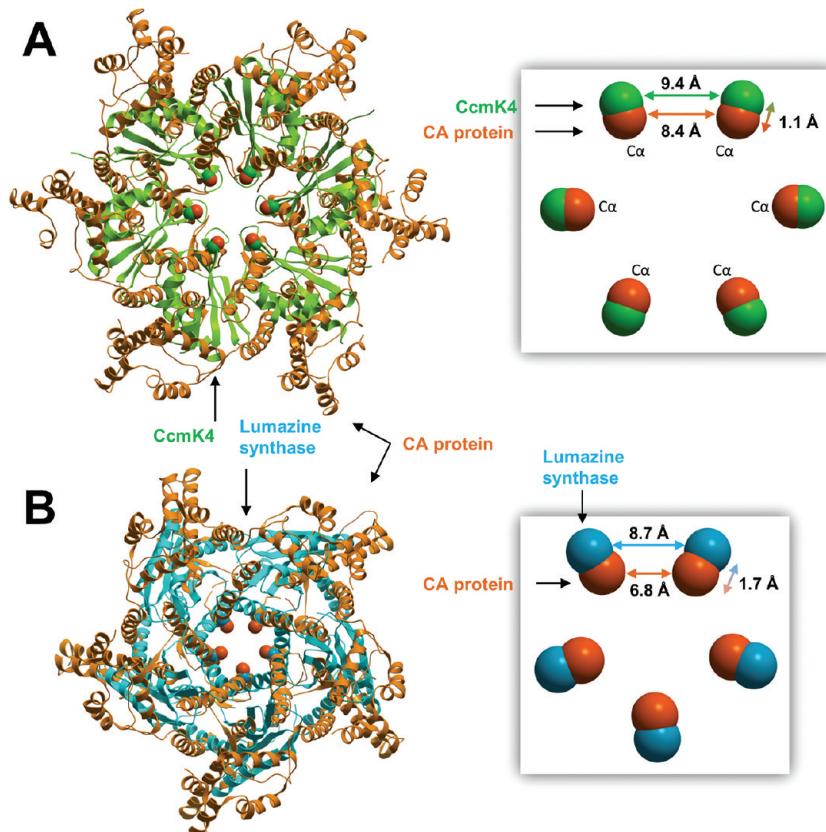


Figure 3. Three-dimensional superimposition of hexamers (A) and pentamers (B) assembled from the HIV CA protein (orange ribbon, PDB entries 3h4e and 3p05), carboxysome shell protein (green ribbon, PDB entry 2a10), and lumazine synthase (cyan ribbon, PDB entry 2f59). The proteins assemble into large closed-shell structures and share a cluster of positively charged arginines at the pentamer/hexamer ring centers. Hexamers and pentamers were superimposed using alpha carbons from the arginine rings, represented with a space filling model. C_{α} - C_{α} distances between adjacent arginines are shown.

confirm that the oligomeric assembly of hexamers is preferred over the pentamers.³⁶ Because the arginine residue is highly conserved in the HIV CA proteins from multiple strains, it has been proposed to function as an electrostatic switch, favoring the assembly of hexamers over pentamers.

Our analysis of the protein structure database identified other cases of arginine rings at the oligomerization interface of proteins involved in closed shell formation. Figure 1G shows the hexameric assembly of the carboxysome shell protein CcmK4, a basic building block for the construction of bacterial subcellular compartments involved in carbon fixation. A ring of six arginines with C_{ζ} - C_{ζ} distances <5 Å is found at the center of the hexamer.³⁷ Moreover, carboxysome shell proteins assemble as hexamers and pentamers, and the final carboxysome compartments are approximately icosahedral in shape.³⁸ The negative charge at the center of the ring is conserved among different family members, suggesting a similar electrostatic switch mechanism for assembly of hexamers and pentamers.

