

ARTICLE

<https://doi.org/10.1038/s42003-019-0748-0>

OPEN

# Structure and dynamics of the *E. coli* chemotaxis core signaling complex by cryo-electron tomography and molecular simulations

C. Keith Cassidy <sup>1,2\*</sup>, Benjamin A. Himes<sup>3</sup>, Dapeng Sun <sup>3</sup>, Jun Ma<sup>3</sup>, Gongpu Zhao <sup>3</sup>, John S. Parkinson <sup>4</sup>, Phillip J. Stansfeld <sup>1,5</sup>, Zaida Luthey-Schulten <sup>2,6</sup> & Peijun Zhang <sup>3,7,8\*</sup>

To enable the processing of chemical gradients, chemotactic bacteria possess large arrays of transmembrane chemoreceptors, the histidine kinase CheA, and the adaptor protein CheW, organized as coupled core-signaling units (CSU). Despite decades of study, important questions surrounding the molecular mechanisms of sensory signal transduction remain unresolved, owing especially to the lack of a high-resolution CSU structure. Here, we use cryo-electron tomography and sub-tomogram averaging to determine a structure of the *Escherichia coli* CSU at sub-nanometer resolution. Based on our experimental data, we use molecular simulations to construct an atomistic model of the CSU, enabling a detailed characterization of CheA conformational dynamics in its native structural context. We identify multiple, distinct conformations of the critical P4 domain as well as asymmetries in the localization of the P3 bundle, offering several novel insights into the CheA signaling mechanism.

<sup>1</sup> Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. <sup>2</sup> Department of Physics and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>3</sup> Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260, USA. <sup>4</sup> School of Biological Sciences, University of Utah, Salt Lake City, UT 84112, USA. <sup>5</sup> School of Life Sciences & Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK. <sup>6</sup> Department of Chemistry and Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. <sup>7</sup> Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. <sup>8</sup> Electron Bio-Imaging Centre, Diamond Light Sources, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK. \*email: [keith.cassidy@bioch.ox.ac.uk](mailto:keith.cassidy@bioch.ox.ac.uk); [peijun@strubi.ox.ac.uk](mailto:peijun@strubi.ox.ac.uk)

Motile bacteria are able to seek out optimal physiological conditions through a behavior known as chemotaxis, where highly conserved signal transduction pathways couple environmental chemical gradients to cellular motility<sup>1,2</sup>. The streamlined chemotaxis pathway of *Escherichia coli* is especially well studied and provides a powerful tool for investigating the molecular underpinnings of sensory signal transduction<sup>3–5</sup>.

Input to the *E. coli* chemotaxis pathway is mediated by transmembrane chemoreceptors, which operate as homodimers and respond to specific sets of external chemoeffectors<sup>4,6</sup>. Upon ligand occupancy changes, chemoreceptors transmit signals across the inner membrane that modulate the autophosphorylation activity of a homodimeric histidine kinase CheA, which is coupled to receptor control by an adaptor protein CheW. In addition to ligand binding, the methylation state of four cytoplasmic glutamyl or glutaminyl residues, denoted QEQE in wild-type receptors, adjusts output activity<sup>6</sup>. Addition or removal of methyl groups at these sites via specific enzymes, modulates the overall level of CheA activity to offset the effects of ligand binding inputs. Importantly, Q residues mimic the effects of methylated E residues on CheA activity, so a receptor with all four modification sites as glutamine (4Q) promotes high CheA activity, whereas a 4E receptor produces low kinase activity<sup>7–9</sup>.

Each CheA monomer comprises five domains, P1–P5. The P3, P4, and P5 domains are connected by two short linkers<sup>10</sup> and enable dimerization, ATP hydrolysis, and binding to chemoreceptors and CheW, respectively. The P1 and P2 domains, which are connected to the P3–P4–P5 dimer core by long flexible linkers, mediate the transfer of phosphoryl groups from P4 to P1 and from there to response regulators<sup>11,12</sup>. The CheY response regulator binds to the flagellar motors upon phosphorylation to affect their rotational bias and ultimately cellular swimming pattern.

CheA regulation requires the assembly of core-signaling units (CSUs)<sup>4,13,14</sup> with a well-defined structure, comprising two chemoreceptor trimer-of-dimers (TOD), a single CheA dimer and two CheW monomers (Fig. 1d). Upon formation of the CSU, CheA undergoes pronounced tertiary rearrangement<sup>15–17</sup>, increasing its basal activity nearly three orders of magnitude<sup>12,18</sup>. CSUs further cluster into large, hexagonally packed chemosensory arrays, which place each CheA under the influence of as many as 20 receptors, giving rise to highly cooperative signaling responses<sup>15,19–21</sup>. The basic organizational hierarchy of the CSU and the extended hexagonal architecture it forms appear to be universally conserved features of bacterial and archaeal chemotaxis pathways<sup>22,23</sup>. Considering the central role that chemotaxis plays in the infection process of numerous human and plant pathogens<sup>24–27</sup>, a detailed characterization of CSU structure and function would, therefore, provide an excellent source of novel targets for the development of broadly applicable antimicrobial therapies.

Despite a considerable effort, the large size and dynamic nature of the CSU have thwarted the determination of a high-resolution structure via conventional structural biology techniques such as X-ray crystallography, nuclear magnetic resonance, and single-particle cryo-electron microscopy. Previous studies involving cellular cryo-electron tomography (cryoET) have provided snapshots of native chemosensory arrays in varying signaling states, considerably advancing our understanding of chemosensory array architecture, assembly, and function<sup>15,19,22,28,29</sup>. However, cell thickness, even that of minicells, typically limits the resolution of in vivo structures to 20 Å or lower<sup>30</sup>. Hence, detailed descriptions of the molecular mechanisms underlying sensory signal transduction, especially that of receptor-mediated CheA control, remain elusive.

Here, we present in vitro cryoET structures of the QEQE, 4Q, and 4E *E. coli* CSU, derived using our recently developed

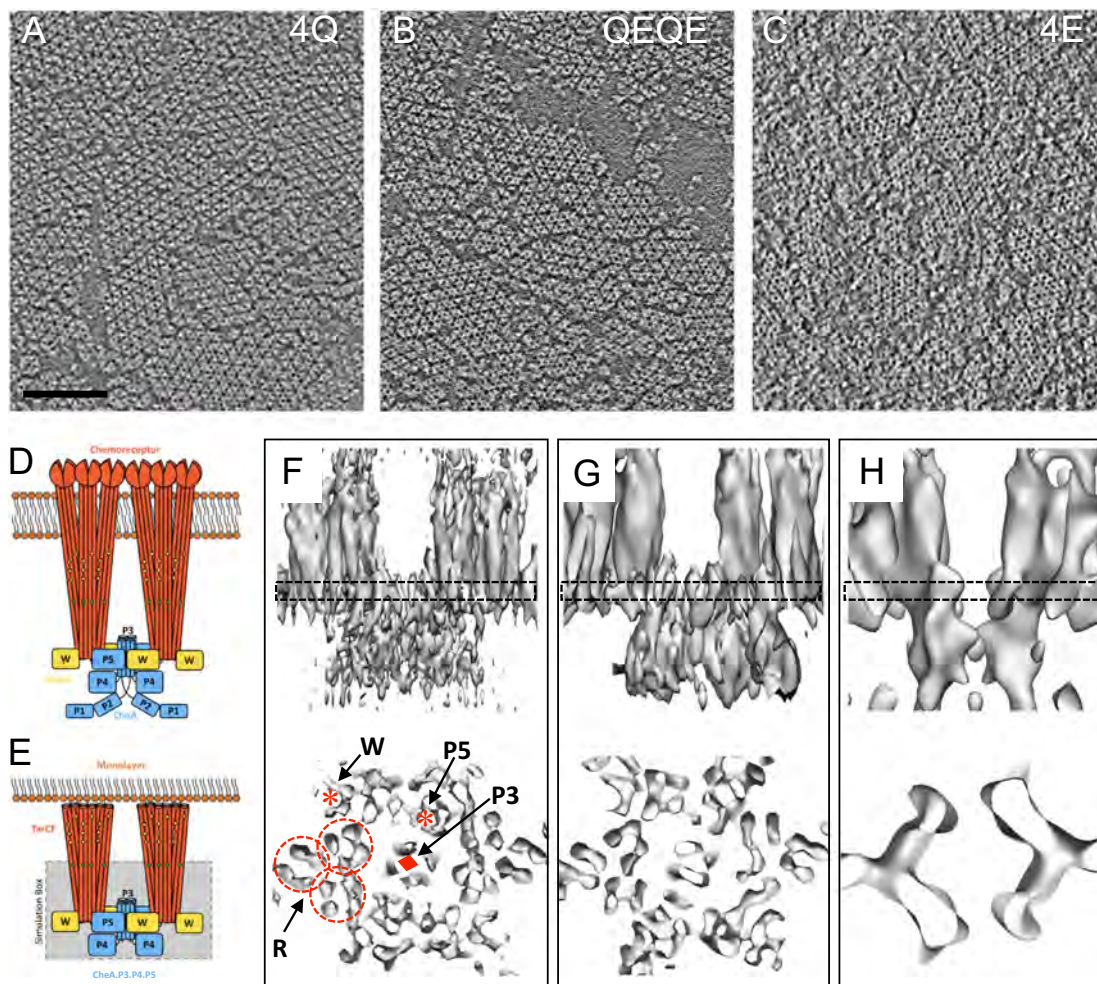
emClarity software<sup>31</sup> package for high-resolution cryoET and sub-tomogram averaging (STA). For the 4Q system, we achieve an overall sub-nanometer resolution, which although anisotropic, enables the construction of an all-atom model of the *E. coli* CSU. Using molecular simulations, we further investigate the conformational landscape of CheA within the CSU, identifying multiple, distinct conformations of the critical CheA.P4 domain as well as asymmetric conformations of the CheA.P3 bundle, highlighting stabilizing features of each. The implications of our results for the CheA signaling mechanism are discussed.

## Results

**CryoET structures of in vitro reconstituted chemosensory arrays.** We previously developed an in vitro reconstitution system comprising *E. coli* CheA.P3–P4–P5, CheW, and a His-tagged cytoplasmic domain of wild-type Tar (TarCF) bound onto Ni<sup>2+</sup>-NTA lipid containing monolayers (Fig. 1e)<sup>16,32,33</sup>. This system produces samples that are suitable for the high-resolution structural analysis of chemosensory arrays by cryoET. In particular, compared to native cellular arrays, in vitro monolayer arrays are thin (~25 nm), compositionally well defined, and homogeneous. Moreover, the abundance of in vitro arrays allows one to extract a few thousand CSUs from a single tomogram, enabling the use of STA to obtain high-resolution structures of the CSU. We have now reconstituted in vitro monolayer systems using 4Q or 4E variants of the aspartate receptor signaling domain (TarCF), giving rise to TarCF<sub>4Q</sub>-CheA.P3.P4.P5-CheW and TarCF<sub>4E</sub>-CheA.P3.P4.P5-CheW monolayer arrays, which should display high and low kinase activity, respectively.

As depicted in Fig. 1a–c, 4Q arrays appear better ordered compared to QEQE arrays (i.e., show larger patches of extended, well-packed CSUs), and both appear highly ordered compared to 4E arrays. The decreased long-range order observed in 4E arrays is generally consistent with previous studies that show the CheA-OFF array state exhibits reduced cooperativity<sup>8,34,35</sup>, likely through the disruption of interactions between CSUs that alter the tightness of their physical and functional coupling<sup>36</sup>. While in vivo cryoET has previously shown that arrays remain intact and hexagonally packed during signaling<sup>37</sup>, the detailed long-range order of CSU packing has not been examined at the tomogram level, making it difficult to say whether in vivo arrays might not also exhibit variations in the size of well-packed patches. At present, however, we cannot rule out the possibility that the differences in order observed in the monolayer array may arise or appear more pronounced due to their in vitro nature. Therefore, the signaling role played by such changes, if any, will require additional study.

We collected 24, 22, and 19 cryo-tomography tilt series of the 4Q, QEQE, and 4E monolayer arrays, respectively (Supplementary Table 1). CSUs were identified within each tomogram using template matching as implemented in emClarity (Supplementary Fig. 1). The resulting CSU sub-tomograms were extracted and processed with emClarity using an iterative approach to refine both the tilt geometry and alignment parameters of each sub-tomogram<sup>31</sup>. Further classification in emClarity was performed to differentiate CSU densities from the neighboring dimers of receptor TODs that lacked CheA density. The three-dimensional (3D) density maps of the 4Q, QEQE, and 4E CSUs derived from STA are shown in Fig. 1 (f–h); the unbiased Fourier shell correlations (FSCs) indicate the overall resolutions of these density maps are 8.4 Å, 10.1 Å, and 14.5 Å, respectively (Supplementary Fig. 2a). Clearly, the relatively well-ordered 4Q and QEQE arrays have produced considerably better-resolved density maps than the disordered 4E arrays.

