Acetylation of Surface Lysine Groups of a Protein Alters the Organization and Composition of Its Crystal Contacts

Kyungtae Kang,[†,§,⊥] Jeong-Mo Choi,[†,⊥] Jerome M. Fox,[†] Phillip W. Snyder,[†] Demetri T. Moustakas,[†] and George M. Whitesides*,[†,‡]

[†]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

[‡]Wyss Institute of Biologically Inspired Engineering, Harvard University, 60 Oxford Street, Cambridge, Massachusetts 02138, United States

[§]Department of Applied Chemistry, Kyung Hee University, 1732 Deogyeong-daero, Giheung, Yongin, Gyeonggi 17104, Republic of Korea

ⓢ Supporting Information

ABSTRACT: This paper uses crystals of bovine carbonic anhydrase (CA) and its acetylated variant to examine (i) how a large negative formal charge can be accommodated in protein−protein interfaces, (ii) why lysine residues are often excluded from them, and (iii) how changes in the surface charge of a protein can alter the structure and organization of protein−protein interfaces. It demonstrates that acetylation of lysine residues on the surface of CA increases the participation of polar residues (particularly acetylated lysine) in protein−protein interfaces, and decreases the participation of nonpolar residues in those interfaces. Negatively charged residues are accommodated in protein−protein interfaces via (i) hydrogen bonds or van der Waals interactions with polar residues or (ii) salt bridges with other charged residues. The participation of acetylated lysine in protein−protein interfaces suggests that unacetylated lysine tends to be excluded from interfaces because of its positive charge, and not because of a loss in conformational entropy. Results also indicate that crystal contacts in acetylated CA become less constrained geometrically and, as a result, more closely packed (i.e., more tightly clustered spatially) than those of native CA. This study demonstrates a physical-organic approach—and a well-defined model system—for studying the role of charges in protein−protein interactions.

$[CA-(NH_3^+)_{18}]^{-3.4}$
$[CA-(NHCOCH_3)_{18}]^{-19}$
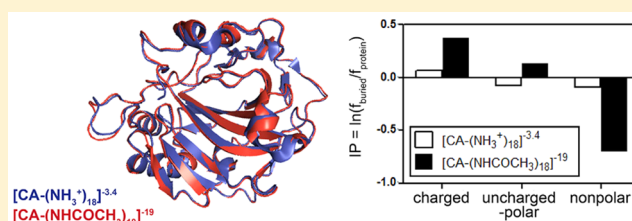
INTRODUCTION

This study uses the contact regions between molecules of bovine carbonic anhydrase (CA; E.C.4.2.1.1) in crystals of this protein as a physical-organic model system with which to study protein−protein interfaces systematically. Four characteristics of the contact regions of crystals make them attractive as a subject for study: (i) They are well-defined structurally by X-ray and neutron crystallography. (ii) They can be systematically modified through mutagenesis and/or chemical functionalization; these modifications often retain native protein structure. (iii) Contact regions generate and define adjacent noncontact regions—regions that provide a set of examples of near-surface films or pools of water that might aid in understanding the role of water in protein−protein association. (iv) A very large number of crystallographic data (from the Protein Data Bank) on protein−protein interfaces are now available. Understanding interactions between the surfaces of proteins in water is, more generally, important for understanding how molecular recognition controls the assembly of functional (and dysfunctional) multiprotein aggregates inside the cell.

CA is a representative and experimentally tractable—although in some respects atypical—model protein.[1−9] It has three characteristics that make it uniquely suited for carrying out a physical-organic examination of protein interfaces: (i) It is structurally rigid; alterations to its surface, thus, tend not to alter its conformation.[10,11] (ii) All 18 lysine residues on the surface of CA can be acetylated. This acetylation results in a large increase in its negative charge (from $Z = -3.4$ for nonacetylated CA, to $Z = -19$ for peracetylated CA;[12] X-ray crystallography and circular dichroism spectroscopy indicate that this peracetylation does not alter the secondary or tertiary structure of CA).[13,14] (iii) Previous studies demonstrate that CA and many of its mutants can be crystallized.[15−21]

Because the structural attributes of CA are unique, the conclusions of this study apply specifically to this protein, and are not, in the absence of examinations of other proteins, general statements about protein−protein interfaces. This caveat not withstanding, CA provides an exceptionally well-defined model system with which to study these interfaces in order to examine biomolecular recognition events involving the

surfaces of proteins outside of the active site. Pursuing a physical-organic approach, we used X-ray crystallography to collect structures of CA with and without its positively charged, surface-exposed lysine residues acetylated; and using these structures, we examined how acetylation altered the organization and composition of crystal contacts.

Background. Protein−protein association plays a central role in a range of biological processes, from the assembly of proteasomes and viral capsids to the binding of antibodies to antigens.[22−24] These processes involve highly tuned structural motifs and/or constellations of amino acids that bind one another at specific positions (and with specific orientations). These protein−protein interactions are unquestionably important in biochemistry, but they are less well understood—as class of biomolecular recognition events—than are ligand-protein interactions. The heterogeneity and (generally) uncharacterizable topography of protein−protein interfaces[25−27] make studies of the influence of systematic perturbations challenging; as a result, their biophysical rationalization, prediction, and synthetic reconstruction remain exceedingly difficult.

Lysine is a particularly interesting residue to examine in the context of protein−protein association. Although it is generally excluded from the interfaces of multiprotein complexes, it is occasionally found in crystallographic interfaces.[28−30] By examining the protein−protein interfaces that form with and without positively charged lysine (and, correspondingly, without and with electrically neutral acetylated lysine), we examined how the charge and chemical functionality of lysine groups affects the position and chemical composition of crystal contacts (i.e., regions on the surface of a protein that are in contact with regions on the surfaces of neighboring proteins).