Another example found in the PDB database is lumazine synthase (Figure 1H), an enzyme involved in the biosynthesis of riboflavin. In contrast to the CA protein, pentamers of lumazine synthase are stable in solution, and the enzyme oligomerizes into typical pentameric rings that exist either in the free state or as different quaternary associations such as decamers (dimers of pentamers), icosahedrons (dodecamers of pentamers), or large closed shell capsids.³⁹

Hexamers of the CA protein were compared with similar types of rings assembled from the carboxysome shell protein, and the structures were superimposed by minimizing the root-mean-square distance between equivalent alpha carbons in the arginine rings. In a similar way, pentamers of the CA protein were compared with pentamers of lumazine synthase. As shown in Figure 3, the C_{α} - C_{α} distance between adjacent arginines in the CA protein hexamers (8.4 Å) is similar to the equivalent distances in the carboxysome shell protein (9.4 Å) and lumazine synthase (8.7 Å). By comparison, the C_{α} - C_{α} distance calculated from arginine pairs in pentamers of the CA protein is significantly lower (6.8 Å), suggesting that stronger electrostatic repulsion is involved.

A large cluster of arginine residues was also found in ferritin, an intracellular iron storage protein present in most living organisms. The protein assembles into spherical shells composed of 24 subunits with 2-, 3-, and 4-fold symmetry axes.⁴⁰ As shown in Figure 1D, large clusters of six arginine residues divided into two rings are found at the 3-fold symmetry axes, stabilized by layers of negatively charged residues.

Arginine Formations with Specific Known Biological Functions. Our analysis of the PDB database identified a number of cases where the positively charged arginine residues are involved in the molecular recognition of specific biological partners. Figure 1I illustrates the 3D crystal structure of a small nuclear ribonucleoprotein component (Sm protein) involved in pre-mRNA processing.⁴¹ A central ring of seven arginines forms