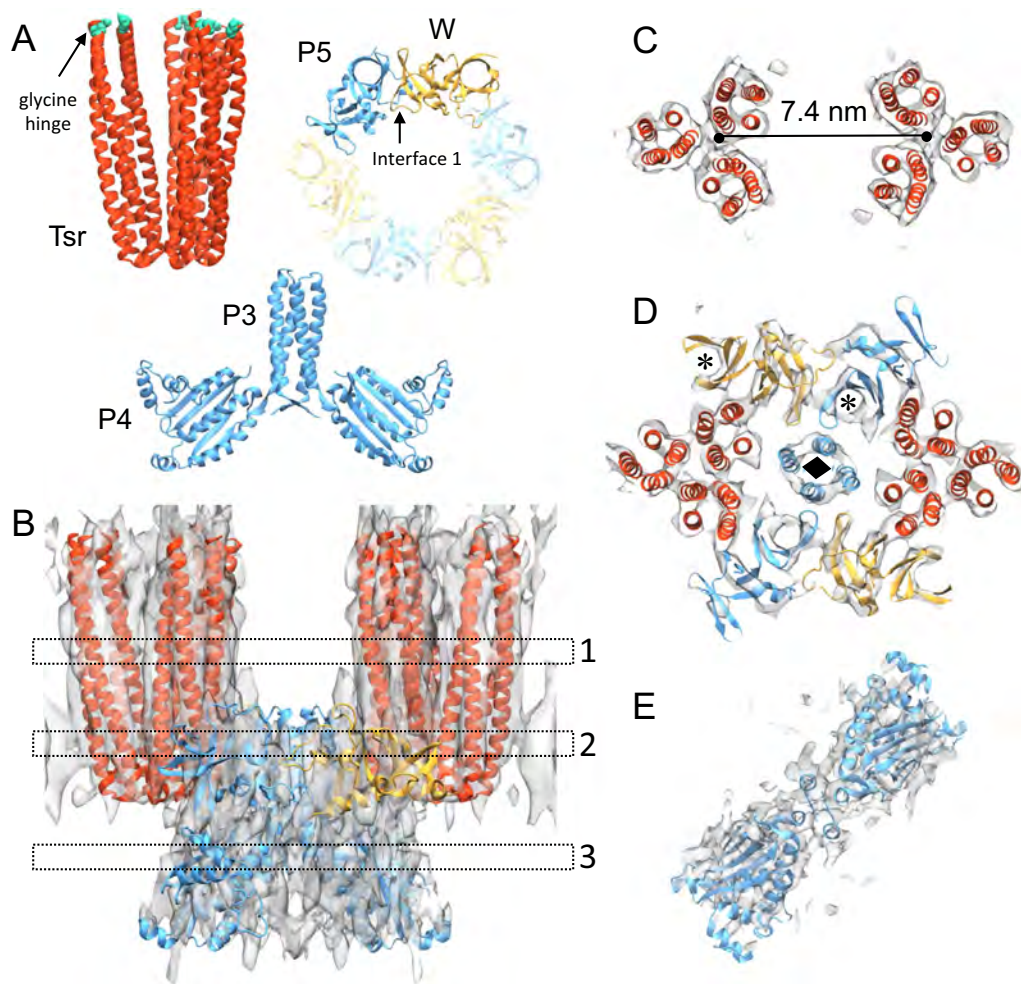


**Fig. 1 Cryo-electron tomography and sub-tomogram averaging of in vitro reconstituted chemosensory arrays.** **a–c** Raw tomographic slices of monolayer arrays of TarCF<sub>4Q</sub>/CheA/CheW **a**, TarCF<sub>QEQE</sub>/CheA/CheW **b**, and TarCF<sub>4E</sub>/CheA/CheW **c**. Scale bar, 100 nm. **d, e** Schematics of the transmembrane **d** and in vitro monolayer **e** core-signaling unit (CSU), consisting of two receptor trimer-of-dimers in red, a single CheA dimer in blue, and four CheW monomers—two essential (bound to CheA.P5) and two ancillary (flanking)—in gold. In **e**, a grey box denotes the region of the CSU modeled and simulated in this study. Receptor modification sites, the glycine hinge, and membrane headgroups are depicted as yellow, teal, and orange circles, respectively. **f–h** Sub-tomogram averages of CSU of TarCF<sub>4Q</sub>/CheA/CheW **f**, TarCF<sub>QEQE</sub>/CheA/CheW **g**, and TarCF<sub>4E</sub>/CheA/CheW **h**. 3D volumes are displayed in surface rendering as side views (top) and top views (bottom). Top views are clipped from the boxed regions surrounding the CheA/CheW baseplate region. In **f**, red asterisks indicate the CheA.P5 and CheW beta barrels, a red diamond indicates the CheA.P3 four-helix bundle, and dashed circles indicate the individual receptor dimers within a trimer.

At the current resolution, the 4Q and QEQE CSU maps show marginal differences overall (Supplementary Movie 1). The high resolution of the 4Q density map, however, reveals several secondary structure features. These include individual alpha helices in four-helix bundles of the chemoreceptor dimers (Fig. 1f, dashed circles) and the CheA.P3 dimerization domain (Fig. 1f, diamond), as well as beta barrels in both the CheA.P5 domain and CheW (Fig. 1f, asterisks). Moreover, as observed previously in the QEQE monolayer system<sup>16</sup>, all three receptors within each TOD, have a close interaction with either CheA.P5 or CheW in a ratio of 1TOD:1CheA:2CheW, confirming the existence of an additional CheW monomer flanking each TOD (Supplementary Fig. 3d, Fig. 1d). In the context of the extended array, these additional CheW monomers form rings that act to network neighboring CSUs<sup>16,19</sup>, thus, reinforcing the notion that CheW plays an important functional role outside the CSU, namely through the formation of complete or partial CheW-only rings that contribute to the ultra-stability and high cooperativity of the chemosensory array<sup>38,39</sup>. The least resolved portion of the 4Q map corresponds to the CheA.P4 domains, indicating

substantial conformational heterogeneity in this region of the CSU.

**Atomistic model of the *E. coli* CSU.** To enable a detailed investigation of CSU structure and dynamics, we constructed an atomistic model of the *E. coli* complex, using the 8.4 Å-resolution 4Q density map to explicitly guide the modeling process<sup>40</sup>. Owing to their lower resolution, the QEQE and 4E maps did not provide additional structural information and were not used for subsequent modeling. Component models of the receptor TOD, CheA.P3.P4 dimer, and CheA.P5-CheW (interface 1) dimer were first generated to make the best use of existing structural information at protein–protein interfaces (Fig. 2a). As no high-resolution structures exist of the cytoplasmic portion of *E. coli* Tar, we instead used a crystal structure of the *E. coli* Tsr TOD (PDB 1QU7<sup>41</sup>), which shares a 96% sequence similarity with Tar. Models of the *E. coli* CheA.P3.P4 dimer and the CheA.P5-CheW dimer were generated using template-based homology modeling from crystal structures of *Thermotoga maritima* CheA and CheW



**Fig. 2 All-atom model of the *E. coli* CSU.** **a** Component models used for initial CSU assembly, including the protein interaction region of the Tsr receptor trimer-of-dimers (TOD) and homology models of the CheA.P3-P4 and CheA.P5-CheW interface 1 dimers. Receptors are colored in red, CheA in blue, and CheW in gold. The residues comprising the glycine hinge are shown in teal. **b** Overlay between the 4Q density map, shown in surface rendering with transparent grey, and the MDFF-refined CSU model. **c–e** Sectional views of the boxed regions in **b** (1–3), highlighting the quality-of-fit at the level of the receptor TODs (1), CheA-CheW baseplate (2), and CheA.P4 domains (3). The distance between the symmetry axes of the receptor TODs is denoted in **c**. In **d**, black asterisks indicate the CheA.P5 and CheW beta barrels; a black diamond indicates the CheA.P3 four-helix bundle.

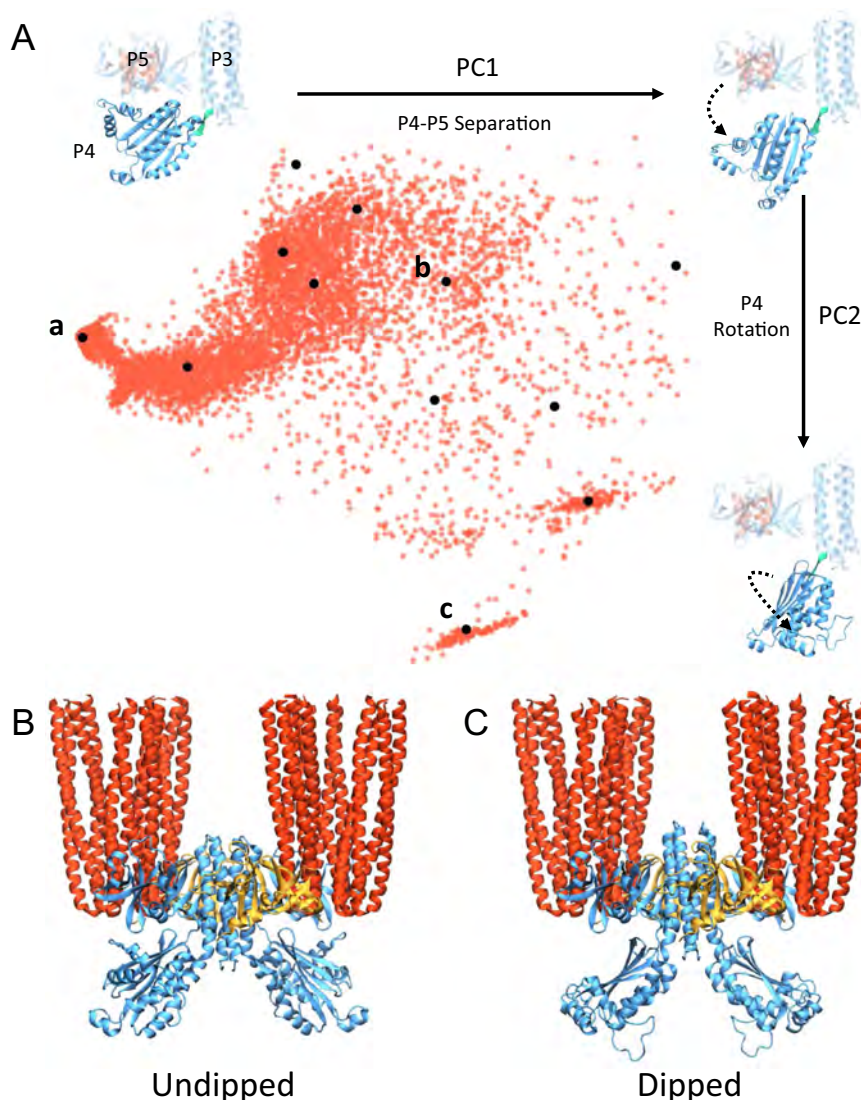
(PDBs 4XIV<sup>17</sup> and 4JPB<sup>42</sup>, respectively), which are sufficiently similar (75% sequence similarity for CheA.P3.P4.P5; 55% for CheW) to permit the construction of reliable models (Methods section). Segmentation of the 4Q map into sub-densities corresponding to each component then provided a basis for their initial rigid docking.

The best-resolved areas of density belong to the receptor TODs, where each receptor can be clearly differentiated, and individual helices are discernable up to the glycine hinge (Fig. 2b). We began assembling the CSU by optimizing the position and conformation of the protein interaction region (residues 340–441) of the receptor TODs using Molecular Dynamics Flexible Fitting (MDFF)<sup>43,44</sup>. The resulting fits minimally altered the conformation of the initial models (backbone root-mean-squared-deviation (RMSD) = 1.8 Å) and oriented the two TODs symmetrically with a separation of 7.4 nm (Fig. 2c). This spacing suggests an extended hexagonal lattice constant of ~12.8 nm, between the 12 nm<sup>22</sup> and 13.2 nm<sup>19</sup> spacings previously reported. To test the robustness of the obtained fit, especially considering resolution anisotropy (Supplementary Fig. 2b), we generated different TOD starting positions by translations of up to ±15 Å along the Z-direction and refitted the models. In all cases, the

TOD structure converged to the same position within the density (Supplementary Fig. 4, Supplementary Movie 2).