Questions. This study addresses three questions pertaining to the influence of surface charge on the structure and organization of protein−protein interfaces: (i) Can a protein with a large formal charge (i.e., $Z = -19$) be crystallized and, if so, how is that charge accommodated and/or compensated to enable stable protein−protein interfaces? (ii) What attributes of lysine residues (e.g., flexibility or charge) cause their often-observed exclusion from protein−protein interfaces? (iii) How does the conversion of a surface of a protein from "polyzwitterionic" to "polyanionic" alter the structure and organization of interfaces that result when the protein is crystallized?

Experimental Design. We used acetic anhydride to acetylate the lysine residues of CA (this procedure is quantitative; see Supporting Information); we refer to crystals of the peracetylated variant as $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$. Capillary electrophoresis confirmed that all 18 residues were acetylated (Figure S1);[31] X-ray crystallography allowed us to determine its structure, and, thus, permitted a comparison with the structure of wild-type CA,[17] which we, hereafter, refer to as $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$. This comparison allowed us to study protein−protein interactions in which the structures of the contacting proteins is very similar, but in which the number of surface-exposed NH_3^+ groups differs by 18, and in which the net charge changes from close to zero ($Z = -3.4$) to highly negatively charged ($Z = -19$) (The small discrepancy between the number of lysine groups acetylated and the change in net charge in solution indicates that the removal of positive charges by peracetylation is partially compensated, possibly by changes in the pK_a of carboxyl groups on the protein surface).[3,4] The increase in net charge that results from acetylation—an increase

that occurs through a decrease in the total number of positive charges—is accompanied by a net decrease (by 18 NH_3^+ groups) in the net surface density of charged groups, and also by a change in the surface of the protein from more "polyzwitterionic" to more "polyanionic" upon acetylation.

■ RESULTS AND DISCUSSION

Can a Protein with a Large Formal Charge (i.e., $Z = -19$) be Crystallized? *Peracetylated CA can be crystallized, but with a different space group than native CA.* We could successfully crystallize both $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$ and $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$, although under slightly different conditions: $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$, 2.4 M ammonium sulfate in 50 mM Tris-HCl, pH = 7.5, and a crystallization period of 2 weeks; and $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$, 1.6 M ammonium sulfate in 50 mM Tris-HCl, pH = 7.0, and a crystallization period of 3−6 months; see Supporting Information. X-ray crystallography shows that $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$ and $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$ have indistinguishable secondary and tertiary structures (the root-mean-square deviation between the heavy atoms of their aligned structures is 0.285 Å, and much of this difference originates from amino acids on the surface of the protein; Figure 1a), but
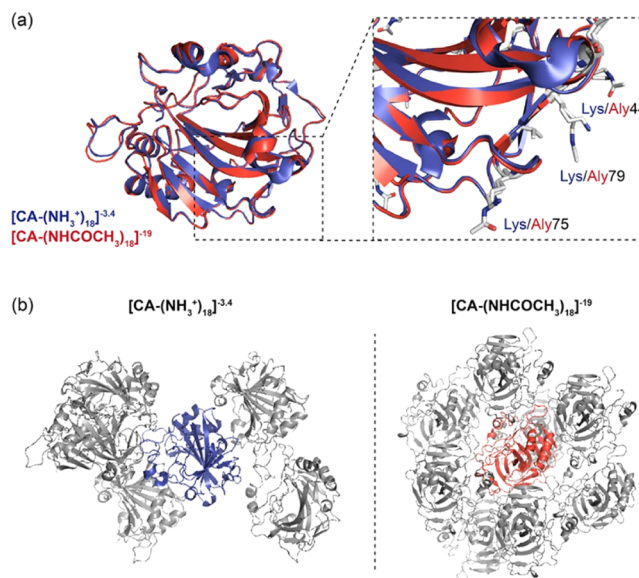

Figure 1. (a) Backbone alignment of $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$ (blue) and $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$ (red). (b) Representation of a unit cell: a central protein (colored) surrounded by its symmetry mates (gray): $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$ (left) and $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$ (right). We considered two proteins to be in contact when there was buried surface area between them.

indicates that crystals of these two proteins have different space groups: $P6_1$ for $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$ (PDBID: 1V9E) and $P2_12_12$ for $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$ (PDBID: 5EZT). Thus, the crystallographic interfaces between adjacent proteins in each structure are entirely different (Figure 1b). In the crystal of $[\mathrm{CA\text{-}(NH_3^+)_{18}}]^{-3.4}$, each molecule of CA contacts five adjacent symmetry mates; in the crystal of $[\mathrm{CA\text{-}(NHCOCH_3)_{18}}]^{-19}$, each molecule contacts eight symmetry mates.

In Crystals of Proteins with Large Formal Charge, How is the Charge Accommodated and/or Compensated to Enable Stable Protein−Protein Interfaces? *Peracetylated CA (the more negatively charged protein) buries more surface area in contact regions than does native CA.* When

proteins crystallize, they bury solvent exposed surface area in contact regions; we calculated the buried surface area (BSA) in crystallographic interfaces using eq 1, where $ASA_{solution}$ is the accessible surface area—the area that can be contacted by a modeled solvent probe (see Supporting Information)—of a single protein in solution,

$$BSA = ASA_{solution} - ASA_{crystal} \qquad (1)$$

and $ASA_{crystal}$ is the accessible surface area of a protein in contact with its crystallographic symmetry mates (we calculated ASA by using the Shrake-Rupley algorithm;[32] see Supporting Information). BSA increased by 1083 Å^2 upon acetylation; this increase indicates that the crystal of $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ buries more surface area—and thus has a smaller crystal void volume—than does $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ (Table 1). This result is unexpected: $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ is a more highly charged protein, but forms a more compact crystal.