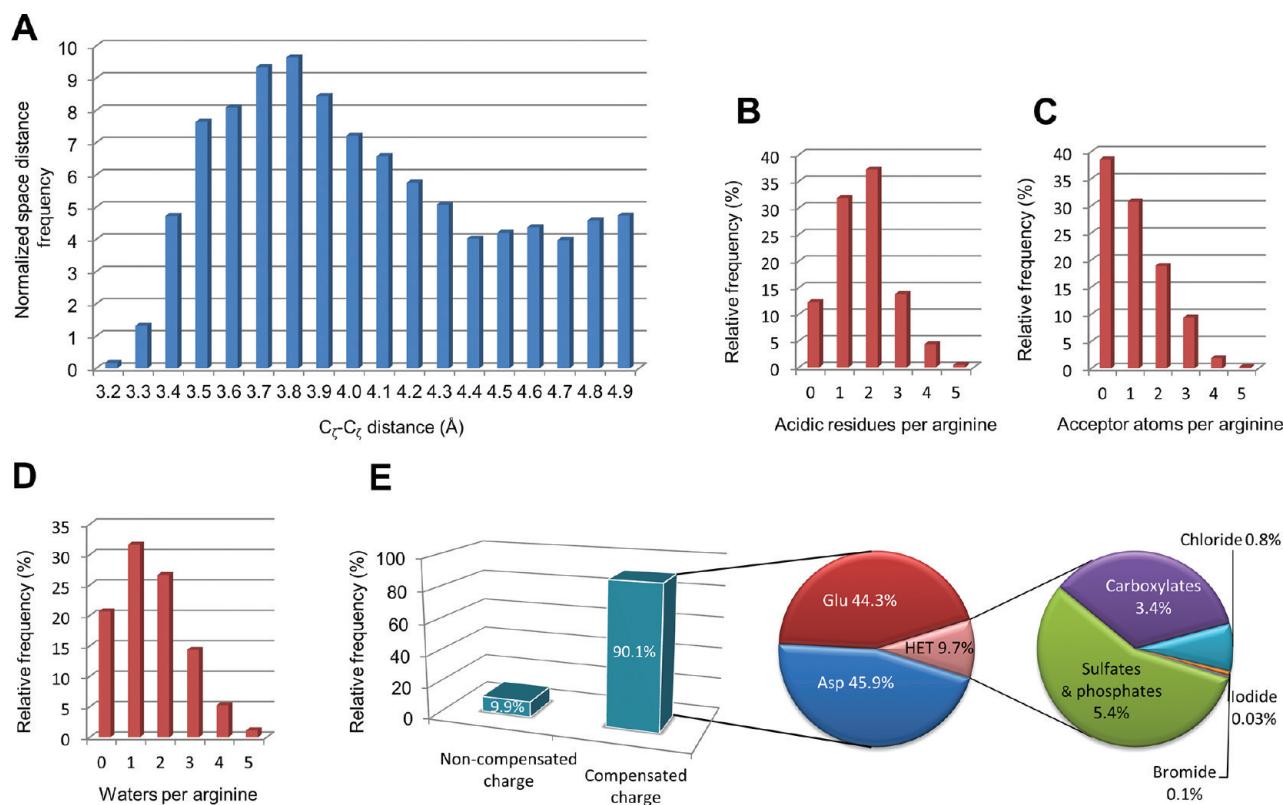


Figure 4. Property distribution of a subset of 13 218 high quality, density-supported, arginine clusters from the PDB database. (A) For each arginine, the C_z–C_z distance to the closest arginine residue was calculated, and the pairs were grouped into distance bins. The frequency was normalized to account for the cubic growth of the spherical volume. The number of arginines that falls into each bin was divided by the square of the distance. Counts for the number of aspartic and glutamic acid residues within a 6 Å distance cutoff (B), the number of hydrogen bond acceptor atoms (C), and crystallographic water molecules (D) within a 3.35 Å distance cutoff are provided. The inset E shows the fraction of arginines surrounded by negatively charged counterions. Common counterions are indicated.

a cationic pore that interacts specifically with single-stranded RNA. The ring is stabilized by salt bridges with an adjacent ring of negatively charged residues (Figure 2C).

Figure 1L shows a string of seven stacked arginines involved in the transport of phosphate anions.⁴² The arginines are stabilized by salt bridges with acidic residues and hydrogen bonds between the guanidinium groups, neighboring residues, and water molecules.

Stabilizing Interactions Derived from a Subset of High Quality Structures. A subset of high resolution arginine clusters with unambiguous fit to the electron density was further analyzed, and the following criteria were used for inclusion: (i) the arginine side chain conformations were unambiguously supported by the electron density, (ii) arginine residues with low occupancy atoms were excluded, and (iii) clusters with arginines from crystallographic neighbors were removed. Using these filters, 13 218 high quality clusters from 6427 PDB entries were found. The sequences were grouped into 2978 unique clusters with <90% identity between supporting protein sequences.

As shown in Figure 4A, the minimum distance between adjacent C_z–C_z pairs is 3.2 Å. This value is in good agreement with the stable 3.0 Å ion pair distance described by Scheraga and co-workers.¹³ Visual inspection revealed a consistent planar stacking of staggered guanidinium groups, exemplified in Figure 2D with an arginine pair at the interface of a nucleoside diphosphate kinase dimer (C_z–C_z distance = 3.2 Å).⁴³ Planar stacking avoids steric clashes between the guanidinium groups.

Close to 60° torsion angles along the N–C_z–C_z–N axis minimize the electrostatic repulsion by maximizing the distance between δ-nitrogens. Moreover, the electrostatic repulsion between positively charged arginines is perfectly balanced by the surrounding environment. As shown in Figure 2D, there are two counterions from aspartic acid residues, and all guanidinium donors are involved in hydrogen bonds with water molecules or surrounding residues. Planar stacking of guanidinium groups, previously described for lumazine synthase⁴⁴ and glutathione-S-transferase,⁴⁵ is commonly found in high quality structures from the PDB database. The C_z–C_z distance distribution shows a clear preference for arginine pairs separated by approximately 3.8 Å. Because there is less electrostatic repulsion at this distance, the relative orientation of stacked guanidinium groups is determined primarily by interactions with counterions and hydrogen bonds with the surrounding environment.