The densities corresponding to CheA.P5 and CheW were also well resolved, showing distinct beta barrel organizations coinciding with subdomain 1 of each protein. We docked CheA.P5 and CheW independently without constraints on interface 1. The resulting positioning produced a nearly identical configuration to that seen in crystal structures<sup>15,42</sup> (backbone RMSD = 1.4 Å), suggesting this interface 1 binding mode is prominent in the 4Q map. Finally, docking the CheA.P3.P4 dimer model produced a nonoverlapping structure while also situating the P4 domain such that it could be readily connected to the P5 domain via the short P4–P5 linker. The entire complex was then subjected to MDFF, producing minor refinements to the positioning of the CheA.P5-CheW and CheA.P3.P4 models. The resulting CSU model is shown in Fig. 2(b–e) and Supplementary Movie 3.

**CheA.P4 adopts multiple, distinct conformations within the CSU.** A growing body of evidence suggests that CheA is highly dynamic, undergoing pronounced interdomain rearrangements during CSU assembly and throughout its catalytic cycle<sup>11,17,45,46</sup>.



**Fig. 3 CheA.P4 adopts multiple distinct conformations within the CSU.** **a** Projection of the 9247 GSA-derived P4 conformations onto the first two PCs of the ensemble (red dots). Black dots denote the 12 medoid structures identified by clustering analysis of the ensemble. Three medoid projections and corresponding structures are shown, illustrating the primary degree of freedom represented by each PC. In each structure, the portion held static during the GSA simulations is shown with a transparent representation and the P3–P4 hinge (residues 326–328) about which rigid-body rotations were made is shown in teal. **b–c** CSU models containing a representative MDFF-refined undipped **b** and dipped **c** CheA dimer.

Indeed, within the 4Q map, the density corresponding to the CheA.P4 domain is considerably less resolved, suggesting substantial conformational heterogeneity. We previously used molecular dynamics (MD) simulations to identify an alternative P4 conformation in *T. maritima* CheA, which we termed dipped and subsequently validated in live *E. coli* cells<sup>16</sup>. Thus, to systematically explore the conformational landscape of *E. coli* CheA in its native structural context, we used generalized simulated annealing (GSA)<sup>47,48</sup> to generate an ensemble of P4 positions within our CSU model through rigid-body rotations about the P3–P4 hinge (residues 326–328). The full details of our GSA protocol are provided in the Methods section.

We analyzed a total of 9247 GSA-derived structures. Figure 3a shows each structure projected onto the first two principal components (PCs) of the full ensemble, which capture 89% of its total variance. The resulting distribution clearly illustrates that as P4 becomes increasingly separated from P5 (movement along PC1), it is able to adopt a much wider range of rotational states (movement along PC2) in which the active site and ATP-lid are

markedly reoriented with respect to the rest of the CSU. This suggests that the regulatory ability of P5 may arise from its capacity to constrain P4 to a highly specific orientation through direct contact, which if disrupted could allow P4 to access a much broader conformational landscape. Indeed, construction of CheA dimer models based on the GSA ensemble shows that P4–P4 separation, as measured by the distance between active sites, can vary substantially (~35–70 Å; Supplementary Fig. 5). In particular, at distances of less than ~40 Å, the P4 domains come close enough to directly interact (Supplementary Fig. 5d), a possibility previously reported<sup>11,45</sup>.

We then considered whether the P4 density in our 4Q map could account for additional P4 conformations. To define a subset of the GSA-derived conformations that preserved the structural variability of the full ensemble, hierarchical clustering was used to group conformations based on pairwise RMSD<sup>49</sup>. This analysis resulted in 12 structures with a wide range of P4 positions (Supplementary Fig. 6). Symmetric CheA dimers constructed from each were then substituted into our CSU model and MDFF

was used to refit the P4 domains to the 4Q map. Strikingly, the resulting fits collapsed into just two classes of CheA conformation with similar P4 CCC (Supplementary Fig. 7a). We term these classes undipped and dipped to denote the relative position of P4 with respect to the rest of the CSU structure (Fig. 3b, c). A third, intermediate class was also observed (Supplementary Fig. 7b–d), which is qualitatively similar to the undipped class. We note that the symmetry of individual P4 monomers within the obtained conformational classes stems from the symmetric nature of the CSU density map, a product of the STA procedure, and does not imply that the P4 domains must move in unison.

The undipped structures essentially reproduce the initial MDFF-refined CheA conformation, showing a relatively planar P3–P4 configuration with considerable P4–P5 interaction. The dipped structures, on the other hand, do not exhibit a planar P3–P4 conformation, but rather show the P4 domains completely separated from P5 and rotated such that both active sites face downward and are separated by ~55 Å. These structures closely resemble the dipped conformation previously predicted in *T. maritima* CheA<sup>16</sup>, providing further evidence that it is a highly conserved, and likely integral, conformational feature of the CheA catalytic cycle. Considering the appreciable difference in orientation of the P4 active sites in the planar, undipped, and intermediate conformations compared to the rotated, dipped conformation, it's likely that transitions between these conformations play a role in regulating P1/P4 interactions necessary for phosphotransfer. Interestingly, Greenswag et al. have suggested that the inhibited state of CheA may involve close association between both P1 and P4 domains in a dimer<sup>17</sup>. Thus, the decreased separation between dipped P4 domains may provide a substrate for formation of inhibiting P1–P4 interactions across the dimer, which are not possible between undipped P4 conformations.

**Structural elements stabilizing alternate CheA.P4 conformations.** To assess the stability of the identified CheA classes and probe the dynamics of the CSU, we conducted 3 × 250-ns all-atom MD simulations of a representative CSU structure from the undipped and dipped CheA classes. The most striking differences are observed at the P3–P4 interface where an antiparallel beta sheet interaction involving residues M327 and L362 in the undipped state (Fig. 4a–b) was disrupted by GSA in the dipped state (Fig. 4c). This permits the formation of a continuous, albeit kinked, helical connection between the two domains during equilibration of the dipped CSU model (Fig. 4c), which appears to stabilize the dipped P4 conformation and is accompanied by a shift in salt-bridge patterns, involving several highly conserved residues (Supplementary Table 2, Supplementary Fig. 8a). In particular, in the undipped simulations, R325 forms salt bridges with D363 and D272, while K270 forms a weaker contact with E361 in several monomers (Fig. 4d); these interactions are in line with those previously seen in crystal structures of *T. maritima* CheA<sup>10,17</sup>, including a catalytically active P3–P4 dimer<sup>17</sup>. In the dipped simulations, however, the rotational orientation of the P4 domain disfavors the R325–D363 and K270–E361 interactions, leading instead to the formation of a salt bridge between R325 and E361 as well as increased interaction between K364 and D272 (Fig. 4e, Supplementary Table 2). Notably, these changes involve or directly affect the P3–P4 linker (residues 322–327) that was previously shown to be crucial for CheA activation and basal activity<sup>46,50</sup>, further suggesting that transitions between undipped and dipped P4 conformations play a critical role in the CheA catalytic cycle.

In addition, as noted previously, the degree of interaction between the P4 and P5 domains differs considerably between the

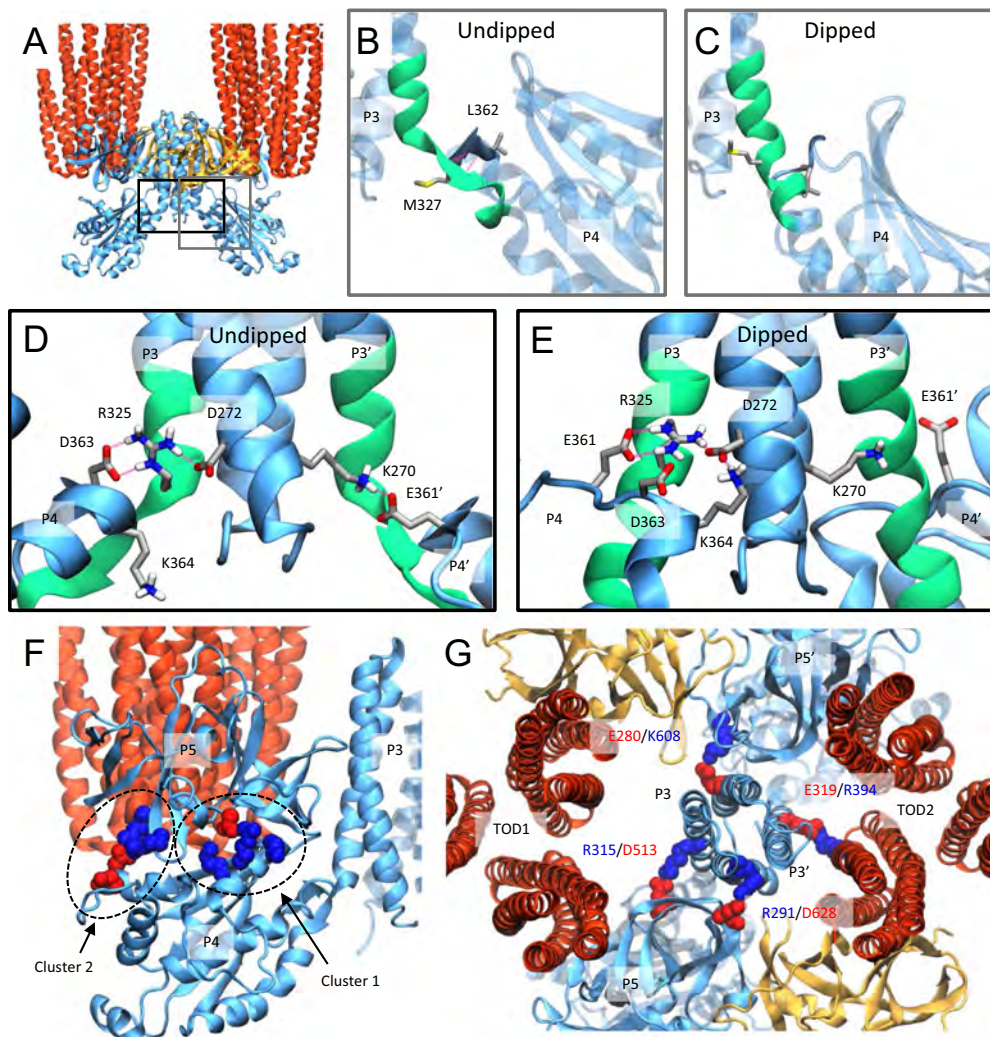
undipped and dipped conformations. In particular, whereas P4 makes few contacts with P5 in the dipped simulations, interactions between two residue clusters with complementary electrostatics are observed in the undipped simulations (Fig. 4f, Supplementary Table 2). However, although interactions between one or both clusters are often present, the specific interaction partners within a cluster vary. Considering this lack of specificity, as well as the imperfect conservation of certain cluster residues (Supplementary Fig. 8a), it's unlikely that any single P4–P5 interaction cited is critical for CheA function. Rather, both clusters of interactions appear to contribute to the general stabilization of the same planar P4–P5 binding mode, lending further support to the notion that P5 acts principally to constrain P4 orientation.

### CheA.P3 can adopt an asymmetric position within the CSU.

The simulation data also report on the dynamic behavior of the P3 domain. Despite the structural signatures described above, CheA does not adopt strictly symmetric conformations in simulations of either conformational class. In particular, the P3 four-helix bundle is observed to drift and/or lean toward one or the other receptor TOD in multiple undipped and dipped simulations (Supplementary Fig. 9). While these shifts are slight in terms of the center-of-mass movement of the P3 bundle, several asymmetric contacts with the receptors and P5 form repeatedly (Fig. 4g, Supplementary Table 2). In addition, the tight structural coupling between the P3 and P4 domains, both in the undipped and dipped conformations, cause asymmetries in the position of the P3 bundle to be transmitted directly to both P4 domains, giving rise to differing P4–P5 interactions within each monomer (Supplementary Movie 4). It is, thus, conceivable that changes in the position or orientation of the P3 bundle could disengage the P4 domains from other parts of the CSU, in particular P5. As a result, subtle signaling-related rearrangements in CSU structure could alter the aforementioned contacts and/or the effective space between TODs to regulate the magnitude of fluctuations in the position of the P3 bundle and, therefore, the interconversion between key conformations in the CheA catalytic cycle. We note that the receptors were positionally restrained in our simulations to prevent structural artifacts due to their truncation, thus, additional simulations will be required to work out the full details of the effects of receptor conformation on the observed changes in CheA.

