Article

The Solvation of the *E. coli* CheY Phosphorylation Site Mapped by XFMS

Maham Hamid [1] , Muhammad Farhan Khalid [1] , Safee Ullah Chaudhary [1],* and Shahid Khan [2,3],*

1 Biomedical Informatics and Engineering Research Laboratory (BIRL), Lahore University of Management Sciences (LUMS), Lahore 54792, Pakistan
2 Syed Babar Ali School of Science & Engineering, Lahore University of Management Sciences (LUMS), Lahore 54792, Pakistan
3 Molecular Biology Consortium, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
* Correspondence: safeeullah@lums.edu.pk (S.U.C.); khansh@lums.edu.pk (S.K.)

Abstract: The *Escherichia coli* CheY protein belongs to a large bacterial response regulator superfamily. X-ray hydroxy radical foot-printing with mass spectroscopy (XFMS) has shown that allosteric activation of CheY by its motor target triggers a concerted internalization of aromatic sidechains. We reanalyzed the XFMS data to compare polar versus non-polar CheY residue positions. The polar residues around and including the 57D phosphorylated site had an elevated hydroxy radical reactivity. Bioinformatic measures revealed that a water-mediated hydrogen bond network connected this ring of residues with the central 57D. These residues solvated 57D to energetically stabilize the apo-CheY fold. The abundance of these reactive residues was reduced upon activation. This result was supported by the bioinformatics and consistent with the previously reported activation-induced increase in core hydrophobicity. It further illustrated XFMS detection of structural waters. Direct contacts between the ring residues and the phosphorylation site would stabilize the aspartyl phosphate. In addition, we report that the ring residue, 18R, is a constant central node in the 57D solvation network and that 18R non-polar substitutions determine CheY diversity as assessed by its evolutionary trace in bacteria with well-studied chemotaxis. These results showcase the importance of structured water dynamics for phosphorylation-mediated signal transduction.

Keywords: solvent accessible surface area (SASA); mass spectroscopy; protein phosphorylation; protection factor (PF); MATLAB toolbox; structured water; protein fold energetics

1. Introduction

The folding problem—the compaction of a 1D polymer to a 3D globule—is a fundamental problem in molecular biology. It has been known for some time that hydrophobic interactions form the protein core [1]—with globular proteins approximated as oil-drops coated by a polar shell that contacts the aqueous solvent. It is now appreciated that water-mediated interactions are also important for the protein fold [2], highlighted by the visualization of immobilized water molecules in high-resolution crystal structures. Studies have also revealed the role of structured water in protein allostery—the ability of proteins to change shape [3]. The role of structured waters is being extensively studied for drug design [4,5], with experiments, bioinformatic data mining, and simulations to detect and characterize water dynamics in hydrogen bonded networks and internal cavities.

Protein folding and allosteric transitions are controlled by dynamic networks that integrate local transitions—the melting of α-helices, the rotation of aromatic sidechains, and the immobilization of surface loops [6] to trigger collective motion [7]. Time-resolved foot-printing techniques are an important tool for mapping global protein transitions with single residue resolution based on the characterization by mass spectroscopy (MS). X-ray hydroxy radical foot-printing (XF) with mass spectroscopy (XFMS) is one such method

that monitors the solvent accessibility of residue sidechains to map conformational transitions in solution [8]. XFMS studies have shown that residues in proximity to structured water have anomalous oxidation rates that are elevated relative to residues in contact with bulk water (reviewed by reference [9]). These studies have targeted membrane G protein coupled receptors (GPCRs) and channel proteins, where the anomalous reactivity is readily correlated with the presence of the few buried water molecules seen to be near the affected sidechains [10–12]. For soluble proteins, the temperature dependence and O^{16}/O^{18} exchange have distinguished between bulk versus structured water residue oxidations. Oxidation of a cytochrome-C (Cyt-C) residue (67Y) hydrogen-bonded to water in the crystal had a markedly longer response time to fast O^{16}/O^{18} water exchange than other Cyt-C residues [13]. Since this work, the interpretation of XFMS dose response curves has advanced by the development of protection factors (PFs) [14,15], a quantitative metric for the relation of oxidation rate to solvent accessibility [12].

We wondered whether the analysis of protection factors (*PF*s) might offer a rapid diagnostic to separate bulk and structured water oxidations in soluble proteins, where the large number of protein-associated waters in the crystal structures makes a straightforward correlation between anomalous reactivity and water proximity difficult. We reworked XFMS data acquired previously on the *Escherichia coli* CheY bacterial response regulator [16] to test this idea. CheY, a 14 kilodalton protein with a α/β flavodoxin-type fold, is the founding member of a large aspartyl phosphate response regulator superfamily involved in diverse intracellular signal transduction circuits [17]. The CheY crystal structures have revealed an extensive hydrogen-bonded protein-water network modulated by binding targets. CheY binds to the N-terminal peptide of FliM$_N$, a rotor component of the flagellar motor, to switch counterclockwise (CCW) to clockwise (CW) rotation sense [18–20]. The XFMS study reported aromatic residue rotations triggered by FliM$_N$ in the CheY13DK106YW mutant. The phosphorylated CheY state has been difficult to characterize due to the short lifetime of the aspartyl phosphate. Consequently, chemical modification and residue substitution have been used extensively as a substitute to reproduce aspects of the physiology and biochemistry obtained upon phosphorylation of native CheY. The *CheY*13DK106YW gene resets motor rotation from CCW -> CW in absence of phosphorylation when expressed in Δ*CheY* mutant *E. coli*. In this study, it will be assumed to be the activated CheY form.

SPECTRUM [21] is an open-source software suite for high-throughput proteoform identification. Its extension, SPECTRUM-XFMS, a custom high-throughput algorithm, was developed for XFMS analysis. Recently developed bioinformatic algorithms, in addition to inspection of the crystal structures, were utilized to infer how the water associated with the polar residue positions detected by XFMS affected the energetics and the hydrogen-bonded network of the protein fold. We found that these residues formed an ion ring around the 57D phosphorylation site. The ring was reduced upon activation consistent with water depletion due to the stabilization of the hydrophobic core by concerted internalization of aromatic sidechains [16]. The evolutionary trace of CheY proteins specifically involved in chemotaxis revealed that non-polar substitutions in one residue position (*E. coli* 18R), a central node in the hydrogen-bonded network, were central to evolutionary diversity of CheY chemotactic function. These results advance knowledge of CheY signal transduction and, more generally, develop XFMS methodology as a diagnostic tool for structured waters.

2. Results

2.1. Polar Residues Have Lowered Protection against Hydroxy Radical Oxidation

The *E. coli* apo-CheY (3CHY.pdb) and CheY13DK106YW.FliM$_N$ (1U8T.pdb) crystal structures, henceforth cited as Mg^{2+}.free *CheY$_I$* and *CheY$_A$* respectively, reveal an extensive hydrogen-bonded network. The CheY 57D phosphorylation site is ringed with polar residues (E, D, K, R, N) for coordination of the magnesium ion and the aspartyl phosphate. The XFMS oxidation dose response plots for these 10 residues, determined in Mg^{2+} buffer solutions for both states (Materials and Methods) are shown in Figure 1. The oxidation of nine other ring residues was detected for one state alone. One residue (75D) external to the