Statistical analysis revealed that 90.1% of the arginines are surrounded by negatively charged counterions (Figure 4E). As expected, ion pairs with carboxylate groups from aspartate and glutamate are found most frequently (87.7% of the arginines have at least one acidic amino acid within a 6 Å cutoff distance, Figure 4B). On the other hand, hetero molecules account for 9.7% of the ion-pairs (phosphates and sulfates 5.4%, carboxylates 3.4%, chlorides 0.8%, and other halogenated ions 0.1%, Figure 4E). Arginine clusters are involved in binding of phosphate-containing molecules such as nucleotides, NADP, FMN, FAD, and acetyl-CoA.^{46–48} Moreover, arginine clusters

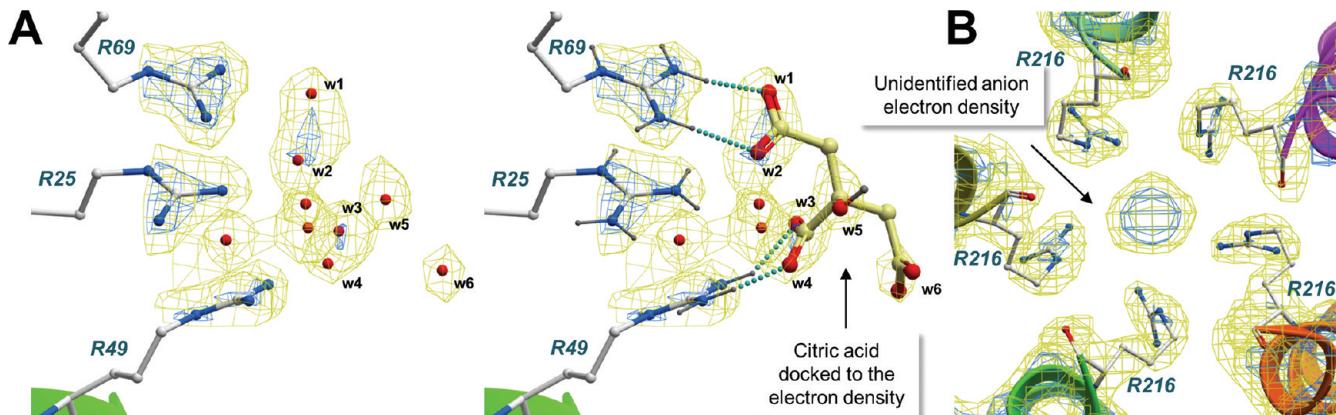


Figure 5. Ambiguous water molecule fit and unidentified anion electron density. In both cases, negatively charged molecules from the buffers used for protein purification and crystallization are likely involved in stabilizing salt bridge interactions. (A) A molecule of citric acid docked to an arginine triplet fits the electron density better than the original water molecules (PDB entry 1d3b). (B) The unidentified electron density could easily accommodate anions used in the buffer, such as, for example, a sulfate molecule (PDB entry 1wuq).

play a crucial role in stabilizing the propionic acid side chains of heme groups and bind frequently to organic acids and negatively charged bioactive compounds.^{49,50} Finally, we found a number of arginine clusters stabilized by chlorides, sulfates, phosphates, acids, and other molecules commonly present in the buffers used for protein purification and crystallization.

Noncompensated positive charge is found in 9.1% of the arginines in a cluster (Figure 4E). Guanidinium groups tend to be solvent exposed in these cases, or fit into a polar pocket with highly complementary physicochemical properties. The guanidinium cation acts as a strong hydrogen bond donor via the guanidine NH₂ and NH groups, with complementary hydrogen bond acceptor counterparts, such as the amide carbonyl of glutamine and asparagine, the hydroxyl groups of serine, threonine, and tyrosine, and the backbone carbonyl (Figure 4C). Crystallographic water molecules have an important stabilizing effect and are involved in hydrogen bonds or dipole–dipole interactions with 79% of the arginines (Figure 4D). However, because the electron density for placement of water oxygen atoms is often an ambiguous sphere of unknown origin, small anions such as chloride might be misinterpreted as water molecules or left uninterpreted. Figure 5A shows a cluster of three arginines surrounded by nine crystallographic water molecules in PDB entry 1d3b.⁵¹ While the electron density is unambiguous for the arginine side chains, water molecules are placed along a continuous duct of electron density, indicating that they could be replaced by a small molecular weight chemical used in the crystallization buffer. Docking results confirmed that citric acid fits into the electron density surrounding the arginine triplet, replacing six water molecules and establishing two stabilizing salt-bridges. Figure 5B shows a cluster of five arginines with a central lobe of unidentified density.⁵² Anions used in the crystallization media, such as sulfate, could easily fit into the electron density. Taken together, these results suggest that water molecules are important for the stability of positively charged arginines, but they might be overestimated in some cases due to ambiguous electron density and incorrect placement.