### Receptors form strong interactions with CheW but not CheA.

**P5.** Characterization of the remaining protein–protein interfaces highlighted three strong contacts at the CheW–receptor interface, which were present in the outer subunits of the dimers in all simulations: R40–E402, D107–R404, and E18–R394 (Supplementary Fig. 10a, Supplementary Fig. 8b–c). Notably, receptor residues E402 and R404 were previously shown to be critical for stabilizing the CheA–ON state, an effect attributed primarily to a predicted E402–R404 inter-receptor salt bridge<sup>51</sup>, which our simulations confirm (Supplementary Fig. 10a). In contrast, the CheA.P5–receptor interface, which is rotated slightly compared to the CheW–receptor interface despite being pseudosymmetric (Supplementary Fig. 10b), does not appear to contain any strong or specific interactions between the two proteins. This is in line with previous studies that suggest the CheA.P5–receptor interface is weaker and only plays a passive role in receptor–CheA communication<sup>52</sup>. Interestingly, receptor residue R404 was shown to play a critical, yet unresolved, role in achieving high cooperativity in the extended array<sup>51</sup>. The D107–R404 salt bridge, which is located directly adjacent to interface 2, suggests an allosteric pathway also involving CheW residue R140 (Supplementary



**Fig. 4** Characterization of the undipped and dipped CheA conformational classes using all-atom MD simulation. **a–e** Structural changes at the CheA P3–P4 interface stabilizing alternate P4 conformations. **a** Gray and black rectangles denote the location of zoomed **b** and **c** (gray), and **d** and **e** (black) within the context of the larger core-signaling unit (CSU), shown here with an undipped CheA dimer. **b–c** Gray rectangles highlight an observed change in the secondary structure of the P3–P4 connector region (residues 320–332, shown in green), which forms a beta sheet interaction with nearby P4 residues in the undipped conformation **b** and a continuous helix in the dipped conformation **c**. **d–e** Alterations in salt-bridge patterns at the P3–P4 hinge, involving several highly conserved residues (Supplementary Fig. 8), further stabilize the structural changes in the P3–P4 connector region. In **b–e**, only the P3 and P4 domains are shown for clarity. **f** Two clusters of residues with complementary electrostatics stabilize a planar P4–P5 binding mode (Supplementary Table 2). **g** A representative asymmetric P3 conformation, highlighting P3–P5 and P3–receptor interactions differing between CheA monomers.

Fig. 10c), by which R404 exerts influence on interactions between neighboring CSUs. Although the focus in the present manuscript has been on the structure and dynamics of a single CSU, our model may be used to construct higher-order assemblies, enabling a detailed investigation of the conformational coupling between CSUs and the origins of high cooperativity in extended arrays.

## Conclusion

The highly conserved bacterial chemotaxis pathway represents one of the best-studied signal transduction systems in all of biology. Consequently, a wealth of genetic, biochemical, and biophysical data has been assembled, especially in the model organism *E. coli*, which provide excellent opportunities for developing and comparing mechanistic signaling models. Owing to the complexity of the chemosensing machinery, however, high-resolution structural data have proven relatively hard to come by,

hindering integration of the aforementioned insights into a comprehensive signal transduction model. In this work, we present an interdisciplinary approach, where we combine in vitro cryoET and STA with molecular modeling and all-atom simulations to provide detailed, residue-level insight into the dynamics of the CSU.

Our results shed new light on several conformational features of CheA in its native structural context. In particular, we show that the CheA.P4 domain is structurally able to adopt a wide range of conformations within the CSU. Consistent with this observation, the P4 region in the 4Q map is less resolved, suggesting that the P4 domain is, indeed, quite mobile. Nevertheless, flexible fitting simulations revealed two, qualitatively different P4 states, the undipped and dipped conformations. Thus, our data support a signaling model in which CheA regulation involves the selection and stabilization of a few specific conformations from a broad landscape of possible conformations. Moreover, the existence of multiple, distinct conformations within our 4Q

(CheA-ON) cryoET data, additionally, suggests that chemoreceptors do not modulate CheA activity by exclusively isolating one or the other of these critical CheA conformations, but rather by biasing the relative occupancies of each state within the population.

We, additionally, provide key atomistic insights into the structural elements responsible for stabilization of the identified states. In particular, we offer a structural explanation for the role of the P3–P4 linker in CheA signaling, showing that this region is involved in different and specific secondary structural motifs at the P3–P4 hinge, which must interconvert in order to stabilize alternate P4 conformations. Moreover, we predict a number of residue pairs that are unique to either conformation, providing specific targets for mutagenesis and cross-linking experiments as well as potential *in vivo* reporters, which should enable elucidation of their roles in the catalytic cycle. Finally, we suggest a role for the P3 domain in CheA signaling, namely that through shifts between symmetric and asymmetric positions within the CSU, the P3 four-helix bundle can alter the interactions and localizations of both P4 domains simultaneously to affect the conformational dynamics of the CheA dimer as a whole. In general, the all-atom molecular models emerging from the present study should provide a valuable structural platform on which to compare, integrate, and expand current models of bacterial signal transduction using experimental and computational techniques alike.

## Methods

**Materials.** Plasmids and cell strains used in this study came from the Parkinson lab, except for plasmid pHTCF, which was a kind gift from Dr. Weis, University of Massachusetts, Amherst. Plasmid pHTCF is an isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG)-inducible expression vector for the N-terminal His<sub>6</sub>-tagged cytoplasmic fragment of the wild-type aspartate receptor (TarCF, residues 357–553), which was also used for producing 4Q and 4E mutant TarCF. Plasmids pKJ9<sup>53</sup> (FL), pQM12 (based on pKJ9 with del P1–P2), and pPA770<sup>54</sup> are IPTG-inducible for the expression of CheA and CheW, respectively.

**Protein expression and purification.** *E. coli* strain RP3098, which lacks all Che proteins and chemoreceptors<sup>55</sup>, was transformed with plasmid pKJ9, pQM12, or pPA770 for CheA or CheW expression, respectively. CheA expression was induced at an OD<sub>600</sub> of 0.6–0.8, with 1 mM IPTG, overnight at 15 °C. CheA was purified using an Affi-gel Blue column (Bio Rad, Hercules, CA) followed by gel filtration on a Superdex 200 column. Further purification with a Mono Q ion exchange column resulted in >99% homogeneity with an overall yield of 50 mg L<sup>-1</sup> of cells. CheW expression was induced by the addition of IPTG (0.5 mM), at an OD<sub>600</sub> of 0.4–0.6, at 37 °C. CheW was purified through 20–40% ammonium sulfate precipitation, a DEAE column followed by a MonoQ ion exchange column and a Superdex 75 size exclusion column. His<sub>6</sub>-tagged wild-type TarCF (His<sub>6</sub>-TarCF<sub>QEQE</sub>), 4Q TarCF (His<sub>6</sub>-TarCF<sub>4Q</sub>), and 4E TarCF (His<sub>6</sub>-TarCF<sub>4E</sub>) were expressed in DH5alpha cells with plasmid pHTCF. TarCF was induced by the addition of IPTG (0.5 mM) at an OD<sub>600</sub> of 0.4–0.6 at 37 °C and purified with a Ni<sup>2+</sup>-NTA affinity column followed with a mono Q column for quick removal of imidazole. The yield for TarCF was excellent (120 mg L<sup>-1</sup> of cells).

**Monolayer reconstitution.** A Ni<sup>2+</sup>-NTA lipid-containing monolayer system was used to reconstitute the chemotaxis core-signaling complex arrays, as described previously<sup>16</sup>. A mixture of 9:18:18  $\mu$ M of TarCF:CheA:CheW in a buffer containing 75 mM Tris-HCl, pH 7.4, 100 mM KCl, and 5 mM MgCl<sub>2</sub> was applied to a Teflon well, over which we immediately laid a lipid monolayer containing 2:1 DOPC:DOGS-NTA-Ni<sup>2+</sup> lipid mixture, at 2 mg mL<sup>-1</sup> concentration. The monolayer set up was left undisturbed in a humidity chamber overnight. The monolayer specimen was picked up with holey carbon grids, stained with 1% uranyl acetate, and examined with an FEI T12 microscope operated at 120KV.

**CryoET.** Reconstituted monolayers were picked up with perforated R2/2 Quantifoil grids (Quantifoil Micro Tools, Jena, Germany) precoated with 10 nm fiducial gold beads and plunge-frozen using a manual gravity plunger. This method prevents disruption of the monolayer by using single-side blotting that eliminates the contact between the blotting filter paper and the delicate monolayer. The frozen-hydrated electron microscopy grids were loaded into FEI Polara cartridges and imaged under low-dose conditions using a Tecnai Polara microscope (FEI Corp., OR.) operating at 300 kv. A series of low-dose projection images were recorded with tilt angles ranging from 70° to –70° with a Gatan 4 K × 4 K CCD camera

(Gatan, Inc., PA), at a nominal magnification of 39,000×, with a defocus value of 3–7  $\mu$ m and an accumulated dose of ~60 e<sup>-</sup> Å<sup>-2</sup>.

**3D reconstruction, sub-tomogram classification, and averaging.** Collected tilt series were aligned and tomograms were reconstructed using patch tracking and weighted back projection, respectively, in IMOD<sup>56</sup>. To extract sub-tomograms, initial positions and orientations of the receptor complexes were estimated via template matching implemented in emClarity<sup>31</sup>. Both the template and tomograms were low-pass filtered to 4 nm and binned to ~1.2 nm per pixel. Following template matching and sub-volume extraction, the data were randomly split into two groups, which were processed independently for all subsequent steps. Sub-tomogram alignment and classification were carried out using emClarity<sup>31</sup>. Sub-tomograms corresponding to CSUs, as judged by the presence of density corresponding to the CheA dimer and CheW, were retained for the final average.

It should be noted that due to the preferred orientation of monolayer arrays, there exists a resolution anisotropy in the resulting density maps as previously observed. This anisotropy leads to better resolution in the X–Y plane and worse resolution in the Z-direction than the overall resolution indicated. To assess the degree of resolution anisotropy, conical FSCs from the two independent half data sets of CSUs were calculated along each of the principal axes as well as the ten axes bisecting them (Supplementary Fig. 2b)<sup>57</sup>. The averaged density map of CSU was then low-pass filtered according to the conical FSCs by using cones with a 42° half-angle, adjusted for any overlapping regions in reciprocal space.

**Component models and homology modeling.** High-resolution structural information for the *E. coli* chemosensory proteins is sparse. In particular, no high-resolution structures exist of the cytoplasmic portion of Tar. As Tar and Tsr possess highly similar sequences (99% similarity and 86% identity in the modeled region), atomic coordinates from *E. coli* Tsr (PDB 1QU7)<sup>41</sup> were taken as an initial model for the receptor TOD (residues 340–441). Receptors were truncated near the glycine hinge to reflect the extent of well-resolved receptor density in the 4Q map. Similarly, no high-resolution structures exist for the P3, P4, or P5 domains of *E. coli* CheA. However, crystal structures of these domains are available from *T. maritima* CheA, which is sufficiently similar to *E. coli* to permit the construction of reliable homology models (75% similarity and 42% identity for CheA.P3.P4.P5). Thus, a crystal structure of the *T. maritima* CheA.P3.P4 dimer (PDB 4XIV<sup>17</sup>) was used to generate atomic coordinates for the corresponding portion of *E. coli* CheA (residues 264–510), while a crystal structure of the *T. maritima* CheA.P5-CheW dimer (PDB 4JPB<sup>42</sup>) was used to simultaneously generate coordinates for *E. coli* CheA.P5 (residues 511–646) and CheW (residues 14–158), thereby, capturing the binding mode at interface 1. All homology models were constructed using Modeller v9.2<sup>58</sup> based on pairwise sequence alignments between the target and template structures. Side chains were subsequently refined using SCWRL4<sup>59</sup> followed by conjugate gradient minimization and all-atom MD.

**MD simulations.** All molecular simulations (i.e., GSA, MDFF, and standard MD) were carried out using NAMD 2.12<sup>60</sup> and the CHARMM36 force field<sup>61</sup>. Equilibrium MD simulations were conducted in the NPT ensemble. Conditions were maintained at 1 atm and 310 K using the Nosé–Hoover Langevin piston (period = 200 fs, relaxation time = 50 fs) and Langevin thermostat (temperature coupling = 5 ps<sup>-1</sup>), respectively. The r-RESPA integrator scheme was employed with an integration time step of 2 fs and SHAKE constraints applied to all hydrogen atoms. Short-range, non-bonded interactions were calculated every 2 fs with a cutoff of 12 Å; long-range electrostatics were evaluated every 6 fs using the particle-mesh-Ewald method.