ring was oxidized in both states. Its oxidation rate increased in the active form as opposed to the rate decrease for the majority of the 57D ring residues. The *E. coli* protein sequence covered (89%) by the tryptic peptides submitted for MS analysis included 35 of the 42 total CheY (E, D, K, R, N) residues. The oxidation of the remaining 15 (E, D, K, R, N) residues was not detected.
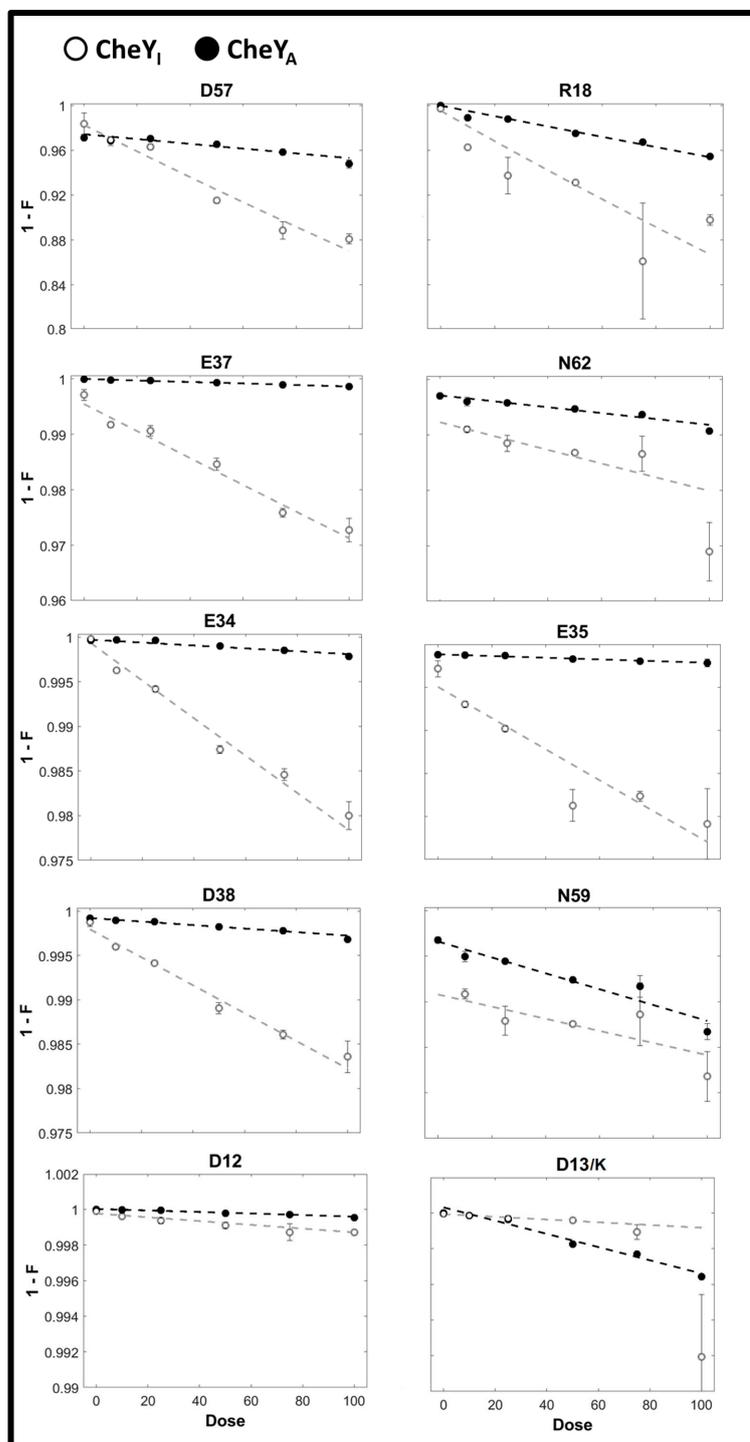

Figure 1. XFMS dose-response curves for residues around the 57D phosphorylation site. Oxidation rates were computed from the slopes of single exponential fits, weighted by reciprocal of the standard error, for all ten residues, including 57D. The $CheY_A$ rates (black dashed lines) were lower than those for $CheY_I$ (grey dashed lines) in all cases, except for 13D/K and 59N. The intrinsic reactivity to hydroxy radical oxidation is greater for K (2.2) than D (0.42).

The solvent accessible surface area (SASA)–Log (PF) regression was constructed from the linear best-fit to $CheY_I$ non-polar methionine and aromatic residue oxidations plus the $CheY_A$ aromatic residue oxidations alone, as methionine was replaced by selenomethionine in the $CheY_A$ construct used for crystallization (Figure 2). The Pearson's correlation (-0.6) was comparable to that reported previously [16]. The corresponding values for the polar sidechains were then superimposed on the plot for comparison. They partitioned into two distinct sets demarcated by the lower 95% confidence limit to the best-fit relation for the non-polar sidechains. One set ((12D, 13D/K, 37E, 59N, 62N), $CheY_I$ (27E, 64D, 89E, 91K, 106K), $CheY_A$ (38D, 118D, 119K)) followed the (SASA)–Log (PF) relation obtained for the non-polar residues. The second set with Log (PF) values below the lower limit consisted of three residues 18R, 57D (phosphorylation site), and 75D. The $CheY_I$ residues 34E, 35E, 38D, 41D, 94N also belonged to this set.
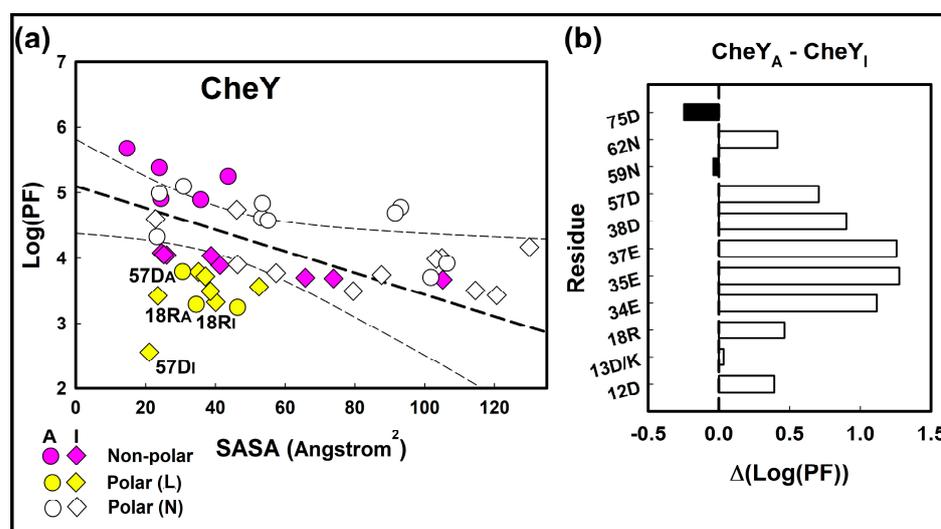

Figure 2. PF differences between residue types and activated states. (**a**). The solvent accessibility of non-polar residues (magenta symbols) varies linearly with SASA. The least-square best fit linear regression (thick dashed line) and 95% confidence limits (thin dashed lines) are shown. Most polar residues (white symbols (N)) follow the best-fit relation, but a significant fraction (yellow symbols (L)) have anomalously high solvent accessibility (low Log (PF)) A = $CheY_A$, I = $CheY_I$. (**b**). The activation-dependent change in polar residue Log (PF) values. There is a marked positive Log (PF) ($\Delta(CheY_A - CheY_I)$) increase for the 34E, 35E, 37E, 38D residue quadruplet ringing 57D, modest increases for 12D, 18R, 57D, 62N, and no significant change for 13D/K, 59N. The $\Delta(CheY_A - CheY_I)$ Log (PF) for 75D outside the 57D solvation network has a negative value.

The non-polar $CheY_A$ residues spanned a smaller SASA relative to their $CheY_I$ states, in line with activation-dependent burial of aromatic sidechains. The non-polar sidechains had increased protection in $CheY_A$ relative to $CheY_I$ (Figure 2b). There was no change (ΔSASA = -1.0 ± 6.9) for the polar residues, except the change due to the 13D/K substitution, in contrast to the aromatic residues (ΔSASA = 28.3 ± 10.8).

2.2. CheY Activation Stabilizes the 57D Phosphorylation Site

Next, we analyzed the PDB structure files to study the role of local waters in the deviation from the (SASA)–Log (PF) relation established for the non-polar residue positions.