CONCLUSION

In this work, unusual formations of positively charged arginine residues were found in the PDB database, suggesting their

important role in protein–protein oligomerization and higher-order assembly, protein–ligand binding, and nucleic acid recognition. Controlling a countercharge to these clusters may be used as a biological regulatory mechanism. Despite the prevalence of short-range interactions between arginine pairs, higher-order clusters with up to eight guanidinium groups were found at protein oligomerization interfaces, as well as at sites involved in molecular recognition. The positive charge is properly balanced by negatively charged counterions in about 90% of the cases. Arginine clusters stabilize enzymatic cofactors and bind to nucleotides, glutamates, aspartates, and phosphorylated residues, and bioactive compounds bearing carboxylate, sulfate, and phosphate moieties. The arginines form rings, stacks, and strings of stacked side-chains. Ring formations are often located between layers of acidic amino acids, or at highly exposed sites easily accessible to water molecules and counterions. Arginine pairs adopt a common planar stacking conformation in part to avoid steric clashes. At the lower C_ζ–C_ζ distance limit, staggering of guanidinium groups helps minimize the electrostatic repulsion between negatively polarized nitrogens. The guanidinium groups are commonly involved in five hydrogen bonds with water molecules and acceptor groups from surrounding amino acids. Water molecules have a bridging effect on the arginine pairs, but in some cases, their electron density might be misinterpreted.

Taken together, our results suggest that clusters of interacting arginine residues play an important role in protein structure and function. They can be stabilized and/or regulated by multiple mechanisms such as stacking, hydrogen bonds, and electrostatics, as well as complementary side chains, backbone carbonyls, water molecules, anions or ligands with negatively charged groups, or hydrogen bond acceptors. Our results may be particularly relevant for the development of new scoring functions for protein–protein docking, identification of druggable protein sites, and the prediction of oligomerization interfaces.

AUTHOR INFORMATION

Corresponding Author

*E-mail: rabagyan@ucsd.edu (R.A.); yeager@virginia.edu (M.Y.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Irina Kufareva for valuable advice and some tools used in this study. The authors thank Owen Pornillos, Barbie Ganser-Pornillos, Fung-Yi Chan, Manuel Rueda, Winston Chen, Chayan Acharya, Fiona McRobb, Falgun Shah, and Chris Edwards for useful discussions and comments. Supported by postdoctoral grant SFRH/BPD/64216/2009 from the Fundação para a Ciência e a Tecnologia (FCT), Portugal (M.A.C.N.), and NIH grants R01 HL48908 (M.Y.), R01 GM066087 (M.Y.), R01 GM084545 (M.Y.), P50 GM082545 (Wesley Sundquist and M.Y.), R01 GM071872 (R.A.), U01 GM094612 (R.A.), and U54 GM094618 (R.A.).