Table 1. Values of $ASA_{solution}$ and BSA for Proteins^a

	$[CA\text{-}(NH_3^+)_{18}]^{-3.4}$	$[CA\text{-}(NHCOCH_3)_{18}]^{-19}$
$ASA_{solution}$	11400	12087
BSA	1678	2761

^{a}The unit is Å^2.

Acetylation increases the ratio of charged to uncharged, polar to nonpolar, and polar (charged) to polar (neutral) groups in buried interfaces. To investigate the influence of acetylation on the amino acid composition of buried surface area, we defined the interface propensity (IP; eq 2),[33] a metric for the tendency of an amino acid to be present in contact regions. This metric is uniquely useful with CA

$$IP_i = \ln \frac{f_{buried,i}}{f_{protein,i}} \qquad (2)$$

because its tertiary structure—and, thus, the composition of solvent-exposed amino acids—does not change after acetylation. In eq 2, $f_{protein,i}$ is the fraction of surface-exposed residues of amino acid i (e.g., i = Lys for lysine residues), and $f_{buried,i}$ is the fraction of buried residues (i.e., residues that partially or entirely participate in interfaces) that are residues of amino acid i (Table S2). To examine general trends in IP values, we classified the most common amino acids into three groups—charged (and thus polar: Lys, Arg, His, Asp, and Glu), uncharged and polar (Ser, Asn, Gln, Thr, and acetylated Lys, to which we assign the three-letter code Aly), and nonpolar (Val, Ile, Leu, Trp, and Phe)—and we examined the influence of acetylation on the summated IP value of each group. (The numbers of buried residues for each group in the crystals of both proteins appear in Table 2 and Table S2 in the SI). Figure 2a, which shows the results of these calculations, indicates that acetylation increases the propensity of polar—both uncharged and, remarkably, charged—residues to participate in contact

Table 2. Number of Buried Residues for Each Group in the Crystals of Each Protein

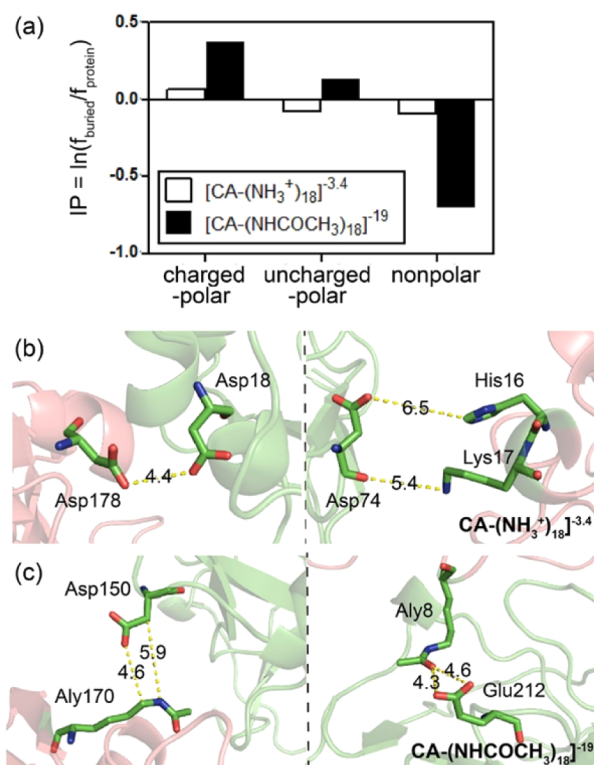|	$[CA\text{-}(NH_3^+)_{18}]^{-3.4}$	$[CA\text{-}(NHCOCH_3)_{18}]^{-19}$
charged-polar	20	29
uncharged-polar	16	34
nonpolar	26	29

Figure 2. (a) The values of interface propensity—a metric for the tendency of amino acids to participate in interfaces—for three categories of residues (charged: Lys, Arg, His, Asp, and Glu; uncharged-polar: Ser, Asn, Gln, Thr, and Aly; and nonpolar: Val, Ile, Leu, Trp, and Phe), calculated for $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$; acetylation increases the propensity of charged residues to participate in interfaces and decreases the propensity of hydrophobic residues to do so. (b,c) Examples of different types of crystallographic interfaces in (b) $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and (c) $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ that involve Asp and Glu residues. Highlighted interactions (dashed lines with corresponding measurements), which are longer than average distances observed for hydrogen bonds or salt bridges, may reflect long-range van der Waals type interactions or water-mediated hydrogen bonds. The numbers with dotted lines denote the distances between neighboring residues. Colors denote atoms as follows: carbon (green), oxygen (red), and nitrogen (blue).

regions, while decreasing the propensity of nonpolar residues to do so.

Contact regions accommodate charged residues through salt bridges and/or hydrogen bonds. The increased participation of charged residues in the crystal contacts of $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$—where proteins have a larger net charge than unacetylated CA, but fewer surface-exposed charged residues—is counterintuitive. To examine how charged residues might be differentially accommodated in the crystal contacts of $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$, we examined their interaction partners in each crystal. In the crystal contacts of $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$, most of the buried, negatively charged Asp and Glu residues (9 of 11) interact either with positively charged Lys or His residues or with negatively charged Asp residues (Figure 2b). In $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$, on the other hand, approximately half (9 of 18) of the buried Asp and Glu residues engage in similar interactions with charged amino acids, while the other half associate with polar residues, notably (and predominantly) Aly, which participates in six out of eight of these interactions (Figure 2c).