First, the energetics of the contacts between the oxidized residue positions was estimated by Frustratometer2, based on knowledge of the local dielectric, and packing density in the CheY fold (Figure 3). Energetically stressed contacts with $CheY_I$ 18R, 34E, 38D, 37E, 41D, 94N (low PFs) ringed the 57D phosphorylation site. The oxidized residue positions (normal PFs) completed the ring, with 75D the sole exception. The energetics of a Mg^{2+}-bound $CheY_I$ state (2CHE.pdb [22]) were also analyzed. The Mg^{2+} ion did not notably alter

the ring of energetically stressed contacts. In contrast, the ring was reduced in $CheY_A$. This reduction correlated with the reduction in the number of low PF residue positions to 18R, consistent with water depletion from the more hydrophobic core due to burial of aromatic residue sidechains. The decrease in frustrated contacts provides an independent metric to the ionizable residue electrostatics used previously [16] to document the stabilization of the CheY core upon activation.
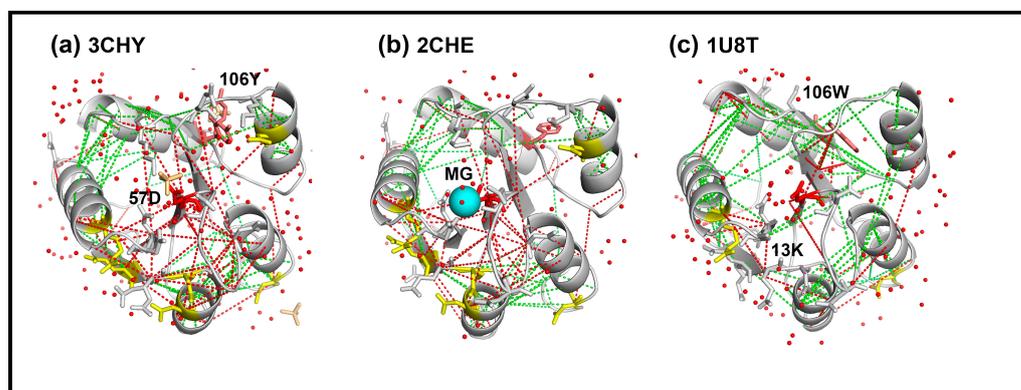

Figure 3. Activation-dependent change in fold energetics. The Frustratometer2 algorithm computes residue contact energies relative to a decoy contact distribution with a coarse-grained AWSEM force-field. The PDB structures with superimposed high (red lines) and low (green lines) contact energies relative to the decoy distributions are shown. (**a**) Apo-CheY. High-energy contacts radiate out from 57D (red sidechain) to surrounding residues that include residue positions detected by XFMS (stick sidechains) with anomalous low (yellow) or comparable (white) Log (PF) values to non-polar residues (see text). 106Y sidechain (salmon). (**b**) Mg^{2+} bound apo-CheY. Magnesium ion (cyan sphere). (**c**) 13DK,106YW CheY. Activated by bound $FliM_N$ (not shown). XFMS residue positions in (**b**,**c**) as in (**a**). There is a reduction in the high energy contacts of 13D/K, 106YW CheY 57D with surrounding residue positions relative to the apo-CheY forms. Waters (red dots).

Frustratometer2 does not report on water positions but infers water-mediated contacts based on the protein packing density. It assessed that all polar residues detected by XFMS had water-mediated contacts with their neighbors that included 57D in several cases. The energetically stressed contacts may have been maintained for formation of the aspartyl phosphate, with the waters acting to bridge negatively charged sidechains. In $CheY_A$, the 13D/K substitution would reduce the water requirement for such purpose.

2.3. Graphical Analysis of the Protein-Water Hydrogen Bonded Network

There was a qualitative correlation between the Frustratometer2 inferred water-mediated contacts with water positions seen in the crystal structures, but the coarse-grained AWSEM (associative memory, water mediated, structure and energy model) force-field on which the algorithm is based precluded a rigorous analysis. Therefore, we analyzed the network architecture of the crystal structures with Bridge2 to better characterize the local solvent density. The Bridge2 algorithm, developed since publication of these structures, is a graphical tool for representation and analysis of the full-atom protein-water hydrogen bond network. The bonds included intervening water molecules [19,20]. There are 147 and 136 water molecules associated with the CheY monomer in the Mg^{2+} free, 1.7 angstrom resolution 3CHY.pdb and 1U8T structures, respectively. In addition, Mg^{2+} bound structures of a $CheY_I$ form (2CHE.pdb (1.8 angstrom resolution, 99 waters [22])) and a beryllium fluoride (BeF_3) $CheY_A$ F_3 form in complex with CheZ peptide (2FQA.pdb (2.0 angstrom resolution, 80 waters [23])) were analyzed. We obtained maps of their protein-water networks characterized by their centrality (Figure 4). The networks were superimposed on the 3D structures and displayed as a 2D map. Inspection of these representations yielded insights into the rearrangement of the hydrogen bonds involved in allosteric activation of CheY.
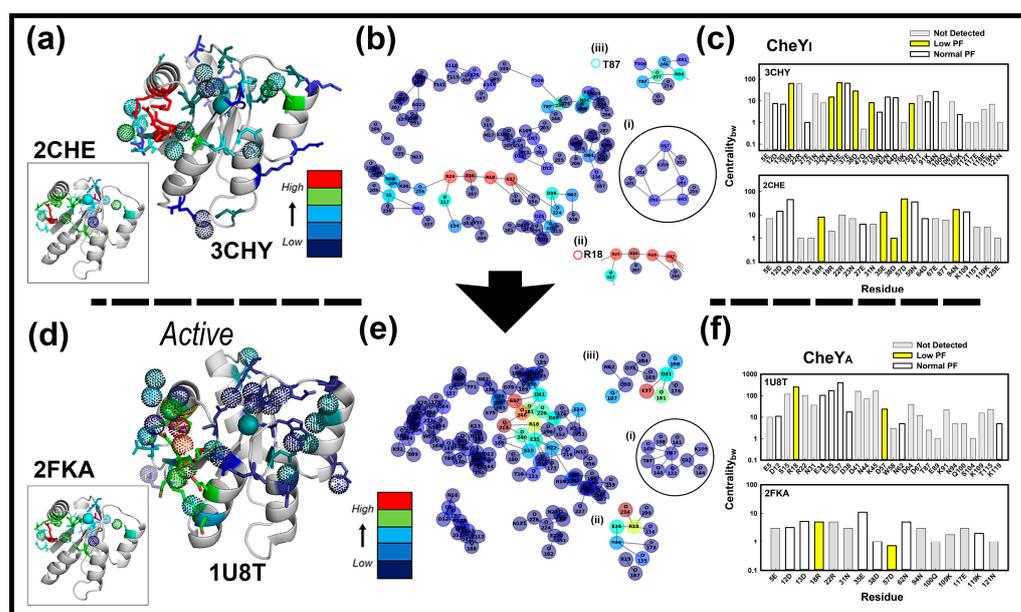
Figure 4. Changes in the Bridge2 CheY hydrogen-bonded protein-water network upon activation. Bars = Color-coded C_{bc} values (Blue (Low) -> Red (High)). (**a**–**c**). Inactive apo-CheY. (**a**) The C_{bc} nodes mapped onto the 3D structure. Residues detected by XFMS (stick sidechains). Water (dotted spheres). Mg^{2+} (cyan spheres), BeF_3 (magenta sticks). (**b**) The 2D map of the network. Local nodes (**i**) 109K -> 57D (circled [19]) (**ii**) 18R. (**iii**) 94N -> 106Y. (**c**) The C_{bc} nodes targeted by XFMS detected residues (low (yellow), normal (white)). (**d**–**f**). Activated CheY13DK106YW in complex with FliMN. (**d**) The C_{bc} nodes mapped onto the 3D structure. Residues detected by XFMS and waters (dotted spheres) as in (**a**). (**e**) The 2D map of the network. Local nodes (**i**) 109K -> 57D (circled [19]) (**ii**) 18R. (**iii**) 37E -> 41D, 38D -> 62N, 75D (**f**) The C_{bc} nodes targeted by XFMS detected residues (anomalous (yellow), comparable (white)). Bridge2 residue nomenclature, where type precedes position, is different from nomenclature used elsewhere in the text.