REFERENCES

- (1) Calnan, B. J.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A. D. *Science* **1991**, *252*, 1167–1171.
- (2) Matsuo, M.; Huang, C.-H.; Huang, L. C. *Biochem. J.* **1978**, *173*, 441–447.
- (3) Williams, M. L.; Gready, J. E. *J. Comput. Chem.* **1989**, *10*, 35–54.
- (4) Tsai, C.-J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. *Protein Sci.* **1997**, *6*, 53–64.
- (5) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1983**, *168*, 867–885.
- (6) Burley, S. K.; Petsko, G. A. *FEBS Lett.* **1986**, *203*, 139–143.
- (7) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709–717.
- (8) Guillén Schlippe, Y. V.; Hedstrom, L. *Arch. Biochem. Biophys.* **2005**, *433*, 266–278.
- (9) Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1–9.
- (10) Ben-Shimon, A.; Eisenstein, M. *J. Mol. Biol.* **2010**, *402*, 259–277.
- (11) Bordner, A. J.; Abagyan, R. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 353–366.
- (12) Kufareva, I.; Budagyan, L.; Raush, E.; Totrov, M.; Abagyan, R. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 400–417.
- (13) Magalhaes, A.; Maigret, B.; Hofack, J.; Gomes, J. N. F.; Scheraga, H. A. *J. Protein Chem.* **1994**, *13*, 195–215.
- (14) Ganser-Pornillos, B. K.; Yeager, M.; Sundquist, W. I. *Curr. Opin. Struct. Biol.* **2008**, *18*, 203–217.
- (15) Pornillos, O.; Ganser-Pornillos, B. K.; Yeager, M. *Nature* **2011**, *469*, 424–427.
- (16) Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M. *Cell* **2009**, *137*, 1282–1292.
- (17) Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comput. Chem.* **1994**, *15*, 488–506.
- (18) Neves, M. A. C.; Totrov, M.; Abagyan, R. Submitted for publication, 2011.
- (19) Rose, P. W.; Beran, B.; Bi, C.; Bluhm, W. F.; Dimitropoulos, D.; Goodsell, D. S.; Prlić, A.; Quesada, M.; Quinn, G. B.; Westbrook, J. D.; Young, J.; Yukich, B.; Zardecki, C.; Berman, H. M.; Bourne, P. E. *Nucleic Acids Res.* **2011**, *39*, D392–D401.
- (20) Kleywegt, G. J.; Harris, M. R.; Zou, J.-Y.; Taylor, T. C.; Wählby, A.; Jones, T. A. *Acta Crystallogr., Sect. D* **2004**, *D60*, 2240–2249.
- (21) Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. *J. Med. Chem.* **2009**, *52*, 397–406.
- (22) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (23) Argos, P. *Protein Eng.* **1988**, *2*, 101–113.
- (24) Janin, J.; Miller, S.; Chothia, C. *J. Mol. Biol.* **1988**, *204*, 155–164.
- (25) Wang, W.-C.; Chiu, W.-C.; Hsu, S.-K.; Wu, C.-L.; Chen, C.-Y.; Liu, J.-S.; Hsu, W.-H. *J. Mol. Biol.* **2004**, *342*, 155–169.
- (26) Chiu, W.-C.; You, J.-Y.; Liu, J.-S.; Hsu, S.-K.; Hsu, W.-H.; Shih, C.-H.; Hwang, J.-K.; Wang, W.-C. *J. Mol. Biol.* **2006**, *359*, 741–753.
- (27) Maita, N.; Hatakeyama, K.; Okada, K.; Hakoshima, T. *J. Biol. Chem.* **2004**, *279*, 51534–51540.
- (28) Munch-Petersen, B.; Tyrsted, G.; Cloos, L. *J. Biol. Chem.* **1993**, *268*, 15621–15625.
- (29) Nishimasu, H.; Fushinobu, S.; Shoun, H.; Wakagi, T. *J. Bacteriol.* **2006**, *188*, 2014–2019.
- (30) Nishimasu, H.; Fushinobu, S.; Shoun, H.; Wakagi, T. *J. Biol. Chem.* **2007**, *282*, 9923–9931.
- (31) Welin, M.; Kosinska, U.; Mikkelsen, N.-E.; Carnrot, C.; Zhu, C.; Wang, L.; Eriksson, S.; Munch-Petersen, B.; Eklund, H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17970–17975.