Salt bridges between Asp or Glu and His or Arg are common between (and within) proteins. Interactions between Asp and/or Glu residues, which are less common, are often stabilized by hydrogen bonds.[34,35] In $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$, however, these interactions involve side chains separated by 4−5 Å (which exceeds the 1.5−2.5 Å of a typical hydrogen bond) and are, thus, probably mediated by long-range van der Waals interactions (perhaps with involvement of water molecules), and with protons at the interface neutralizing negative charges.[36,37] Interactions between Asp or Glu and polar residues (notably Aly) seem, by contrast, usually to be mediated by hydrogen bonds—either direct (between side chains) or indirect (through molecules of water between those side chains)—and van der Waals forces. Accordingly, Aly residues—perhaps, because of their conformational flexibility (and prevalence)—may be more likely to participate in such interactions and, thus, be better able to facilitate the burial of (i.e., to associate with) Asp and Glu at interfaces.

What Attributes of Lysine Residues (e.g., Flexibility or Charge) Cause their Often-Observed Exclusion from Protein−Protein Interfaces? *Aly is more common in contact regions of $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ than is Lys in contact regions of $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$.* Aly residues in $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ contribute much more to BSA (691 Å²) than do the Lys residues in $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ (173 Å²). The difference in contributions (518 Å²) suggests that Aly—perhaps as a result of its charge neutrality, increased surface area, and/or ambidexterity in hydrogen bonding—can more easily accommodate a range of binding partners in crystal contacts.

Aly is responsible for the increased participation of uncharged polar residues in the interfaces of $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$. (That is, if the IP value for polar residues did not include contributions from Aly, it would be similar between $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$. At interfaces, Aly interacts with both negatively charged residues, such as Glu and Asp (of which there are six), and with other uncharged-polar residues (predominantly Aly) and nonpolar residues (Leu and Tyr) —via hydrogen bonds and van der Waals forces. The size and conformational flexibility of Aly may, again, make it more likely—and more able—(than other amino acids) to engage in such interactions, suggesting that the charge of Lys—not its conformational flexibility—reduces its propensity to participate in the contact regions of proteins.

How Does the Conversion of a Surface of a Protein from "Polyzwitterionic" to "Polyanionic" Alter the Structure and Organization of Interfaces that Result When the Protein is Crystallized? *Acetylation has little influence on the size and geometry of crystal contacts, but reduces their charge complementarity.* To investigate the influence of acetylation on the configuration and chemical composition of crystal contacts, we analyzed individual crystallographic interfaces (and the crystal contacts of which they are composed). An atom whose BSA value is larger than zero was considered to be in a crystal contact. For $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$, there are three crystallographic interfaces (which incorporate all five crystal contacts): two heterogeneous interfaces (A1−A2 and C1−C2) and a homogeneous interface (B−B). Similarly, for $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$, there are five crystallographic interfaces: three heterogeneous (A1−A2, D1−D2, and E1−E2) and two homogeneous (B−B and C−C; Figure 3a and Table S3). The crystal contacts from the two crystals have similar morphology. Their average gap volume index (the ratio of the surface area of an interface to the volume
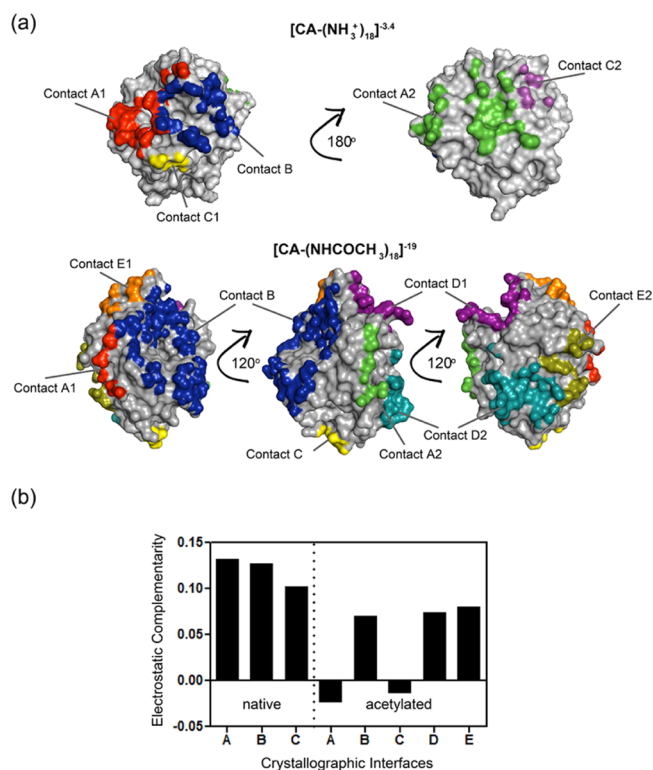

Figure 3. (a) Illustrations of the crystal contacts for $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$. Different colors denote different contact regions. (b) Values of the electrostatic complementarity (EC) of various interfaces in the two crystal systems. A negative EC value means that the averaged correlation between the electrostatic surface potentials of two contacting atoms at a crystal interface is positive; they tend to have the same sign. If charges are randomly distributed, and there is no correlation between the surface potentials of two surfaces, the EC value would be zero. This plot shows that the charge complementarity of interfaces at crystal contacts decreases significantly upon acetylation.

of a cavity between the surfaces that comprise that interface—a measure of shape complementarity) and size (BSA$_{avg}$ = 336 Å² for $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$; BSA$_{avg}$ = 345 Å² for $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$, respectively) are similar for the two crystals (See Supporting Information for full analysis).