Inspection of the 3D maps (Figure 4a,d) superimposed on the crystal structures showed both protein and water nodes clustered around, and including, 57D. A structurally conserved dome of water molecules above 57D was a feature of this water map. The prominent 18R residue node connected the 57D network, with nodes mapped on the α_1-β_2 surface.

The Bridge2 2D maps (Figure 4b,c,e,f) reported three features of interest. First, the algorithm faithfully documented the published [19], hydrogen-bonded rearrangements triggered by allosteric activation. Second, the dominant clusters in both cases were centered around 18R/35E. Additional clusters were the water-mediated 87T-94N-106Y cluster in *CheY$_I$* 3CHY and the 37E, 41D, 38D-62N, 75D clusters in *CheY$_A$* 1U8T. Third, the anomalous, low (PFs) residue positions (ALPFs) revealed by the XFMS experiments targeted the centrality (C_{bc}) profiles of the Bridge2 mediated protein-water networks. The C_{bw} values report high connectivity within the protein-water network not local solvent density, but we hoped that high values reporting hydrogen-bonded clusters would be correlated with solvent density. The *CheY$_I$* 3CHY network (15 water, 30 (detected 17) amino acid) residue node C_{bc} values were 31.3 $\pm$ 9.6 (low), 15.1±6.1(normal), 11.1 $\pm$ 4.4 (not detected). The 2CHE network (7 water, 23 (11 detected) amino acids) residue node C_{bc} values were 17.5 $\pm$ 8.2 (low), 19.9 + 6.8 (normal), 4.3 $\pm$ 0.9 (not detected). The merged *CheY$_I$* networks had a (Detected/Undetected) C_{bc} value ratio of 2.6 $\pm$ 0.6. The residue C_{bc} node values for the 1U8T network (35 water, 27 (detected 9) amino acids), were 137.9 $\pm$ 114.1 (low), 101.6 + 54.9 (normal), 43.8 $\pm$ 13.0 (not detected). The residue C_{bc} node values for the 2FKA network (5 water, 16 (detected 8) amino acids) were 2.8 $\pm$ 2.1 (low), 4.6 + 1.4 (normal), 2.7 $\pm$ 0.9 (not detected). The merged *CheY$_A$* networks had a (Detected/Undetected) C_{bc} value ratio of 2.4 $\pm$ 1.2. The separation of the ALPFs 18R and 57D from other residues

detected by XFMS for $CheY_A$ may not have statistical significance. Nevertheless, the (Detected/Undetected) C_{bc} ratios for both $CheY_I$ and $CheY_A$ were significant, consistent with a correlation between C_{bc} and solvent density.

2.4. The ALPF 18R Is Important for CheY Chemotactic Function

The 57D, 12D, and 13D residue positions are important for phosphorylation, hence chemotaxis [17]. There are no known residue substitutions to our knowledge that affect chemotaxis in the ALPF 18R or the $CheY_I$ ALPFs 34E, 35E, 37E, and 38D. Therefore, we used EV trace to assess the importance of these residues to chemotactic function; in particular, residue 18R prominent is mediating both connectivity and stabilization of the CheY core.

The EV-Trace for *E. coli* CheY matched the trace deposited in UniProt for the closely related *Salmonella typhimurium* CheY. Each trace collected 421 sequences for a multiple sequence alignment (MSA) and phylogenetic tree construction (Figure 5a). The rvET scores combined both positional conservations, as determined by the MSA, and evolution, as determined by the phylogenetics. The scores for the 12 top-ranked residue positions (10% of the protein) are listed in Figure 5b. Residue 18R ranked tenth on this list. Five of the other remaining residues typically determine fold architecture and dynamics. The prolines (61P, 110P) disrupt α-helix formation and the glycines (39G, 65G, 110G) form flexible hinges. Mutational studies have established the importance of 109K and 87T for chemotaxis, in addition to 12D, 13D, and 57D [24]. Thus, 18R together with 60M are the only residues in the list whose functional relevance is not known.

We parsed out CheY sequences dedicated to chemotaxis from the large and functionally diverse bacterial response regulator superfamily to address this issue. A total of 56 CheY sequences were collected from 22 species whose chemotaxis machinery is under experimental investigation. Most species have multiple CheY proteins known to interact with flagellar motor or chemoreceptor complex proteins, in contrast to *E. coli*. The chemotaxis-specific CheY MSA determined the residue conservation for the 12 positions identified by EV Trace (Figure 5c), while the rvET scores from the larger, 421 sequence CheY dataset were superimposed on the 3CHY structure to show the range of scores for the 11 residue positions detected in Mg^{2+} bound apo-CheY ($CheY_I$) and the $FliM_N$-bound 13DK106YW ($CheY_A$) (Figure 5d). Finally, the phylogenetic tree constructed from the chemotaxis specific MSA, less diverse than the EV Trace tree, was analyzed to identify residue substitutions at its branch points (Figure 5e). We found 18R residue substitutions to non-polar residues (I, L, V) determined the tree morphology. Residue substitutions for 39G, 61P, 87T, and 102G also influenced tree morphology. In contrast, residue 60M was invariant. We conclude, based on this evidence, that evolutionary variation of 18R, but not 60M, is important for modulation of CheY chemotactic function.
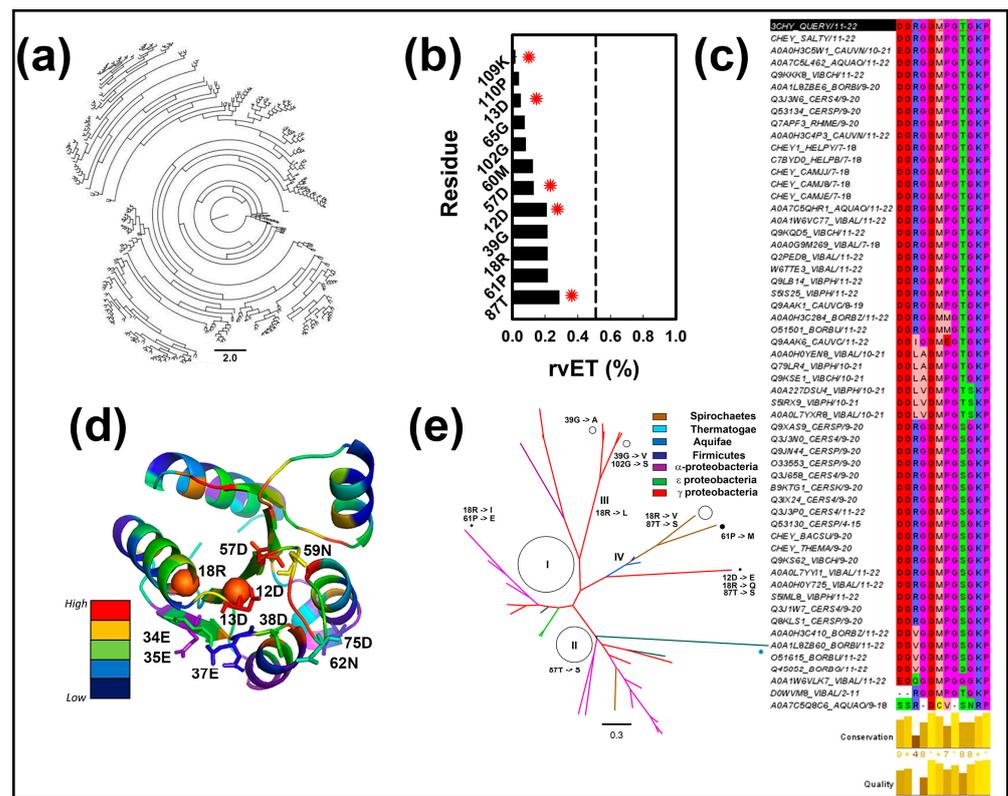
Figure 5. The evolutionary importance of ALPF 18R in CheY evolution. (**a**) The phylogenetic tree for *E. coli* CheY. Based on 443 sequences used by EV-Trace. Scale bar = 2.0. (**b**) Top (10%) residue positions. The *E. coli* CheY residue positions were ranked by rv.ET score. Asterisks denote residues known previously to be important for chemotaxis. (**c**). Residue conservation of the top rv.ET ranked positions. From an MSA of 56 sequences from 12 distinct species. (**d**) The 3D CheY structure color-coded by ev.ET score. The residues oxidized by XFMS are labelled (stick sidechains). Spheres mark the residues 12D and 18R (spheres) important for CheY evolution under selection for chemotaxis. (**e**) The phylogenetic tree constructed from the MSA of CheY chemotaxis proteins. 56 CheY sequences. The non-polar 18R substitutions localize at the internal tree nodes. Roman numerals denote major branches. Circle diameter = branch size. Note difference in scale with the EV-Trace tree (**a**).