CSU models were hydrated with TIP3P and neutralized with 150 mM KCl using VMD<sup>62</sup>, producing systems containing ~320,000 atoms. Each model was subjected to a conjugate gradient energy minimization (10,000 steps) and a 20-ns NPT equilibration simulation with backbone restraints. Production simulations of CSU models were conducted with weak harmonic restraints (force constant = 0.25 kcal mol<sup>-1</sup> Å<sup>-2</sup>) placed on the receptor alpha carbons to prevent distortions arising from truncation; the remainder of the complex was not restrained.

**MDFF.** MDFF simulations were performed in the NVT ensemble at 310 K with additional simulation parameters as detailed above. All fittings were carried out in explicit solvent using a scaling factor of 0.1 to couple backbone atoms to the MDFF potential. To prevent the loss of secondary structure as well as the formation of cis-peptide bonds and chirality errors during the fitting process, additional harmonic restraints were applied to the protein backbone using the default force constants provided by the MDFF plugin in VMD. The cascade-MDFF protocol<sup>63</sup> with symmetry restraints<sup>64</sup> was used to fit P4 conformations resulting from clustering analysis of the GSA ensemble; all other fittings utilized the standard, single-density MDFF protocol<sup>65</sup>. Quality-of-fit was assessed in real space by computing the local cross-correlation coefficient (CCC) between the experimental density and a simulated density map derived from the fitted model using the MDFF plugin in VMD<sup>43,66</sup>. To prevent an artificially high assessment of model-map agreement, a threshold was applied to the experimental map to remove relatively low-density values (corresponding to solvent) and only pixels surrounding the localized region of interest in the model (e.g., an individual CheA.P4 domain) were considered in



the calculation<sup>44,66</sup>. UCSF Chimera v1.13<sup>67</sup> was used for map-segmentation and rigid-docking procedures conducted prior to MDFF.

**GSA.** GSA simulations were performed in implicit solvent using previously established parameters for biomolecular applications<sup>47,48,68</sup>. To provide a minimal system that reproduced the structural environment of the P4 domain within the CSU, a CheA.P3.P4.P5 monomer, along with the receptor dimer bound to P5 and the P3 domain from the adjacent monomer, were extracted from the MDFF-refined CSU model (Fig. 3a, Supplementary Fig. 6c). GSA was then used to sample different values of the backbone Phi and Psi dihedral angles of P3–P4 hinge residues 326–328. For this purpose, the GSA approach offers a substantial advantage over standard MD simulation in that much larger conformational transitions may be generated, owing to its Monte Carlo nature, without sacrificing structural fidelity<sup>69</sup>. As the MDFF-refined P4 domain was evidently in a favorable conformation, an initial GSA simulation was conducted to generate several different P4 positions from which to seed an additional round of GSA simulations. The conformations resulting from each of these were combined to produce the analyzed ensemble.

During sampling, the P4 domain and hinge residues (aside from the sampled dihedrals) were held rigid, while the remainder of the structure was held fixed. To prevent unphysical energies arising from steric clashes within the narrow confines of the P3–P4 hinge, which would lead to low acceptance of the conformations proposed by GSA, side chain atoms of residues 326–328 were removed. Additionally, the P4–P5 loop (residues 507–510) was removed to increase sampling efficiency. A minor population of the generated structures had a P4–P5 distance >18 Å (i.e., greater than the length of four residues) and were not included in the analyzed ensemble. Both the side chain atoms and P4–P5 loop residues were subsequently restored using Modeller prior to additional simulation and analysis.

**Data Analysis.** Simulation trajectories were visualized and subjected to basic structural analyses using VMD<sup>62</sup>. Structural clusterings were performed on dissimilarity matrices constructed from pairwise C $\alpha$  RMSD values using the QCP method implemented in MDAnalysis v0.19.0<sup>49,70</sup>. Clusters were formed via an agglomerative hierarchical approach using the so-called centroid linkage criterion (i.e., UPGMC), employing functions implemented in the python module *scipy.cluster.hierarchy*. The distance cutoff parameter, which determines the number of returned clusters, was chosen as 20–40% of the maximum distance based on visual inspection of the clustering dendrogram. In cases where it was desirable to select a single cluster representative, the *k*-medoids algorithm (with *k* = 1) was used to identify the medoid structure of each identified cluster. Potential salt-bridge interactions were identified by calculating the distance between the outermost side-chain carbon atom of acidic and basic residues using MDAnalysis. If this distance was <7.0 Å for more than a third of the total simulation time, hydrogen bonding between the pair was further analyzed in VMD. Images of molecular structures and density maps used in figures were rendered with VMD.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The cryoET density map of the *E. coli* 4Q CSU is deposited in the EM Data Bank under accession code EMD-10050. The raw tilt series will be deposited EMPIAR database. Coordinates for the corresponding atomic model are deposited in the RCSB Protein Data Bank under accession code 6S1K. All relevant data are available from the authors upon reasonable request.

Received: 7 August 2019; Accepted: 19 December 2019;

Published online: 10 January 2020

## References

- Wadhams, G. H. & Armitage, J. P. Making sense of it all: bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **5**, 1024–1037 (2004).
- Eisenbach, M. *Chemotaxis*. (World Scientific Publishing Company, 2004).
- Bi, S. & Sourjik, V. Stimulus sensing and signal processing in bacterial chemotaxis. *Curr. Opin. Microbiol.* **45**, 22–29 (2018).
- Parkinson, J. S., Hazelbauer, G. L. & Falke, J. J. Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. *Trends Microbiol.* **23**, 257–266 (2015).
- Hazelbauer, G. L. Bacterial chemotaxis: the early years of molecular studies. *Annu. Rev. Microbiol.* **66**, 285–303 (2012).
- Hazelbauer, G. L., Falke, J. J. & Parkinson, J. S. Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends Biochem. Sci.* **33**, 9–19 (2008).
- Bornhorst, J. A. & Falke, J. J. Evidence that both ligand binding and covalent adaptation drive a two-state equilibrium in the aspartate receptor signaling complex. *J. Gen. Physiol.* **118**, 693–710 (2001).
- Li, G. & Weis, R. M. Covalent modification regulates ligand binding to receptor complexes in the chemosensory system of *Escherichia coli*. *Cell* **100**, 357–365 (2000).
- Borkovich, K. A., Alex, L. A. & Simon, M. I. Attenuation of sensory receptor signaling by covalent modification. *Proc. Natl Acad. Sci. USA* **89**, 6756–6760 (1992).
- Bilwes, A., Alex, L., Crane, B. & Simon, M. Structure of CheA, a signal-transducing histidine kinase. *Cell* **96**, 131–141 (1999).
- Gloor, S. L. & Falke, J. J. Thermal domain motions of CheA kinase in solution: disulfide trapping reveals the motional constraints leading to trans-autophosphorylation. *Biochemistry* **48**, 3631–3644 (2009).
- Mo, G., Zhou, H., Kawamura, T. & Dahlquist, F. W. Solution structure of a complex of the histidine autokinase CheA with its substrate CheY. *Biochemistry* **51**, 3786–3798 (2012).
- Li, M. & Hazelbauer, G. L. Core unit of chemotaxis signaling complexes. *Proc. Natl Acad. Sci. USA* **108**, 9390–9395 (2011).
- Li, M. & Hazelbauer, G. L. Selective allosteric coupling in core chemotaxis signaling complexes. *Proc. Natl Acad. Sci. USA* **2014**, 1–6 (2014).
- Briegel, A. et al. Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins. *Proc. Natl Acad. Sci. USA* **109**, 3766–3771 (2012).
- Cassidy, C. K. et al. CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling. *Elife* **4**, e08419 (2015).
- Greenswag, A. R., Muok, A., Li, X. & Crane, B. R. Conformational transitions that enable histidine kinase autophosphorylation and receptor array integration. *J. Mol. Biol.* **427**, 3890–3907 (2015).
- Pan, W., Dahlquist, F. W. & Hazelbauer, G. L. Signaling complexes control the chemotaxis kinase by altering its apparent rate constant of autophosphorylation. *Protein Sci.* **26**, 1535–1546 (2017).
- Liu, J. et al. Molecular architecture of chemoreceptor arrays revealed by cryoelectron tomography of *Escherichia coli* minicells. *Proc. Natl Acad. Sci. USA* **109**, E1481–E1488 (2012).
- Sourjik, V. & Armitage, J. P. Spatial organization in bacterial chemotaxis. *EMBO J.* **29**, 2724–2733 (2010).
- Falke, J. J. & Piasta, K. N. Architecture and signal transduction mechanism of the bacterial chemosensory array: progress, controversies, and challenges. *Curr. Opin. Struct. Biol.* **29**, 85–94 (2014).
- Briegel, A. et al. Universal architecture of bacterial chemoreceptor arrays. *Proc. Natl Acad. Sci. USA* **106**, 17181–17186 (2009).
- Briegel, A. et al. Structural conservation of chemotaxis machinery across Archaea and Bacteria. *Environ. Microbiol. Rep.* **7**, 414–419 (2015).
- Matilla, M. A. & Krell, T. The effect of bacterial chemotaxis on host infection and pathogenicity. *FEMS Microbiol. Rev.* **42**, 40–67 (2018).
- Korolik, V. The role of chemotaxis during *Campylobacter jejuni* colonisation and pathogenesis. *Curr. Opin. Microbiol.* **47**, 32–37 (2019).
- Gotoh, Y. et al. Two-component signal transduction as potential drug targets in pathogenic bacteria. *Curr. Opin. Microbiol.* **13**, 232–239 (2010).
- Aihara, E. et al. Motility and chemotaxis mediate the preferential colonization of gastric injury sites by *Helicobacter pylori*. *PLoS Pathog.* **10**, e1004275 (2014).
- Briegel, A. et al. New insights into bacterial chemoreceptor array structure and assembly from electron cryotomography. *Biochemistry* **53**, 1575–1585 (2014).
- Briegel, A. et al. The mobility of two kinase domains in the *Escherichia coli* chemoreceptor array varies with signalling state. *Mol. Microbiol.* **89**, 831–841 (2013).
- Briegel, A. & Jensen, G. Progress and potential of electron cryotomography as illustrated by its application to bacterial chemoreceptor arrays. *Annu. Rev. Biophys.* **46**, 1–21 (2017).
- Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat. Methods* **15**, 955–961 (2018).
- Montefusco, D. J., Shrout, A. L., Besschetnova, T. Y. & Weis, R. M. Formation and activity of template-assembled receptor signaling complexes. *Langmuir* **23**, 3280–3289 (2007).
- Shrout, A. L., Montefusco, D. J. & Weis, R. M. Template-directed assembly of receptor signaling complexes. *Biochemistry* **42**, 13379–13385 (2003).
- Sourjik, V. & Berg, H. Receptor sensitivity in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **99**, 123–127 (2002).
- Frank, V. & Vaknin, A. Prolonged stimuli alter the bacterial chemosensory clusters. *Mol. Microbiol.* **88**, 634–644 (2013).
- Piñas, G. E., Frank, V., Vaknin, A. & Parkinson, J. S. The source of high signal cooperativity in bacterial chemosensory arrays. *Proc. Natl Acad. Sci. USA* **113**, 3335–3340 (2016).