The charge complementarity between opposing crystal contacts within each contact region—that is, the tendency of opposing interfacial charges to be complementary to one another—however, differ between $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$. We calculated the electrostatic complementarity (EC) for each crystallographic interface; EC is a metric for the correlation between (i) electrostatic surface potentials in a crystallographic interface and (ii) electrostatic surface potentials that account only for charges induced on one surface by the other (see Supporting Information). Figure 3b shows the results of our calculations; the EC values for the interfaces of $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ are uniformly more positive and larger (0.10−0.13) than the EC values for the interfaces of $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ (−0.02−0.08). These results suggest that Coulombic attraction (the origin of charge complementary) may guide the formation of crystal contacts in wild-type CA, but plays much less of a role in the formation of contacts in acetylated CA, where there are fewer positively charged residues (specifically, half of the 36 available in native CA) available to complement opposing negatively charges.

The crystal contacts of $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ *are more condensed than those of* $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$. We hypothesized that crystal contacts, without the constraint of charge complementarity, might optimize the spatial density of their constituent atoms: That is, atoms involved in any particular contact might appear in more distinct, geometrically well-defined patches (rather than appearing in diffuse patches that permit opposing charges to interact). To test this hypothesis with $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ and $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$, we used a hierarchical clustering algorithm to group atoms involved in crystal contacts by their proximity (Figure 4a and b; see Supporting Information). The idea: If a contacting region is spatially distinct, the algorithm will group its constituent atoms together; if not, it will divide them into multiple groups. For each crystal, we limited the number of groups to the number of known crystal contacts (five for $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and eight for $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$). To test the performance of the clustering algorithm, we calculated S (the Jaccard index, eq 3), a metric for the overlap between group C (the algorithm-derived grouping of contacting atoms) and group I (the contacting atoms of a known crystal contact).

$$S(C, I) = n(C \cap I)/n(C \cup I) \qquad (3)$$

In eq 3, n ($C \cap I$) represents the number of overlapping atoms in sets C and I (i.e., the atoms present in the both sets), and $n(C \cap I)$ represents the number of total atoms (overlapping and nonoverlapping) in sets C and I. If cluster C overlaps perfectly with crystal contact I, $S = 1$; if there is no overlap, $S = 0$. Figure 4c shows values of S for every combination of group C and crystal contact I. The highest value of each column (S_c) represents the best match between a specific cluster and a crystal contact; the highest value of each row (S_r) represents the best match between a specific crystal contact and a cluster. To determine representative performance indices for our clustering algorithm, we calculated $\overline{S}_c$ and $\overline{S}_r$ (the average values of S_c and S_r for the contact regions) for each crystal. The upper bound of $\overline{S}_c$ and $\overline{S}_r$ is 1; the lower bound is $1/n$, where n is the total number of crystal contacts (see SI for proof). These indices (Figure 4c) suggest that the clustering algorithm is better able to identify atoms in the interfaces of $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ ($\overline{S}_c = 0.55$ and $\overline{S}_r = 0.57$) than it is able to identify atoms in the interfaces of $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ ($\overline{S}_c = 0.44$ and $\overline{S}_r = 0.43$). This result suggests that crystal contacts, when less constrained by the requirement of charge complementarity, tend to appear in more distinct patches (i.e., they optimize the proximity of their constituent atoms).

The influences of acetylation on the charge complementarity and organization of crystal contacts are supported by a third structure. While we were conducting this study, another structure of BCA (5A25) was added to the Protein Data Bank. To ask if our observations are, in fact, the result of different crystallization conditions/protocols, we analyzed this structure in the same way that we analyzed the other two structures in our study. Structure 5A25 has a different space group than 1V9E (the other unacetylated structure), and a higher BSA value (2631 Å^2; this is although still lower than that of acetylated BCA.), despite very similar crystallization conditions. Nonetheless, the organization of the crystallographic interfaces of 5A25 are surprisingly similar to those of 1V9E (Figure S3): the average EC values of the crystallographic interfaces of 5A25 and 1V9E are 0.11 and 0.12, respectively—both far from the EC value of acetylated BCA (0.04). In addition, a hierarchical clustering analysis of 5A25 yields values of $\overline{S}_c$ and $\overline{S}_r$ of 0.43 and 0.42,
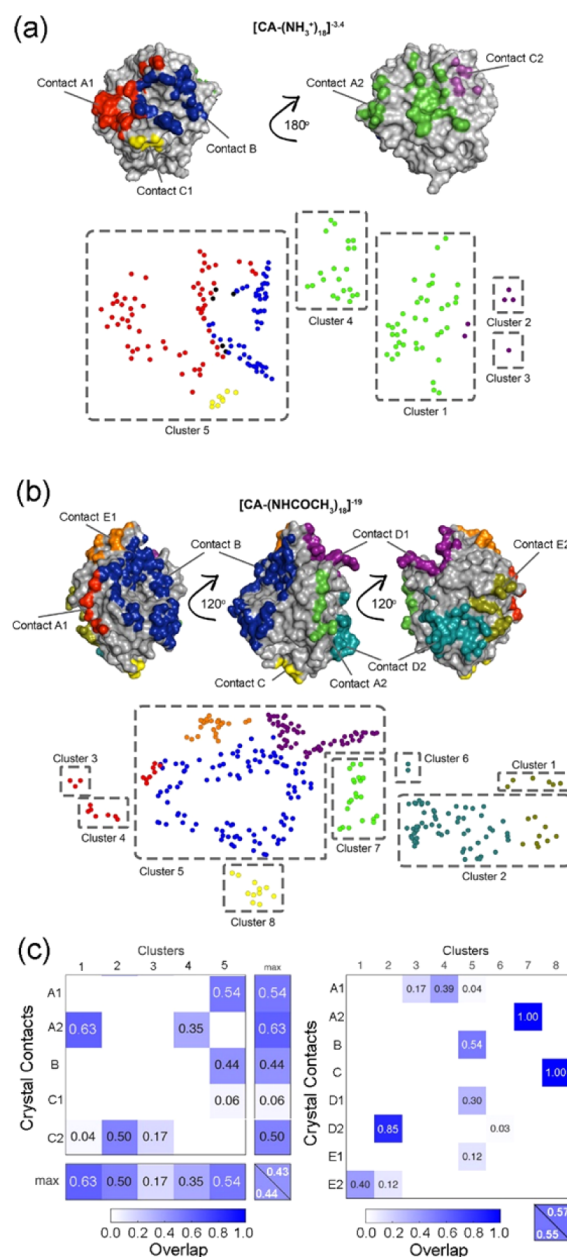