3. Discussion

We have identified polar residues around the 57D phosphorylation site and 57D itself that have anomalous *PF*s compared against previously analyzed non-polar methionine and aromatic residues. The solvent accessibility of the latter residue positions scales with the SASA computed from the crystal structures as reported previously [16]. Two of the polar residue positions, 18R and 57D, retain anomalous, low protection, regardless of activation state. Bioinformatic metrics supplemented inspection of the crystal structures suggest that the anomalous reactivity of these and other residues is due to hydrogen-bonded water molecules that solvate and stabilize the phosphorylation site.

The Frustratometer2 metrics are a first attempt to interpret XFMS data in the context of the energetics and architecture of the protein fold. Several studies have shown that energetically frustrated contacts are diagnostic of binding interfaces that are stabilized upon association of the binding partner [25] or ligand [26]. The present study suggests that this tool may be used in the same way to diagnose covalent modification sites in proteins.

Amide and carbonyl groups play key roles in water-mediated hydrogen bonds. So, as expected, acidic and basic amino acids dominated the list compiled by Bridge2. The inclusion of asparagine (N) in the list is consistent with the fact that its amide group can accept as well as donate two hydrogen bonds. X-ray crystallography has shown that the Mg^{2+} ion, essential as cofactor for phosphorylation, is coordinated to 13D, 57D, and 59N,

with water-mediated contact to 13D [22,23,27]; it has also documented hydrogen-bond rearrangements between 12D, 57D, 87T, and 109K triggered by activation. Studies have also explored the effects of charge substitutions on chemotaxis [28] and phosphorylation kinetics [29]. However, a structural or functional role for 18R and the β_2-α_2 loop residues (34E, 35E, 37E, 38D) has not previously been recognized to our knowledge. The present study makes the case that they are important for solvation of the aspartyl phosphorylation site.

The lack of insight provided by sequence alignments on the diverse motor responses to CheY obtained in different species [30] or the interactions of multiple CheY proteins with the same flagellar motor in several species [31] has been a longstanding conundrum. The role of 18R as a determinant of phylogenetic diversity, together with the centrality of this residue position in the *E. coli* CheY hydrogen-bonded network, offers a tantalizing clue that merits further investigation. The unique properties of the arginine guanidium group in forming out-of-plane as well as in-plane interactions with neighboring residues as noted for channel gating [32], may also be relevant to 18R centrality in the CheY network.

The present study adds to evidence [13] that hydrogen-bonded waters facilitate hydroxy radical oxidation due to slower exchange with bulk phase water. The application of bioinformatic measures to diagnose structural water is an advance, but it remains reliant on the detection of water positions in crystal structures. Water positions may have been missed during refinement, particularly for the 2.0 resolution 2FKA crystal structure. In addition, the occupancy and dynamics of detected water positions are not known, let alone the fact that undetected waters are less mobile than bulk water [33]. While water positions detected in crystal structures will remain relevant, it should be possible, moving forward, to couple time-resolved measures of hydroxy radical oxidation that readout O^{16}/O^{18} oxygen exchange and the temperature dependence [9] with full-atom, femtosecond simulations of water dynamics [34]. The contact map of fold energetics by the AWSEM force-field is another advance that should be extended by a full-atom force field to accurately model energy frustration [35] allowing better estimation of parameters such as packing density [14] for interpretation of measured *PF*s. The XFMS approach, therefore, could then provide quantitative insights on the role of structured waters on protein fold and function.

4. Materials and Methods

4.1. Hydroxy Radical Foot-Printing (XF) with Mass Spectroscopy (MS) Analysis

The XFMS experiments on CheY and CheY13DK106YW.FliM$_N$ in Mg^{2+} buffers were reported previously [16]. In that study, oxidations of aromatic (phenylalanine (F), tryptophan (W), tyrosine (Y)), methionine, proline, and lysine residues from technical ES-MSI duplicates from two XF experiments were identified by MASCOT analysis of the MS-MS spectrum at maximal dose. These data documented the concerted internalization of aromatic sidechains during allosteric activation of CheY by the 13DK106YW substitutions and the N-terminal motor peptide, FliM$_N$ [16]. The quantification of single residue solvent accessibility was possible due to development of an important metric, the residue protection factor (*PF*) [14,15].

$$PF = S/I_R \tag{1}$$

where *S* is the slope of the single exponential fit to the plot of the unoxidized fraction (1-(F, oxidized fraction)) and I_R is the intrinsic reactivity of the residue sidechain to hydroxy radical oxidation [8]. The solvent accessibility is proportional to Log (*PF*), as established by a linear relation between the residue solvent accessible surface area (SASA), determined from the crystal structures, and Log (*PF*).

Here, we reanalyzed three technical ES-MSI replicates from one of the XF experiments selected for greater sampling frequency and lowered flux that minimized secondary effects of radiolysis. We generated new MASCOT files to identify oxidation peaks for polar residues abundant in CheY (arginine (R), lysine (K), glutamate (G), aspartate (D), asparagine (N)) to assess protein water interactions that are typically mediated by carbonyl and amide groups. Oxidation profiles for the previously analyzed methionine (M) and the aromatic residues that lack these groups were used as controls for SASA versus Log (*PF*) plots.

The bottleneck for computation of the oxidized fraction (F) at single residue positions is the identification of multiple oxidized residue peaks in the MS-ESI chromatograms. The multiplicity results from multiple residue oxidations within a peptide. The match of the retention times (RTs) reported for the MS/MS spectrum by the MASCOT file with the eluted RTs in the ESI chromatograms typically retrieves multiple peaks that must be parsed to compile the total oxidation fraction for different residues within a peptide. Accordingly, we have developed a semi-automated algorithm that is over two orders of magnitude more rapid than manual analysis with Excel (or similar) as schematized (Figure 6).
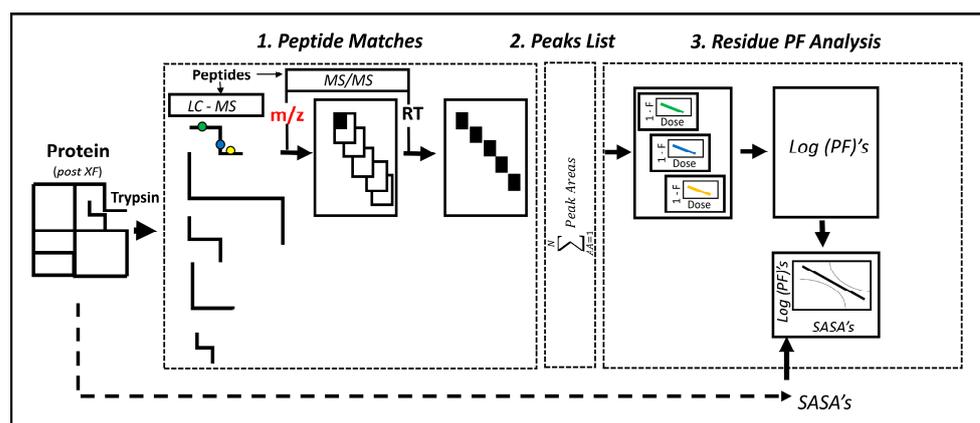
Figure 6. Semi-automated algorithm for determination of PFs and the SASA–Log (PF) relation. There are three distinct stages to the MATLAB code. The code is generally applicable. 1. The proteolyzed (tryptic) peptides from the XF experiment were analyzed by LC-MS and MS/MS. The pipeline illustrated for one peptide was iterated for all peptides. The resulting Mass Hunter (LC-MS) and MASCOT (MS/MS) files were matched by RT (retention time) to parse out the oxidized and unoxidized peak areas (black sections) for each residue from unrelated peaks in the lists obtained by manual entry (red) of the MASCOT *m/z* values. 2. A list with the oxidized and unoxidized peak areas for all detected residues in the protein was compiled. 3. Computed oxidation rates were divided by intrinsic reactivities for determination of residue PFs. Data for all residues were combined for the SASA versus Log (PF) relations. The SASAs were obtained from the protein crystal structure or Alpha-fold model. Residues with Log (PF) values below the 95% confidence limit for the best-fit linear relation were designated as ALPRs (anomalous protection factor residues). See Appendix A for details.