- (32) Baumeister, W.; Walz, J.; Zühl, F.; Seemüller, E. *Cell* **1998**, *92*, 367–380.
- (33) Saibil, H. *Curr. Opin. Struct. Biol.* **2000**, *10*, 251–258.
- (34) Pereira, J. H.; Ralston, C. Y.; Douglas, N. R.; Meyer, D.; Knee, K. M.; Goulet, D. R.; King, J. A.; Frydman, J.; Adams, P. D. *J. Biol. Chem.* **2010**, *285*, 27958–27966.
- (35) Groll, M.; Brandstetter, H.; Bartunik, H.; Bourenkov, G.; Huber, R. *J. Mol. Biol.* **2003**, *327*, 75–83.
- (36) Ganser-Pornillos, B. K.; von Schwedler, U. K.; Stray, K. M.; Aiken, C.; Sundquist, W. I. *J. Virol.* **2004**, *78*, 2545–2552.
- (37) Kerfeld, C. A.; Sawaya, M. R.; Tanaka, S.; Nguyen, C. V.; Phillips, M.; Beeby, M.; Yeates, T. O. *Science* **2005**, *309*, 936–938.
- (38) Tanaka, S.; Kerfeld, C. A.; Sawaya, M. R.; Cai, F.; Heinhorst, S.; Cannon, G. C.; Yeates, T. O. *Science* **2008**, *319*, 1083–1086.
- (39) Ritsert, K.; Huber, R.; Turk, D.; Ladenstein, R.; Schmidt-Bäse, K.; Bacher, A. *J. Mol. Biol.* **1995**, *253*, 151–167.
- (40) Macedo, S.; Romão, C. V.; Mitchell, E.; Matias, P. M.; Liu, M. Y.; Xavier, A. V.; LeGall, J.; Teixeira, M.; Lindley, P.; Carrondo, M. A. *Nat. Struct. Biol.* **2003**, *10*, 285–290.
- (41) Mura, C.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5532–5537.
- (42) Moraes, T. F.; Bains, M.; Hancock, R. E. W.; Strynadka, N. C. J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 85–87.
- (43) Moynié, L.; Giraud, M.-F.; Georgescauld, F.; Lascau, I.; Dautant, A. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 755–765.
- (44) Zhang, X.; Meining, W.; Fischer, M.; Bacher, A.; Ladenstein, R. *J. Mol. Biol.* **2001**, *306*, 1099–1114.
- (45) Xiao, B.; Singh, S. P.; Nanduri, B.; Awasthi, Y. C.; Zimniak, P.; Ji, X. *Biochemistry* **1999**, *38*, 11887–11894.
- (46) Meissner, B.; Schleicher, E.; Weber, S.; Essen, L.-O. *J. Biol. Chem.* **2007**, *282*, 33142–33154.
- (47) Pitcher, R. S.; Brissett, N. C.; Picher, A. J.; Andrade, P.; Juarez, R.; Thompson, D.; Fox, G. C.; Blanco, L.; Doherty, A. J. *J. Mol. Biol.* **2007**, *366*, 391–405.
- (48) Dall'Aglio, P.; Arthur, C. J.; Williams, C.; Vasilakis, K.; Maple, H. J.; Crosby, J.; Crump, M. P.; Hadfield, A. T. *Biochemistry* **2011**, *50*, 5704–5717.
- (49) Erlanson, D. A.; Lam, J. W.; Wiesmann, C.; Luong, T. N.; Simmons, R. L.; DeLano, W. L.; Choong, I. C.; Burdett, M. T.; Flanagan, W. M.; Lee, D.; Gordon, E. M.; O'Brien, T. *Nat. Biotechnol.* **2003**, *21*, 308–314.
- (50) Riise, E. K.; Lorentzen, M. S.; Helland, R.; Smalås, A. O.; Leiros, H.-K. S.; Willlassen, N. P. *Acta Crystallogr., Sect. D* **2007**, *63*, 135–148.
- (51) Kambach, C.; Walke, S.; Young, R.; Avis, J. M.; de la Fortelle, E.; Raker, V. A.; Lührmann, R.; Li, J.; Nagai, K. *Cell* **1999**, *96*, 375–387.
- (52) Tanaka, Y.; Nakagawa, N.; Kuramitsu, S.; Yokoyama, S.; Masui, R. *J. Biochem.* **2005**, *138*, 263–275.
- (53) Mura, C.; Kozhukhovsky, A.; Gingery, M.; Phillips, M.; Eisenberg, D. *Protein Sci.* **2003**, *12*, 832–847.
- (54) Klinke, S.; Zylberman, V.; Bonomi, H. R.; Haase, I.; Guimarães, B. G.; Braden, B. C.; Bacher, A.; Fischer, M.; Goldbaum, F. A. *J. Mol. Biol.* **2007**, *373*, 664–680.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on April 19, 2012. Table 1 and the Table of Contents/abstract graphic were updated. The revised paper was reposted on April 30, 2012.