Figure 4. (a) 2-dimensional topological maps of interface atoms, the atoms with a nonzero BSA value, for $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ and (b) $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ with denotations of the actual crystal contact (color) and artificially generated clusters (rectangles of dotted lines). For the generation of the maps, we first projected a 3-dimensional atomic coordinate of each contact onto its principal surface, and then arranged those maps for each contact in a single plane. The distances and orientations of the maps between different crystal contacts are arbitrarily determined for clear visualization. (c) Overlap matrix for $[CA\text{-}(NH_3^+)_{18}]^{-3.4}$ (left) and that for $[CA\text{-}(NHCOCH_3)_{18}]^{-19}$ (right) where the maximal values for each row (S_r) and column (S_c) are shown at the right and bottom of each matrix. The lower right cell averages maximum values of each row and each column. Shades of blue indicate the extent of overlap S_{IC} between crystal contact I and cluster C.

respectively; these values are very similar to those of 1V9E (0.44 and 0.43, respectively). Again, both sets of values differ significantly from those of acetylated BCA ($\overline{S}_c = 0.55$ and $\overline{S}_r = 0.57$). Accordingly, all three structures examined in this study are compatible with our inferences concerning the influence of

acetylation on the charge complementarity and organization of crystal contacts.

CONCLUSIONS

This study uses CA as a physical-organic model system with which to study the influence of surface charge on the structure and composition of crystallographic interfaces. It shows that peracetylated CA—a protein with a net formal charge of −19 at pH 8.4; an empirical value established in solution by capillary electrophoresis—is capable of assembling into a stable crystal. By comparing the crystallographic interfaces of acetylated and nonacetylated crystals, we determined how acetylation causes the composition and organization of those interfaces to change. Acetylation of lysine residues has three significant effects on crystallographic interfaces: (i) It increases the participation of charged and uncharged polar residues—and decreases the participation of nonpolar residues—in crystal contacts (relative to the contacts of nonacetylated CA). This shift likely results from van der Waals and/or hydrogen bonding interactions between negatively charged residues and Aly, which seems to play an important role in accommodating the burial of negative charge; (ii) Aly participates much more in crystal contacts than does Lys; this difference results in the increase in total buried surface area in crystal contacts upon acetylation. (iii) The crystal contacts of peracetylated CA become less constrained geometrically (perhaps, as a result, of their decreased dependence on charge complementarity) and, as a result, more closely packed (i.e., more spatially distinct) than those of native CA.

We note: The protein−protein interactions that give rise to crystals are generally weaker than—and involve interfaces that are morphologically distinct[38] from—the protein−protein interactions that permit biological function. Crystals of proteins therefore normally form *in vitro* at concentrations of protein that are much higher than those found inside the cell. However, for both types of interactions—the protein−protein interactions that occur *in vivo* and those that result exclusively from protein crystallization—the forces that cause assembly—forces reflecting the hydrophobic effect, Coulombic attraction, hydrogen bonding, and van der Waals interactions—are the same.[38−41] In studying how alterations to the chemical functionality of protein surfaces influence interactions between proteins in crystals, we can, thus, begin to understand how similar alterations might influence interactions between proteins inside the cell.[42−46]

The acetylated protein has several possible ways of compensating the large formal charge generated by acetylation: (i) The values of pK_a of carboxylic acids on the surface of the protein might increase (relative to their values in the nonacetylated protein) and, thus, cause those acids to be uncharged under the conditions of crystallization. (ii) Opposing negative charges in crystallographic interfaces might be neutralized—or bridged—by cationic species (e.g., Na^+, NH_4^+) in the crystallization medium. (iii) The negative charges might be screened by opposing uncharged-polar residues. In this study, we observed that most negative charges in crystallographic interfaces tend to be associated with uncharged polar residues (mainly Aly) and/or bridged to opposing negative charges via intermediary cations. (Again, we did not observe cations in the crystal structure, but the absence of an observable electron density for these ions does not indicate that they are not present.) The influence of acetylation on the pK_a's

of surface-exposed residues remains to be evaluated in future work.

Results of this study suggest an unanticipated rationalization for the infrequent (observed across many proteins) participation of lysine in crystallographic interfaces. Previous studies have attributed the infrequent participation of Lys in interfaces to the entropic cost of restricting its highly mobile side chain.[29,47] This proposal seems to be incompatible with our results (for the specific but very well-defined comparisons of $[CA-(NH_3^+)_{18}]^{-3.4}$ and $[CA-(NHCOCH_3)_{18}]^{-19}$), which show that Aly groups, which have a conformational flexibility similar to or greater than that of Lys, make a large contribution to the buried surface area of interfaces. The presence of Aly groups in interfaces, thus, suggests that the infrequent participation of lysine in crystal contacts may result more from its positive charge than from the entropic penalty associated with immobilizing it.

The results of this study also emphasize the significant role of surface charges in mediating interactions between proteins,[35,48] and suggest that charges, as a constraint in determining protein−protein interfaces, might be modulated (or reduced) to control biomolecular recognition. We showed that acetylation of positively charged lysine residues gives rise to new protein−protein interfaces, which are less complementary electrostatically but which involve contacts that are more well-defined geometrically.