4.2. Structural Bioinformatics

4.2.1. Frustratometer2

The associative memory, water mediated, structure and energy model (AWSEM) is a coarse-grained force-field that utilizes a folding funnel strategy to predict 3D structure from 1D sequences [36]. It endeavors to account for the role of waters in the folding process [2]. Frustratometer2 [37] is based on AWSEM for analysis of the energetics of the 3D fold visualized by the crystal structures. Contact energies between atom pairs are scored relative to a decoy library generated by substitution of all possible amino acids at the two residue positions.

The frustration index, ΔEfr, computes the energies of the native residue contacts relative to the distribution of decoy energies, obtained by randomizing the identities of the residues in the native (ij) contacts with n randomly selected amino acid combinations (h) [37].

$$\Delta Efr_{ij} = \left(\Delta E_{ij}^N - \overline{\Delta E_{i'j'}^D}\right) / \sqrt{(1/n)\sum_{k=1}^{h}\left(\Delta E_{i'j'}^D - \overline{\Delta e_{i'j'}^D}\right)^2} \qquad (2)$$

A high score for the native contact relative to the decoy distribution implies that its evolution is influenced by factors other than fold stabilization. The nature of the contact

is categorized as water mediated or not, based on the local protein density as well as by distance—short or long.

4.2.2. Bridge2

The Bridge2 algorithm is a tool for rapid computation of the protein–water hydrogen-bonded network from either a single structure (static selection) or a multi-frame trajectory [38]. The CheY structures 3CHY.pdb and 1U8T.pdb, diagnostic of the inactive and activated states respectively, were entered as static selections. The default distance (3.5 angstrom) and angle (60°) were used for modeling hydrogen bonds. Bridge2 calculates the betweenness centrality of the network, C_{bw}, that measures how often a node, n, lies on the shortest path between a pair of nodes n1 and n2 [39].

$$C_{bw}{}^n = \sum_{n1=1-N, n2=1-N} (X_{n1n2})/(M_{n1n2}) \tag{3}$$

where X = number of shortest paths between $n1$, $n2$, passing through n, and M = number of total paths between $n1$, $n2$.

The graphical interface can navigate a complex many-body network to display projections that can be selected by residue position and type to identify key network features. The computations are accelerated by the usage of k-d trees, with Dijkstra's algorithm used for identification of the shortest paths. Residue sidechain, but not backbone, atoms were considered in the calculations.

4.3. Phylogenetics

EV-Trace

The real value EV Trace score (*rvET*) considers both the entropy, based on the variability of a residue position in a multiple sequence alignment, with phylogenetics, based on the correlation between residue substitutions and branches within a superfamily tree [40].

The *rvET* integrates the entropy and dendrogram location of each residue position in the MSA, weighted for evolutionary distance.

$$rvET_i = 1 + \sum_{n=1}^{N=1} w_{node}(n) \sum_{g=1}^{n} w_{group}(g) x s_i \tag{4}$$

where w_{node} and w_{group} are the phylogenetic tree nodes and tips, respectively. The s_i is the information entropy that measures the frequency of occurrence, (f_{ia}), of amino acid a in residue position i within the MSA.

$$s_i = -\sum_{a=1}^{20} f_{ia} ln(f_{ia}) \tag{5}$$

It is valuable for dissecting the evolution of a specific function within a large family such as the bacterial response regulator Pfam superfamily (PF00072 (>1.5 million sequences)) (https://www.pfam.org). The EV Trace for *E. coli* CheY was indistinguishable from the EV-Trace for *Salmonella typhimurium* CheY in complex with a CheA kinase peptide (1A0O.pdb) available from UniProt (https://www.uniprot.org). Another phylogenetic tree was constructed from a MSA of CheY sequences for chemotactic bacteria under active study. The top twelve residues based on the EV-Trace scores were mapped onto this tree to evaluate their influence on branch morphology.

Appendix A. Semi-Automated Analysis of XFMS Spectra

1. Peptide Matches.

(i) The ESI MS data were stored as Mass Hunter files with radiation dose data from tryptic peptides organized into distinct folders with the following hierarchy.
"Protein
-> Peptide 1, Peptide 2, —-
-> Replica 1, Replica 2, —-
-> Dose 0, Dose 1, —-"
Each peptide was characterized by its elution (retention time (RT)) and abundance (peak area) extracted from the Mass Hunter GUI (Figure A1). Dose is integrated intensity of the oxidizing X-ray radiation, varied by pulse duration (ms).

(ii) MASCOT analysis of the MS/MS run of the trypsin-digested sample after application of a maximal duration (100ms) pulse. Oxidized residue positions were identified in a MS/MS run by fragmentation of the tryptic peptides. The 1D protein sequence is inferred in MASCOT from the peptide sequences by search of stored database.

(iii) The 3D structure (PDB) files, given the protein, may be retrieved from the protein sequence using UniProt [41] or modelled with Alpha fold [42].

(iv) A table of residue intrinsic reactivities [43] is required to compute residue solvent accessibility from the oxidation rates derived from the plot of the oxidized fraction versus X-ray dose. For any residue (AA), the intrinsic reactivity = I_R^{AA}.

MANUAL read in of the peptide m/z values ($\pm$ 0.1) in Mass Hunter initiated the extraction of chromatograms at the entered values (Figure A1), typically with multiple peaks. The peaks corresponding to the sample peptides were retrieved by matching the RTs in the peak list to the RTs recorded for the MS/MS run in the MASCOT file. Tolerance = $\pm$ 0.21. Oxidized peak m/z values were (($16/z$) × k) Daltons greater than the corresponding m/z values of the unoxidized peaks, where k = oxidation state.
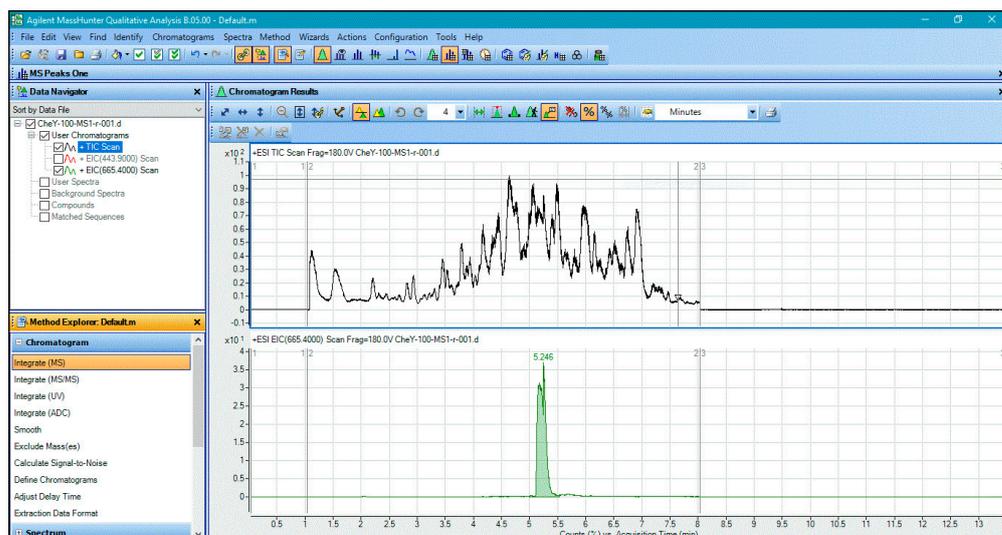
Figure A1. Screenshot of Mass Hunter GUI.