37. Briegel, A., Beeby, M., Thanbichler, M. & Jensen, G. J. Activated chemoreceptor arrays remain intact and hexagonally packed. *Mol. Microbiol.* **82**, 748–757 (2011).
38. Goldman, J. P., Levin, M. D. & Bray, D. Signal amplification in a lattice of coupled protein kinases. *Mol. Biosyst.* **5**, 1853–1859 (2009).
39. Erbe, A. A. H. & Falke, J. J. The core signaling proteins of bacterial chemotaxis assemble to form an ultrastable complex. *Biochemistry* **48**, 6975–6987 (2009).
40. Cassidy, C. K., Himes, B. A., Luthey-Schulten, Z. & Zhang, P. CryoEM-based hybrid modeling approaches for structure determination. *Curr. Opin. Microbiol.* **43**, 14–23 (2018).
41. Kim, K., Yokota, H. & Kim, S. Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature* **400**, 787–792 (1999).
42. Li, X. et al. The 3.2 Å resolution structure of a receptor:CheA:CheW signaling complex defines overlapping binding sites and key residue interactions within bacterial chemosensory arrays. *Biochemistry* **52**, 3866–3880 (2013).
43. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
44. McGreevy, R., Teo, I., Singharoy, A. & Schulten, K. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **100**, 50–60 (2016).
45. Miller, A., Kohout, S., Gilman, K. & Falke, J. CheA kinase of bacterial chemotaxis: chemical mapping of four essential docking sites. *Biochemistry* **45**, 8699–8711 (2006).
46. Wang, X., Wu, C., Vu, A., Shea, J. E. & Dahlquist, F. W. Computational and experimental analyses reveal the essential roles of interdomain linkers in the biological function of chemotaxis histidine kinase CheA. *J. Am. Chem. Soc.* **134**, 16107–16110 (2012).
47. Melo, M. C. R., Bernardi, R. C., Fernandes, T. V. A. & Pascutti, P. G. GSAFold: a new application of GSA to protein structure prediction. *Proteins Struct. Funct. Bioinforma.* **80**, 2305–2310 (2012).
48. Tsallis, C. & Stariolo, D. A. Generalized simulated annealing. *Phys. A Stat. Mech. its Appl.* **233**, 395–406 (1996).
49. Stone, J. E., Perilla, J. R., Cassidy, C. K. & Schulten, K. GPU-accelerated molecular dynamics clustering analysis with OpenACC. *Parallel Program. OpenACC* 215–240 (2016).
50. Wang, X. et al. The linker between the dimerization and catalytic domains of the CheA histidine kinase propagates changes in structure and dynamics that are important for enzymatic activity. *Biochemistry* **53**, 855–861 (2014).
51. Lai, R.-Z. Z., Gosink, K. K. & Parkinson, J. S. Signaling consequences of structural lesions that alter the stability of chemoreceptor trimers of dimers. *J. Mol. Biol.* **429**, 823–835 (2017).
52. Piñas, G. E., DeSantis, M. D. & Parkinson, J. S. Noncritical signaling role of a kinase–receptor interaction surface in the Escherichia coli chemosensory core complex. *J. Mol. Biol.* **430**, 1051–1064 (2018).
53. Garzón, A. & Parkinson, J. S. Chemotactic signaling by the P1 phosphorylation domain liberated from the CheA histidine kinase of Escherichia coli. *J. Bacteriol.* **178**, 6752–6758 (1996).
54. Zhao, J. & Parkinson, J. S. Mutational analysis of the chemoreceptor-coupling domain of the Escherichia coli chemotaxis signaling kinase CheA. *J. Bacteriol.* **188**, 3299–3307 (2006).
55. Smith, R. A. & Parkinson, J. S. Overlapping genes at the cheA locus of Escherichia coli. *Proc. Natl Acad. Sci. USA* **77**, 5370–5374 (1980).
56. Kremer, J. R., Mastrorarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
57. Diebold, C. A., Faas, F. G. A., Koster, A. J. & Koning, R. I. Conical fourier shell correlation applied to electron tomograms. *J. Struct. Biol.* **190**, 215–223 (2015).
58. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **2016**, 5.6.1–5.6.37 (2016).
59. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins Struct. Funct. Bioinformatics* **77**, 778–795 (2009).
60. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Computational Chem.* **26**, 1781–1802 (2005).
61. Huang, J. & Mackerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
62. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
63. Singharoy, A. et al. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *Elife* **5**, 1–33 (2016).
64. Chan, K. Y. et al. Symmetry-restrained flexible fitting for symmetric em maps. *Structure* **19**, 1211–1218 (2011).
65. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).
66. Stone, J. E., McGreevy, R., Isralewitz, B. & Schulten, K. GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting. *Faraday Discuss.* **169**, 265–283 (2014).
67. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
68. Agostini, F. P., Soares-Pinto, D. D. O., Moret, M. A., Osthoff, C. & Pascutti, P. G. Generalized simulated annealing applied to protein folding studies. *J. Comput. Chem.* **27**, 1142–1155 (2006).
69. Bernardi, R. C., Melo, M. C. R. & Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* **1850**, 872–877 (2015).
70. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

## Acknowledgements

We thank Dr. Robin Corey for critical reading of the manuscript and many valuable suggestions, Professor Mark Sansom for fruitful discussions, and D. Bevan for computer technical support. This work was supported by the National Institutes of Health NIGMS Grants R01GM085043 (P.Z.), RO5GM19559 (J.S.P.) and P41GM104601 (Z.L.-S.), the US National Science Foundation grant PHY1430124 (Z.L.-S.), the UK Wellcome Trust Investigator Award 206422/Z/17/Z (P.Z.), the UK Biotechnology and Biological Sciences Research Council grants BB/S003339/1 (C.K.C., P.J.S., and P.Z.), BB/P01948X/1 and BB/R002517/1 (P.J.S.), the Medical Research Council grant MR/S009213/1 (P.J.S.), and the UK Wellcome Trust Technology Award 208361/Z/17/Z (P.J.S.). MD simulations were performed on the Blue Waters supercomputer, which is supported by the National Science Foundation (OCI-0725070 and ACI-1238993) and the state of Illinois. This work is part of the Petascale Computational Resource Grant (ACI-1713784).

## Author contributions

P.Z. conceived the experiments and with C.K.C. designed the experiments. J.M. performed in vitro reconstitution of monolayer arrays of CSUs containing WT, 4Q, and 4E receptor variants. P.Z. and G.Z. collected cryoET data. B.H. and D.S. performed cryoET reconstruction and STA; C.K.C., with P.Z. designed computational approach, constructed structural models, and performed molecular simulations in discussion with Z.L.-S. and P.J.S.; C.K.C. analyzed atomic models and interpreted simulation results in discussion with J.S.P. and P.Z.; C.K.C. and P.Z. wrote the paper with support from other authors.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-019-0748-0>.

Correspondence and requests for materials should be addressed to C.K.C. or P.Z.

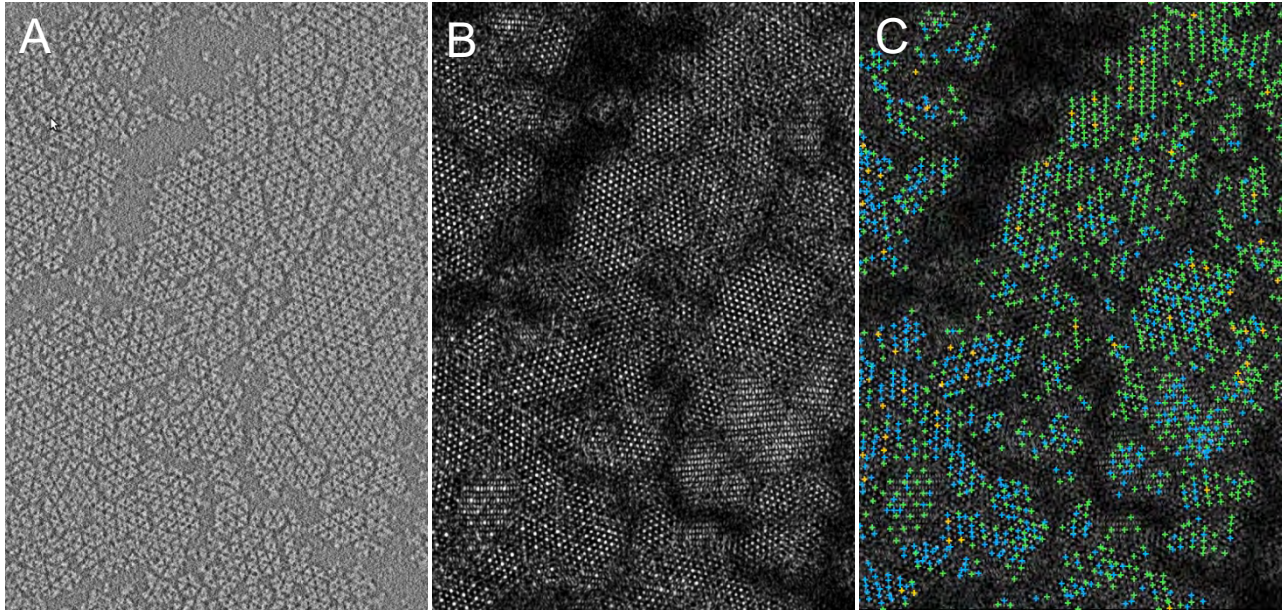
Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

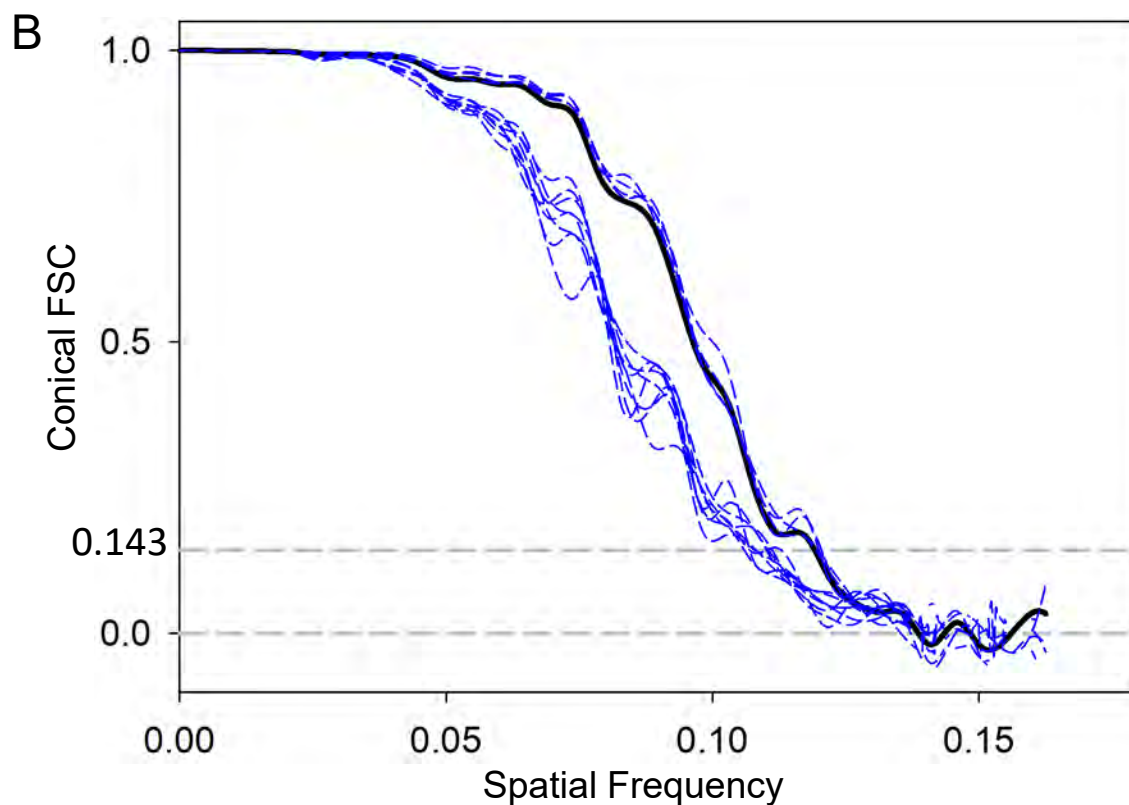
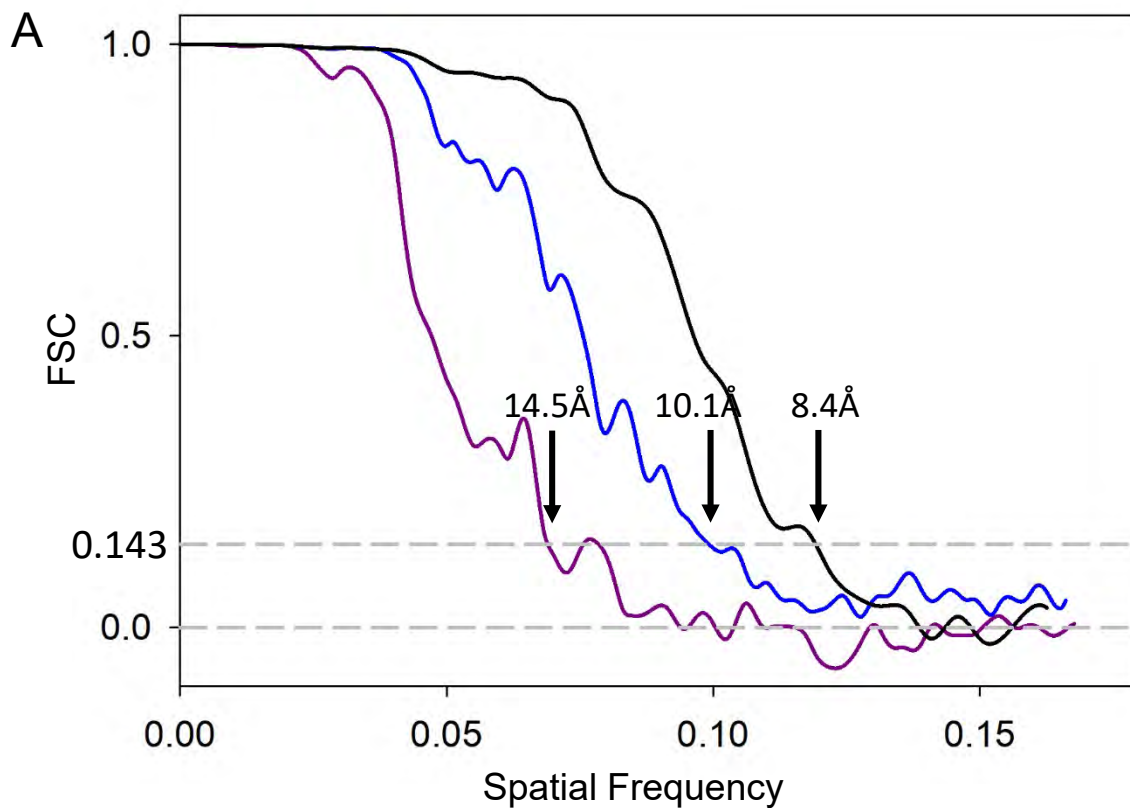