ASSOCIATED CONTENT

ⓈSupporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b01105.

> The experimental procedures, additional characterizations of crystal contacts, and computational analyses details (PDF)

AUTHOR INFORMATION

Corresponding Author
*gwhitesides@gmwgroup.harvard.edu.

Author Contributions
[⊥]K.K. and J.-M.C. contributed equally to this work.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Tu, C. K.; Silverman, D. N.; Forsman, C.; Jonsson, B. H.; Lindskog, S. Role of Histidine 64 in the Catalytic Mechanism of Human Carbonic Anhydrase II Studied with a Site-Specific Mutant. *Biochemistry* 1989, 28, 7913.
(2) Venters, R. A.; Farmer II, B. T.; Fierke, C. A.; Spicer, L. D. Characterizing the Use of Perdeuteration in NMR Studies of Large

Proteins: ^{13}C, ^{15}N and ^{1}H Assignments of Human Carbonic Anhydrase II. *J. Mol. Biol.* **1996**, *264*, 1101.

(3) Pacchiano, F.; Aggarwal, M.; Avvaru, B. S.; Robbins, A. H.; Scozzafava, A.; McKenna, R.; Supuran, C. T. Selective Hydrophobic Pocket Binding Observed Within the Carbonic Anhydrase II Active Site Accommodate Different 4-Substituted-Ureido-Benzenesulfonamides and Correlate to Inhibitor Potency. *Chem. Commun.* **2010**, *46*, 8371.

(4) Supuran, C. T. Structure-Based Drug Discovery of Carbonic Anhydrase Inhibitors. *J. Enzyme Inhib. Med. Chem.* **2012**, *27*, 759.

(5) Domsic, J. F.; Avvaru, B. S.; Kim, C. U.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. Entrapment of Carbon Dioxide in the Active Site of Carbonic Anhydrase II. *J. Biol. Chem.* **2008**, *283*, 30766.

(6) Xu, Y.; Feng, L.; Jeffrey, P. D.; Shi, Y.; Morel, F. M. Structure and Metal Exchange in the Cadmium Carbonic Anhydrase of Marine Diatoms. *Nature* **2008**, *452*, 56.

(7) Maupin, C. M.; McKenna, R.; Silverman, D. N.; Voth, G. A. Elucidation of the Proton Transport Mechanism in Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2009**, *131*, 7598.

(8) Mocharla, V. P.; Colasson, B.; Lee, L. V.; Roper, S.; Sharpless, K. B.; Wong, C. H.; Kolb, H. C. In Situ Click Chemistry: Enzyme-Generated Inhibitors of Carbonic Anhydrase II. *Angew. Chem., Int. Ed.* **2005**, *44*, 116.

(9) Fisher, Z.; Hernandez Prada, J. A.; Tu, C.; Duda, D.; Yoshioka, C.; An, H.; Govindasamy, L.; Silverman, D. N.; McKenna, R. Structural and Kinetic Characterization of Active-Site Histidine as a Proton Shuttle in Catalysis by Human Carbonic Anhydrase II. *Biochemistry* **2005**, *44*, 1097.

(10) Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. Carbonic Anhydrase as a Model for Biophysical and Physical-Organic Studies of Proteins and Protein−Ligand Binding. *Chem. Rev.* **2008**, *108*, 946.

(11) Supuran, C. T. Carbonic Anhydrases: Novel Therapeutic Applications for Inhibitors and Activators. *Nat. Rev. Drug Discovery* **2008**, *7*, 168.

(12) Gitlin, I.; Mayer, M.; Whitesides, G. M. Significance of Charge Regulation in the Analysis of Protein Charge Ladders. *J. Phys. Chem. B* **2003**, *107*, 1466.

(13) Carbeck, J. D.; Severs, J. C.; Gao, J. M.; Wu, Q. Y.; Smith, R. D.; Whitesides, G. M. Correlation Between the Charge of Proteins in Solution and in the Gas Phase Investigated by Protein Charge Ladders, Capillary Electrophoresis, and Electrospray Ionization Mass Spectrometry. *J. Phys. Chem. B* **1998**, *102*, 10596.

(14) Gudiksen, K. L.; Gitlin, I.; Yang, J.; Urbach, A. R.; Moustakas, D. T.; Whitesides, G. M. Eliminating Positively Charged Lysine ε-NH_3^+ Groups on the Surface of Carbonic Anhydrase Has No Significant Influence on Its Folding from Sodium Dodecyl Sulfate. *J. Am. Chem. Soc.* **2005**, *127*, 4707.

(15) Eriksson, A. E.; Jones, T. A.; Liljas, A. Refined Structure of Human Carbonic Anhydrase II at 2.0 Å Resolution. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 274.

(16) Hakansson, K.; Carlsson, M.; Svensson, L. A.; Liljas, A. Structure of Native and Apo Carbonic Anhydrase II and Structure of Some of Its Anion-Ligand Complexes. *J. Mol. Biol.* **1992**, *227*, 1192.

(17) Saito, R.; Sato, T.; Ikai, A.; Tanaka, N. Structure of Bovine Carbonic Anhydrase II at 1.95 Å Resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 792.

(18) Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. A Short, Strong Hydrogen Bond in the Active Site of Human Carbonic Anhydrase II. *Biochemistry* **2010**, *49*, 249.

(19) Aaron, J. A.; Chambers, J. M.; Jude, K. M.; Di Costanzo, L.; Dmochowski, I. J.; Christianson, D. W. Structure of a ^{129}Xe-Cryptophane Biosensor Complexed with Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2008**, *130*, 6942.