Peptides with different oxidized residues frequently could not be separated based on RTs alone since the RTs had a weak dependence on the oxidized residue type. The MASCOT file frequently reported oxidized peptides at the same *m/z* value with similar RTs that gave multiple matches to the same peak in the Mass Hunter ESI file, albeit oxidized at different residue positions. The matched peak was partitioned between these overlapping peptides ($\text{Pept}_{\text{Ox}}^{\text{overlap}}$), whenever $\Sigma AA > k$. In such cases, for a peak with area $P_O A^1$ (Figure A2).
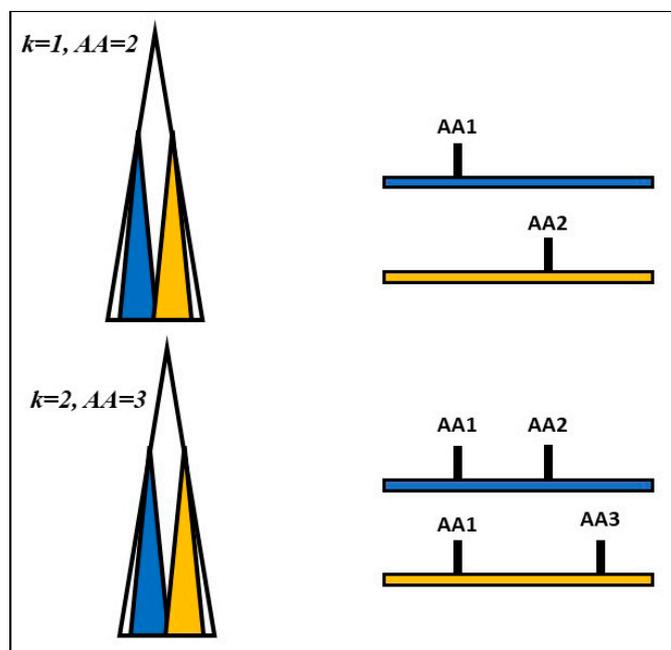

Figure A2. Examples of oxidized peaks (Po's) with $\Sigma AA > k$ for 2 overlapping peptides ($\Sigma\text{Pept}_{\text{Ox}}^{\text{overlap}} = 2$). (a) k = 1, $\Sigma AA = 2$. (b) k = 2, $\Sigma AA = 3$. For $\Sigma\text{Pept}_{\text{Ox}}^{\text{overlap}} = 3$ or >, the equations have extra terms, but the logic is the same. Overlapping peptides do not occur for unoxidized peaks (Pu's).

$$k = 1, \Sigma AA = 2$$

$$P_O A^1 = P_O AA^1 + P_O AA^2$$

$$P_O AA^1 = P_O A^1 \times (AA^{IR1}/(AA^{IR1} + AA^{IR2}))$$

$$P_O AA^2 = P_O A^1 \times (AA^{IR2}/(AA^{IR1} + AA^{IR2}))$$

$$k = 2, \Sigma AA = 3$$

$$P_O A^1 = Y + Z$$

$$Y = (AA^{IR1} \times AA^{IR2})/((AA^{IR1} \times AA^{IR2}) + (AA^{IR1} \times AA^{IR3}))$$

$$Z = (AA^{IR1} \times AA^{IR3})/((AA^{IR1} \times AA^{IR2}) + (AA^{IR1} \times AA^{IR3}))$$

Then,

$$P_O AA^1 = P_O A^1$$
$$P_O AA^2 = P_O A^1 \times ((Y)/(Y + Z))$$
$$P_O AA^3 = P_O A^1 \times ((Z)/(Y + Z))$$

2.   Peaks List

A list of the unoxidized and oxidized peaks for each detected residue position was generated for a peptide. The process was iterated over all peptides. A complete list of the peaks for all residue positions over all peptides was generated for subsequent analysis.

3.   Residue Protection Factor (PF) analysis

The oxidized fraction for any amino acid (AA^i) position at any "Dose" is $F_{AA}{}^i$.

$$F_{AA}{}^i = \Sigma P_O AA^i/\{(\Sigma Pu) + (\Sigma Po)\}$$

where $\Sigma P_O AA^i = \sum_{\frac{m}{z}=1}^{n} P_O AA^i$, $\Sigma Pu = \sum_{\frac{m}{z}=1}^{n} Pu$, $\Sigma Po = \sum_{\frac{m}{z}=1}^{n} \sum P_O AA^i$.

The unoxidized fraction ($1-F_{AA}{}^i$) for any amino acid position (AA^i), weighted with the inverse of the standard error ($SD/(N_R)^{1/2}$), is then plotted as a function of "Dose". The oxidation rates, S, were obtained from the best single exponential fits to the plots. The Log (PF) values, computed for each AA, were then plotted against the AA SASA obtained from the Input 3D structures. Anomalous, low (PFs) AA positions (ALPFs) had values < the lower 95% confidence predicted for the AA SASA from the plot best-fit linear regression.

References

1.   Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. In *Advances in Protein Chemistry*; Elsevier: Amsterdam, The Netherlands, 1959; Volume 14, pp. 1–63. [CrossRef]
2.   Papoian, G.A.; Ulander, J.; Eastwood, M.P.; Luthey-Schulten, Z.; Wolynes, P.G. Water in protein structure prediction. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3352–3357. [CrossRef] [PubMed]
3.   Tsai, C.-J.; Nussinov, R. A Unified View of "How Allostery Works". *PLoS Comput. Biol.* **2014**, *10*, e1003394. [CrossRef] [PubMed]
4.   Samways, M.L.; Taylor, R.D.; Macdonald, H.E.B.; Essex, J.W. Water molecules at protein–drug interfaces: Computational prediction and analysis methods. *Chem. Soc. Rev.* **2021**, *50*, 9104–9120. [CrossRef] [PubMed]
5.   Spyrakis, F.; Ahmed, M.H.; Bayden, A.S.; Cozzini, P.; Mozzarelli, A.; Kellogg, G.E. The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *J. Med. Chem.* **2017**, *60*, 6781–6827. [CrossRef]
6.   Modi, V.; Dunbrack, R.L., Jr. Defining a New Nomenclature for the Structures of Active and Inactive Kinases. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6818–6827. [CrossRef]
7.   McClendon, C.L.; Kornev, A.P.; Gilson, M.K.; Taylor, S.S. Dynamic architecture of a protein kinase. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4623–E4631. [CrossRef]

8.   Xu, G.; Chance, M.R. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **2007**, *107*, 3514–3543. [CrossRef]

9.   Gupta, S.; Feng, J.; Chan, L.J.G.; Petzold, C.J.; Ralston, C.Y. Synchrotron X-ray footprinting as a method to visualize water in proteins. *J. Synchrotron Radiat.* **2016**, *23*, 1056–1069. [CrossRef]

10.   Gupta, S.; Merriman, C.; Petzold, C.J.; Ralston, C.Y.; Fu, D. Water molecules mediate zinc mobility in the bacterial zinc diffusion channel ZIPB. *J. Biol. Chem.* **2019**, *294*, 13327–13335. [CrossRef]

11.   AAngel, T.E.; Gupta, S.; Jastrzebska, B.; Palczewski, K.; Chance, M.R. Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 14367–14372. [CrossRef]

12.   Gustavsson, M.; Wang, L.; Van Gils, N.; Stephens, B.; Zhang, P.; Schall, T.J.; Yang, S.; Abagyan, R.; Chance, M.R.; Kufareva, I.; et al. Structural basis of ligand interaction with atypical chemokine receptor 3. *Nat. Commun.* **2017**, *8*, 14135. [CrossRef] [PubMed]

13.   Gupta, S.; D'Mello, R.; Chance, M.R. Structure and dynamics of protein waters revealed by radiolysis and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14882–14887. [CrossRef] [PubMed]