**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

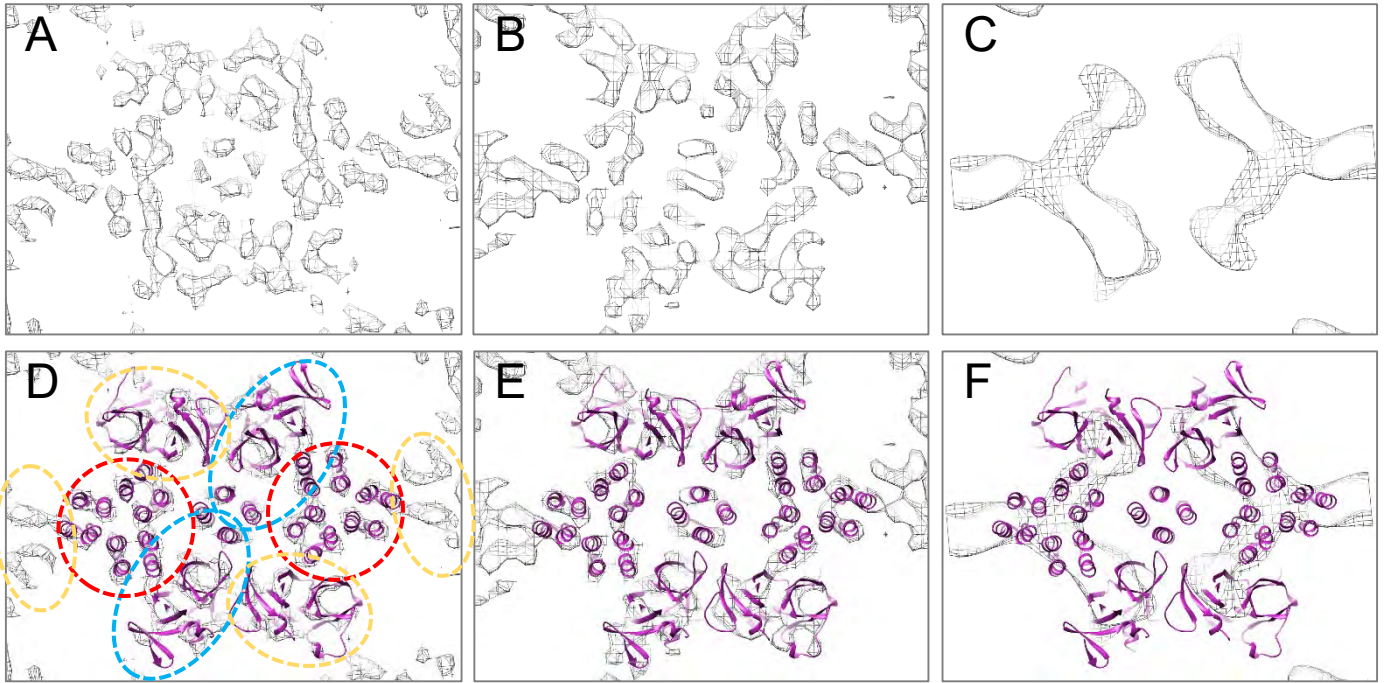
© The Author(s) 2020



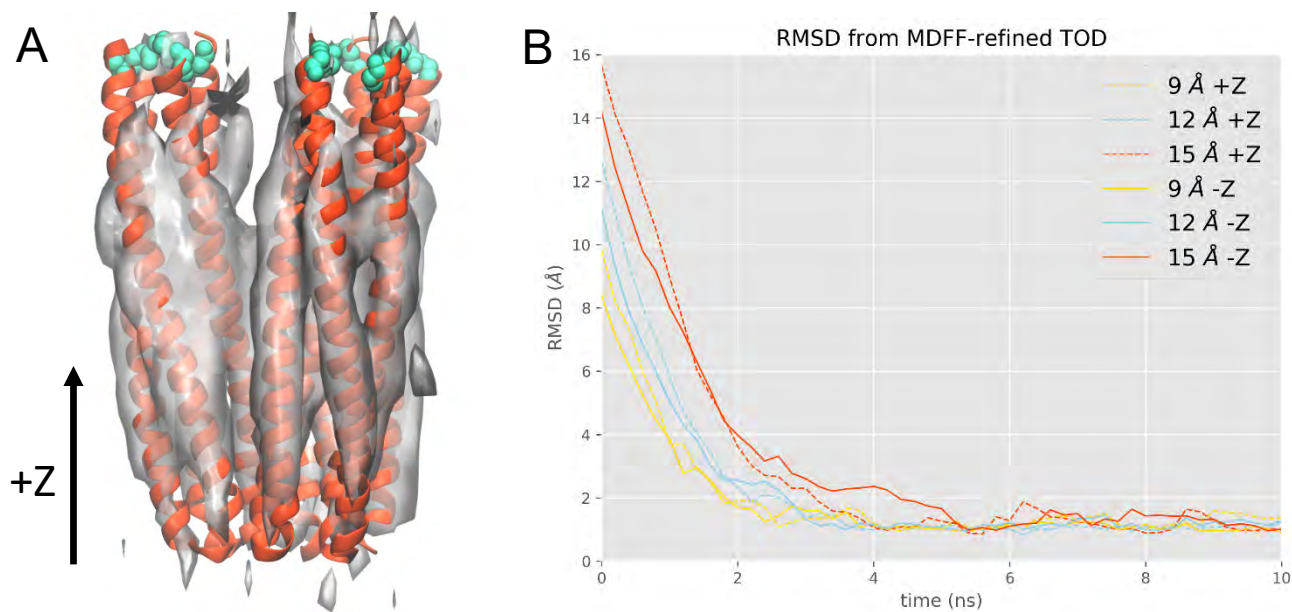
**Supplementary Figure 1.** Extraction of sub-tomograms by template matching. (A) A slice from a 3D tomogram of the TarCF<sub>QE QE</sub>/CheA/CheW monolayer arrays. (B&C) The corresponding slice of a 3D convolution map obtained by template matching (B) and overlaid with the positions of matched sub-tomograms (C). Colors orange, blue, and green indicate strong-to-weak peak intensities in the convolution map.



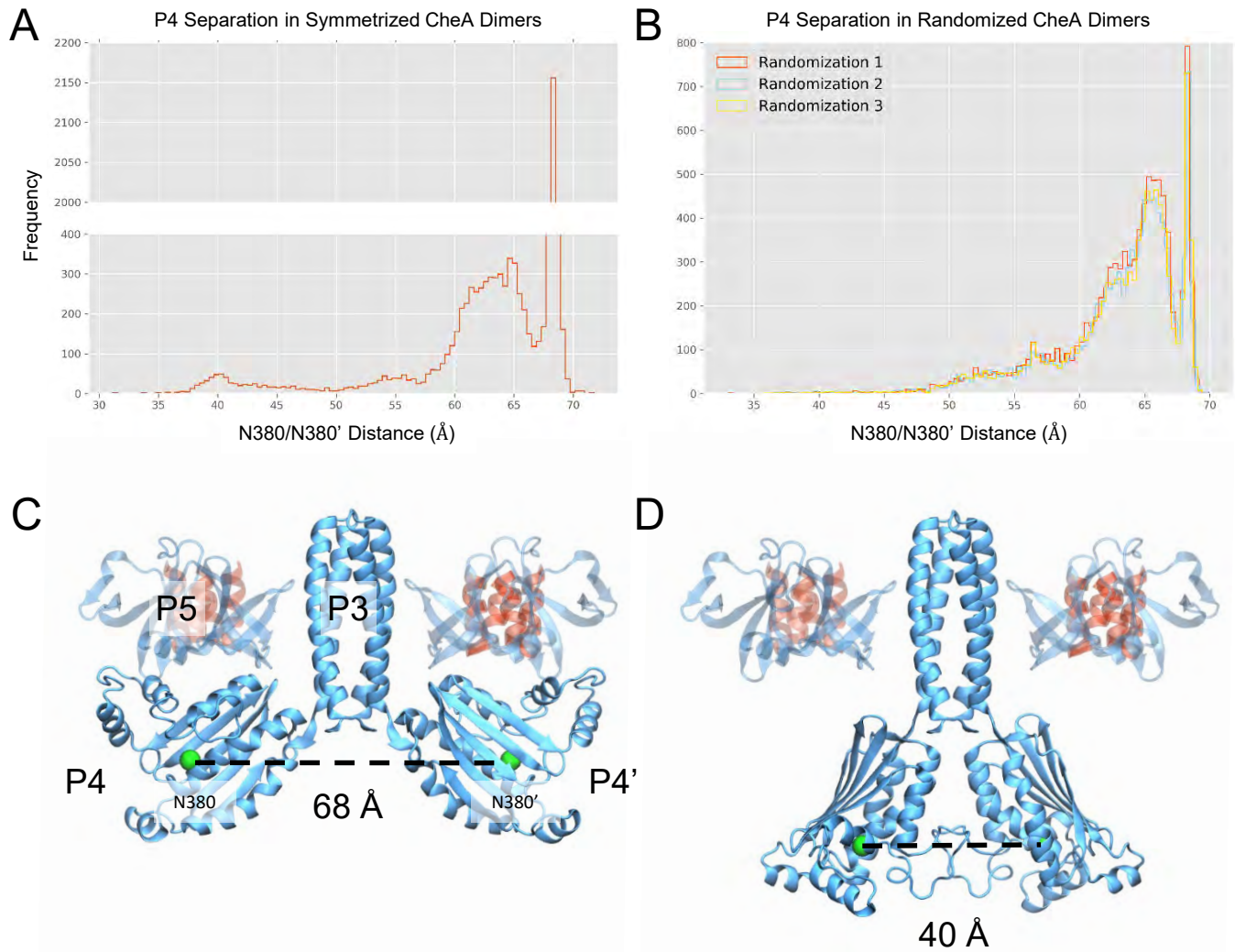
**Supplementary Figure 2.** Fourier shell correlation (FSC) between the two independently aligned and averaged half-datasets of core signaling unit. (A) FSC plots of sub-tomogram averages of TarCF<sub>4E</sub>/CheA/CheW (Purple), TarCF<sub>QE</sub>/CheA/CheW (Blue) and TarCF<sub>4Q</sub>/CheA/CheW (Black). Resolution values are reported at the 0.143 threshold. (B) Conical FSC plots (16 cones, dashed blue) overlaid with the overall FSC (black) of core signaling unit TarCF<sub>4Q</sub>/CheA/CheW.