(20) Fisher, S. Z.; Kovalevsky, A. Y.; Domsic, J. F.; Mustyakimov, M.; McKenna, R.; Silverman, D. N.; Langan, P. A. Neutron Structure of

Human Carbonic Anhydrase II: Implications for Proton Transfer. *Biochemistry* **2010**, *49*, 415.

(21) Fisher, S. Z.; Aggarwal, M.; Kovalevsky, A. Y.; Silverman, D. N.; McKenna, R. Neutron Diffraction of Acetazolamide-Bound Human Carbonic Anhydrase II Reveals Atomic Details of Drug Binding. *J. Am. Chem. Soc.* **2012**, *134*, 14726.

(22) Chothia, C.; Janin, J. Principles of Protein-Protein Recognition. *Nature* **1975**, *256*, 705.

(23) Davies, D. R.; Padlan, E. A.; Sheriff, S. Antibody-Antigen Complexes. *Annu. Rev. Biochem.* **1990**, *59*, 439.

(24) Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13.

(25) Bahadur, R. P.; Zacharias, M. The Interface of Protein-Protein Complexes: Analysis of Contacts and Prediction of Interactions. *Cell. Mol. Life Sci.* **2008**, *65*, 1059.

(26) Schwikowski, B.; Uetz, P.; Fields, S. A Network of Protein-Protein Interactions in Yeast. *Nat. Biotechnol.* **2000**, *18*, 1257.

(27) Nooren, I. M. A.; Thornton, J. M. Diversity of Protein-Protein Interactions. *EMBO J.* **2003**, *22*, 3486.

(28) Nanev, C. N. How Do Crystal Lattice Contacts Reveal Protein Crystallization Mechanism? *Cryst. Res. Technol.* **2008**, *43*, 914.

(29) Derewenda, Z. S.; Vekilov, P. G. Entropy and Surface Engineering in Protein Crystallization. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 116.

(30) Czepas, J.; Devedjiev, Y.; Krowarsch, D.; Derewenda, U.; Otlewski, J.; Derewenda, Z. S. The Impact of Lys– > Arg Surface Mutations on the Crystallization of the Globular Domain of RhoGDI. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 275.

(31) Gitlin, I.; Gudiksen, K. L.; Whitesides, G. M. Peracetylated Bovine Carbonic Anhydrase (BCA-Ac18) is Kinetically More Stable than Native BCA to Sodium Dodecyl Sulfate. *J. Phys. Chem. B* **2006**, *110*, 2372.

(32) Shrake, A.; Rupley, J. A. Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin. *J. Mol. Biol.* **1973**, *79*, 351.

(33) Liddington, R. In *Protein-Protein Interactions*, Fu, H., Ed.; Humana Press, 2004; Vol. *261*, p 3.

(34) Xu, D.; Tsai, C. J.; Nussinov, R. Hydrogen Bonds and Salt Bridges Across Protein-Protein Interfaces. *Protein Eng., Des. Sel.* **1997**, *10*, 999.

(35) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic Aspects of Protein-Protein Interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.

(36) Tanaka, S.; Scheraga, H. A. Medium- and Long-Range Interaction Parameters Between Amino Acids for Predicting Three-Dimensional Structures of Proteins. *Macromolecules* **1976**, *9*, 945.

(37) Roth, C. M.; Neal, B. L.; Lenhoff, A. M. Van der Waals Interactions Involving Proteins. *Biophys. J.* **1996**, *70*, 977.

(38) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. A Dissection of Specific and Non-Specific Protein-Protein Interfaces. *J. Mol. Biol.* **2004**, *336*, 943.

(39) Bordner, A. J.; Abagyan, R. Statistical Analysis and Prediction of Protein-Protein Interfaces. *Proteins: Struct., Funct., Genet.* **2005**, *60*, 353.

(40) Chakrabarti, P.; Janin, J. Dissecting Protein-Protein Recognition Sites. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 334.

(41) Tsuchiya, Y.; Nakamura, H.; Kinoshita, K. Discrimination Between Biological Interfaces and Crystal-Packing Contacts. *Adv. Appl. Bioinform. Chem.* **2008**, *1*, 99.

(42) Janin, J. Specific Versus Non-Specific Contacts in Protein Crystals. *Nat. Struct. Biol.* **1997**, *4*, 973.

(43) Dasgupta, S.; Iyer, G. H.; Bryant, S. H.; Lawrence, C. E.; Bell, J. A. Extent and Nature of Contacts Between Protein Molecules in Crystal Lattices and Between Subunits of Protein Oligomers. *Proteins: Struct., Funct., Genet.* **1997**, *28*, 494.

(44) Janin, J.; Rodier, F. Protein-Protein Interaction at Crystal Contacts. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 580.

(45) Carugo, O.; Argos, P. Protein-Protein Crystal-Packing Contacts. *Protein Sci.* **1997**, *6*, 2261.

(46) Janin, J. Principles of Protein-Protein Recognition from Structure to Thermodynamics. *Biochimie* **1995**, *77*, 497.

(47) Price II, W. N.; Chen, Y.; Handelman, S. K.; Neely, H.; Manor, P.; Karlin, R.; Nair, R.; Liu, J.; Baran, M.; Everett, J.; et al. Understanding the Physical Properties That Control Protein Crystallization by Analysis of Large-Scale Experimental Data. *Nat. Biotechnol.* **2009**, *27*, 51.

(48) Gitlin, I.; Carbeck, J. D.; Whitesides, G. M. Why Are Proteins Charged? Networks of Charge-Charge Interactions in Proteins Measured by Charge Ladders and Capillary Electrophoresis. *Angew. Chem., Int. Ed.* **2006**, *45*, 3022.