14.   Huang, W.; Ravikumar, K.M.; Chance, M.R.; Yang, S. Quantitative Mapping of Protein Structure by Hydroxyl Radical Footprinting-Mediated Structural Mass Spectrometry: A Protection Factor Analysis. *Biophys. J.* **2015**, *108*, 107–115. [CrossRef] [PubMed]

15.   Wang, L.; Chance, M.R. Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Mol. Cell. Proteom.* **2017**, *16*, 706–716. [CrossRef] [PubMed]

16.   Wheatley, P.; Pandini, A.; Gupta, S.; Chen, Y.; Petzold, C.J.; Ralston, C.Y.; Blair, D.F.; Khan, S.M. Allosteric Priming of E. Coli CheY by the Flagellar Motor Protein FliM. *bioRxiv* **2020**, *119*, 1108–1122. [CrossRef]

17.   Gao, R.; Bouillet, S.; Stock, A.M. Structural Basis of Response Regulator Function. *Annu. Rev. Microbiol.* **2019**, *73*, 175–197. [CrossRef]

18.   Dyer, C.M.; Dahlquist, F.W. Switched or Not?: The Structure of Unphosphorylated CheY Bound to the N Terminus of FliM. *J. Bacteriol.* **2006**, *188*, 7354–7363. [CrossRef]

19.   Dyer, C.M.; Quillin, M.L.; Campos, A.; Lu, J.; McEvoy, M.M.; Hausrath, A.C.; Westbrook, E.M.; Matsumura, P.; Matthews, B.W.; Dahlquist, F.W. Structure of the Constitutively Active Double Mutant CheYD13K Y106W Alone and in Complex with a FliM Peptide. *J. Mol. Biol.* **2004**, *342*, 1325–1335. [CrossRef]

20.   Volz, K.; Matsumura, P. Crystal structure of Escherichia coli CheY refined at 1.7-A resolution. *J. Biol. Chem.* **1991**, *266*, 15511–15519. [CrossRef]

21.   Basharat, A.R.; Iman, K.; Khalid, M.F.; Anwar, Z.; Hussain, R.; Kabir, H.G.; Tahreem, M.; Shahid, A.; Humayun, M.; Hayat, H.A.; et al. SPECTRUM—A MATLAB Toolbox for Proteoform Identification from Top-Down Proteomics Data. *Sci. Rep.* **2019**, *9*, 11267. [CrossRef]

22.   Stock, A.; Martinez-Hackert, E.; Rasmussen, B.F.; West, A.H.; Stock, J.B.; Ringe, D.; Petsko, G.A. Structure of the magnesium-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry* **1993**, *32*, 13375–13380. [CrossRef] [PubMed]

23.   Guhaniyogi, J.; Robinson, V.L.; Stock, A.M. Crystal Structures of Beryllium Fluoride-free and Beryllium Fluoride-bound CheY in Complex with the Conserved C-terminal Peptide of CheZ Reveal Dual Binding Modes Specific to CheY Conformation. *J. Mol. Biol.* **2006**, *359*, 624–645. [CrossRef] [PubMed]

24.   Bourret, R.B. Receiver domain structure and function in response regulator proteins. *Curr. Opin. Microbiol.* **2010**, *13*, 142–149. [CrossRef]

25.   Ferreiro, D.U.; Hegler, J.A.; Komives, E.A.; Wolynes, P.G. On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3499–3503. [CrossRef]

26.   Freiberger, M.I.; Guzovsky, A.B.; Wolynes, P.G.; Parra, R.G.; Ferreiro, D.U. Local frustration around enzyme active sites. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4037–4043. [CrossRef] [PubMed]

27.   Lee, S.-Y.; Cho, H.S.; Pelton, J.G.; Yan, D.; Henderson, R.K.; King, D.S.; Huang, L.; Kustu, S.; Berry, E.A.; Wemmer, D.E. Crystal Structure of an Activated Response Regulator Bound to Its Target. *Nat. Struct. Biol.* **2001**, *8*, 52–56. [CrossRef]

28.   Smith, J.G.; Latiolais, J.A.; Guanga, G.P.; Citineni, S.; Silversmith, R.E.; Bourret, R.B. Investigation of the Role of Electrostatic Charge in Activation of the *Escherichia coli* Response Regulator CheY. *J. Bacteriol.* **2003**, *185*, 6385–6391. [CrossRef]

29.   Foster, C.A.; Silversmith, R.E.; Immormino, R.M.; Vass, L.R.; Kennedy, E.N.; Pazy, Y.; Collins, E.J.; Bourret, R.B. Role of Position K+4 in the Phosphorylation and Dephosphorylation Reaction Kinetics of the CheY Response Regulator. *Biochemistry* **2021**, *60*, 2130–2151. [CrossRef]

30.   Pilizota, T.; Brown, M.T.; Leake, M.C.; Branch, R.W.; Berry, R.M.; Armitage, J.P. A molecular brake, not a clutch, stops the *Rhodobacter sphaeroides* flagellar motor. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11582–11587. [CrossRef]

31.   Rossmann, F.M.; Hug, I.; Sangermani, M.; Jenal, U.; Beeby, M. In situ structure of the *Caulobacter crescentus* flagellar motor and visualization of binding of a CheY-homolog. *Mol. Microbiol.* **2020**, *114*, 443–453. [CrossRef]

32.   Armstrong, C.T.; Mason, P.; Anderson, R.; Dempsey, C.E. Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Sci. Rep.* **2016**, *6*, 21759. [CrossRef] [PubMed]

33.   Mayol, E.; García-Recio, A.; Tiemann, J.K.S.; Hildebrand, P.W.; Guixà-González, R.; Olivella, M.; Cordomí, A. HomolWat: A web server tool to incorporate 'homologous' water molecules into GPCR structures. *Nucleic Acids Res.* **2020**, *48*, W54–W59. [CrossRef] [PubMed]

34. Krishnamurthy, S.; Eleftheriadis, N.; Karathanou, K.; Smit, J.H.; Portaliou, A.G.; Chatzi, K.E.; Karamanou, S.; Bondar, A.-N.; Gouridis, G.; Economou, A. A nexus of intrinsic dynamics underlies translocase priming. *Structure* **2021**, *29*, 846–858. [CrossRef] [PubMed]

35. Chen, M.; Chen, X.; Schafer, N.P.; Clementi, C.; Komives, E.A.; Ferreiro, D.U.; Wolynes, P.G. Surveying biomolecular frustration at atomic resolution. *Nat. Commun.* **2020**, *11*, 5944. [CrossRef] [PubMed]

36. Davtyan, A.; Schafer, N.P.; Zheng, W.; Clementi, C.; Wolynes, P.G.; Papoian, G.A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503. [CrossRef] [PubMed]

37. Parra, R.G.; Schafer, N.P.; Radusky, L.G.; Tsai, M.-Y.; Guzovsky, A.B.; Wolynes, P.G.; Ferreiro, D.U. Protein Frustratometer 2: A tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* **2016**, *44*, W356–W360. [CrossRef]

38. Siemers, M.; Bondar, A.-N. Interactive Interface for Graph-Based Analyses of Dynamic H-Bond Networks: Application to Spike Protein S. *J. Chem. Inf. Model.* **2021**, *61*, 2998–3014. [CrossRef]

39. Karathanou, K.; Lazaratos, M.; Bertalan, É.; Siemers, M.; Buzar, K.; Schertler, G.F.X.; Del Val, C.; Bondar, A.-N. A Graph-Based Approach Identifies Dynamic H-Bond Communication Networks in Spike Protein S of SARS-CoV-2. *J. Struct. Biol.* **2020**, *212*, 107617. [CrossRef]

40. Mihalek, I.; Reš, I.; Lichtarge, O. A Family of Evolution–Entropy Hybrid Methods for Ranking Protein Residues by Importance. *J. Mol. Biol.* **2004**, *336*, 1265–1282. [CrossRef]

41. UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

42. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]

43. Xu, A.G.; Chance, M.R. Radiolytic Modification and Reactivity of Amino Acid Residues Serving as Structural Probes for Protein Footprinting. *Anal. Chem.* **2005**, *77*, 4549–4555. [CrossRef] [PubMed]