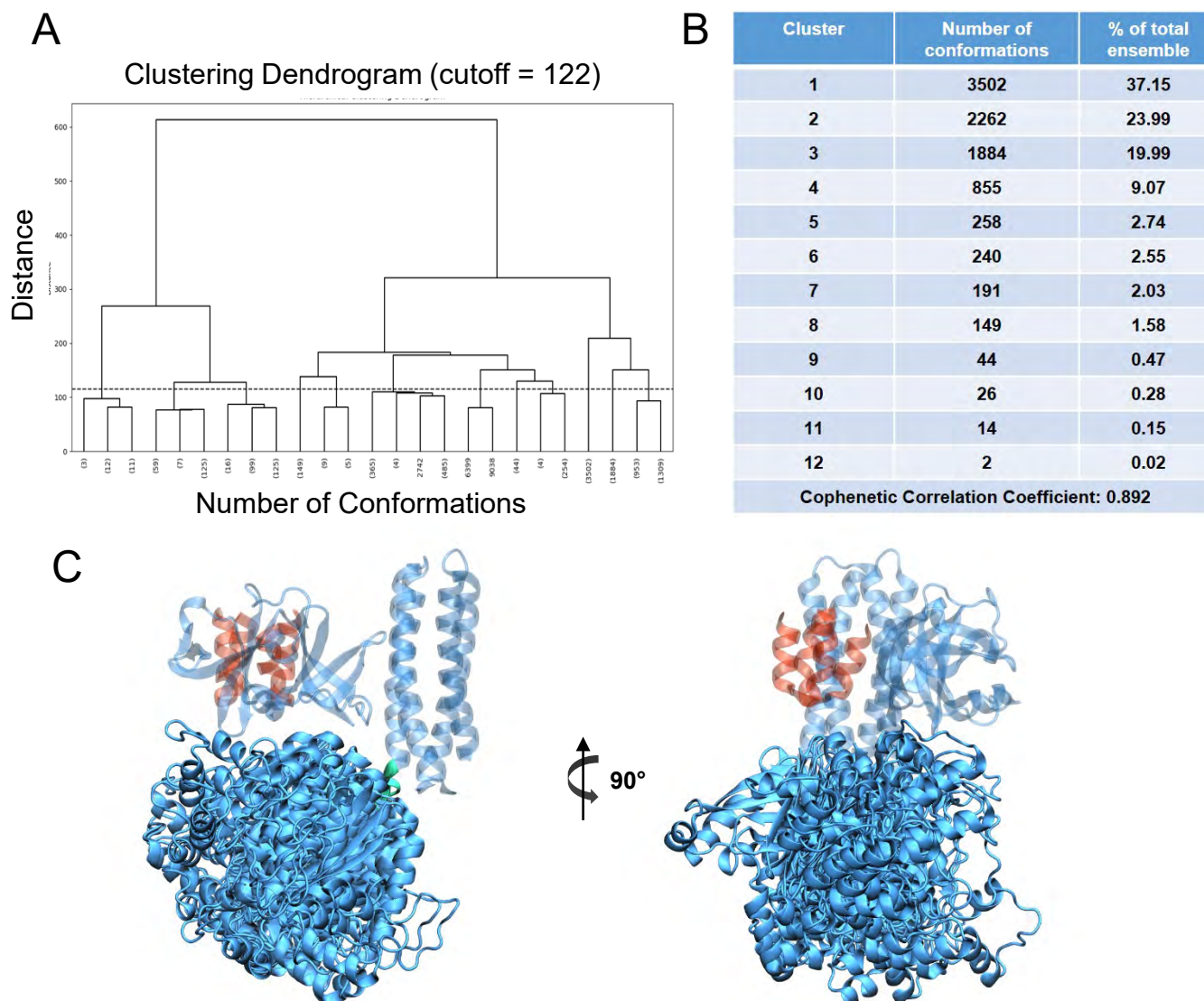
**Supplementary Figure 3.** CheA.P5 and CheW interact with receptors in a ratio of 1:2. (A-C) Clipped sectional views from Fig. 1 in surface mesh. (D-F) Same sectional views overlaid with atomic models of the core signaling unit. Receptor TODs, CheA.P3.P5, and CheW are marked in (D) with red, blue and yellow dashed circles, respectively.



**Supplementary Figure 4.** MDFF robustly positions receptor trimers-of-dimers (TOD) within 4Q map. (A) Overlay between MDFF-refined Tsr TOD model and corresponding density extracted from 4Q map. Receptors are shown in red; residues comprising the glycine hinge are shown in teal. (B) Timeseries of the root-mean-square deviation (RMSD) between the final frame of the initial TOD MDFF simulation and MDFF simulations of TOD models whose starting positions were shifted in the Z-direction by +/- 9 Å (yellow lines), +/- 12 Å (blue lines), or +/- 15 Å (red lines). See also Supplementary Movie 2.

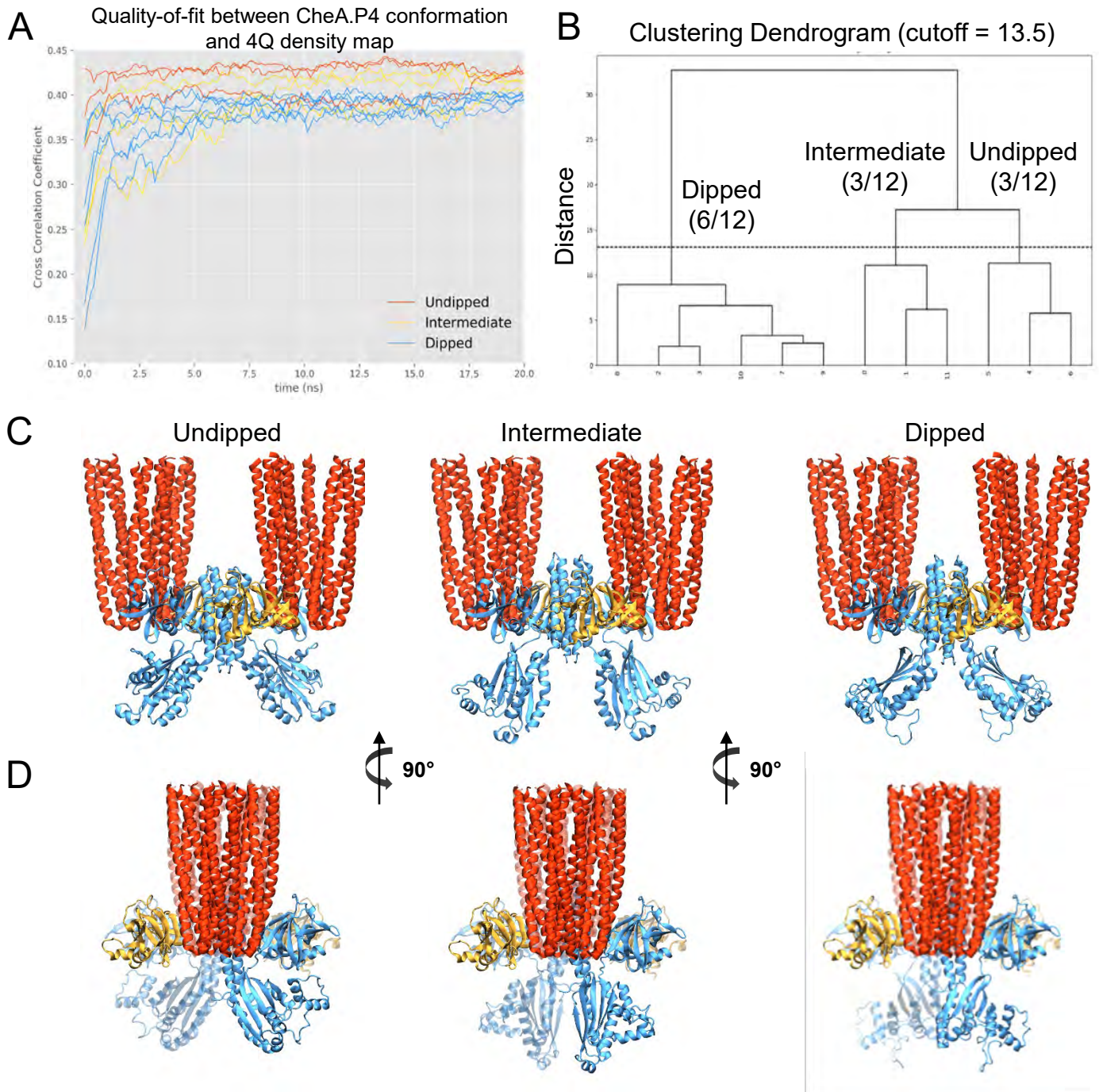


**Supplementary Figure 5.** CheA.P4 can adopt a wide-range of conformations within the core signaling unit structure. (A-B) Separation between P4 active sites, as measure by the C $\beta$ -C $\beta$  distance of residues N380 and N380', in CheA dimers constructed by combining GSA-derived P4 monomer conformations symmetrically (A) and randomly (B). In (B), each trace corresponds to a different random combination of CheA monomers. (C-D) Representative conformations from symmetrized CheA dimer distribution, showing the P4 domains well-separated (C) and directly interacting (D). CheA.P3.P4 is shown in opaque blue; the C $\beta$ 's of residues N380 and N380' are shown as green spheres. For context, CheA.P5 and its bound receptor dimer are shown in transparent blue and red, respectively.

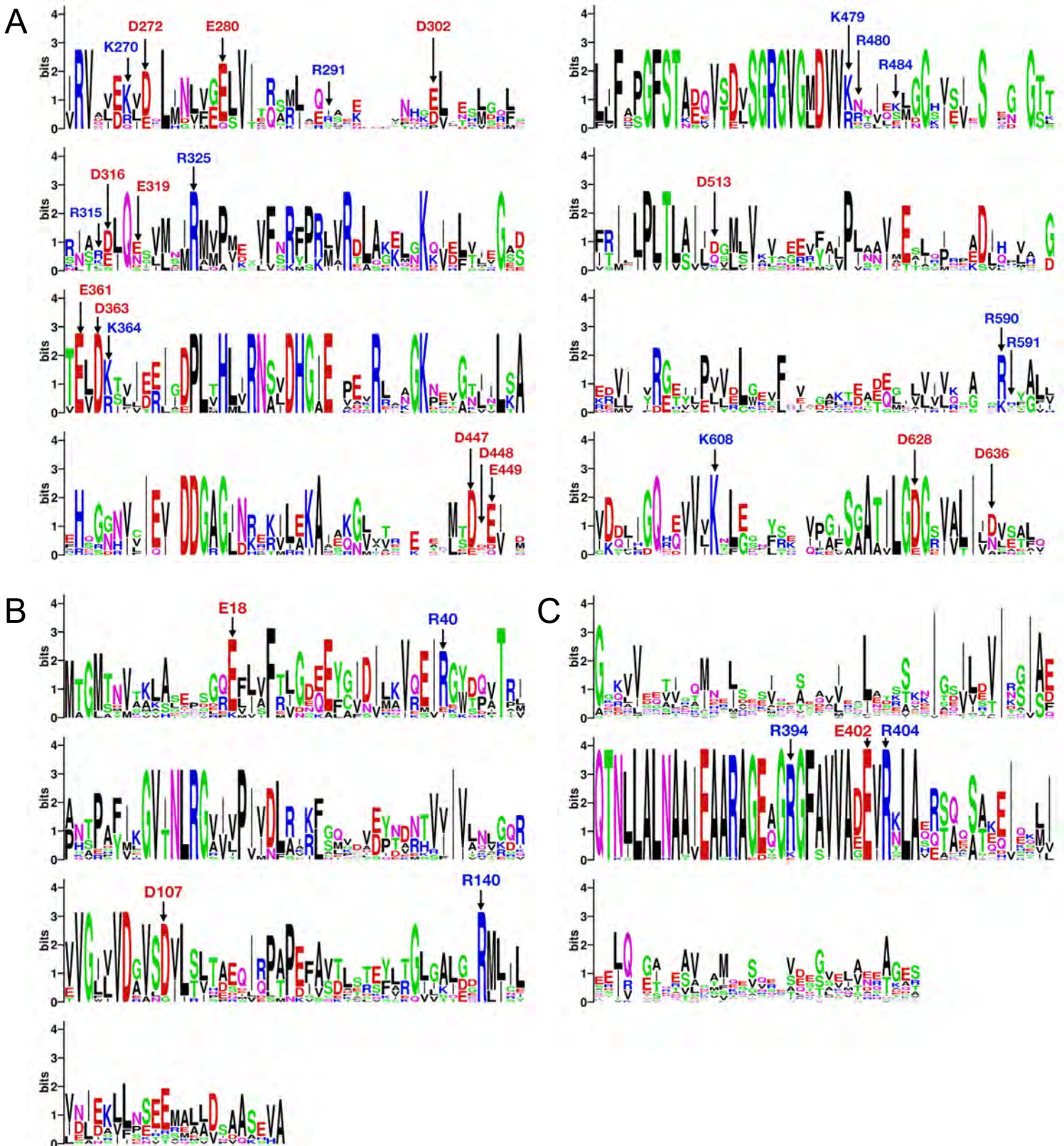


**Supplementary Figure 6.** Structural clustering of Generalized Simulated Annealing ensemble. (A) Dendrogram resulting from hierarchical clustering of GSA ensemble, truncated at 25 leaves for clarity. A horizontal dashed line denotes the distance cutoff, chosen to be 20% of the maximum distance, resulting in 12 clusters. (B) Number of conformations and percentage of total ensemble represented by each cluster. A cophenetic correlation coefficient  $> 0.75$  suggests the clustering preserves well the original RMSD relationships between conformations. (C) Overlay of 12 representative P4 conformations resulting from clustering analysis. See Methods for more details.

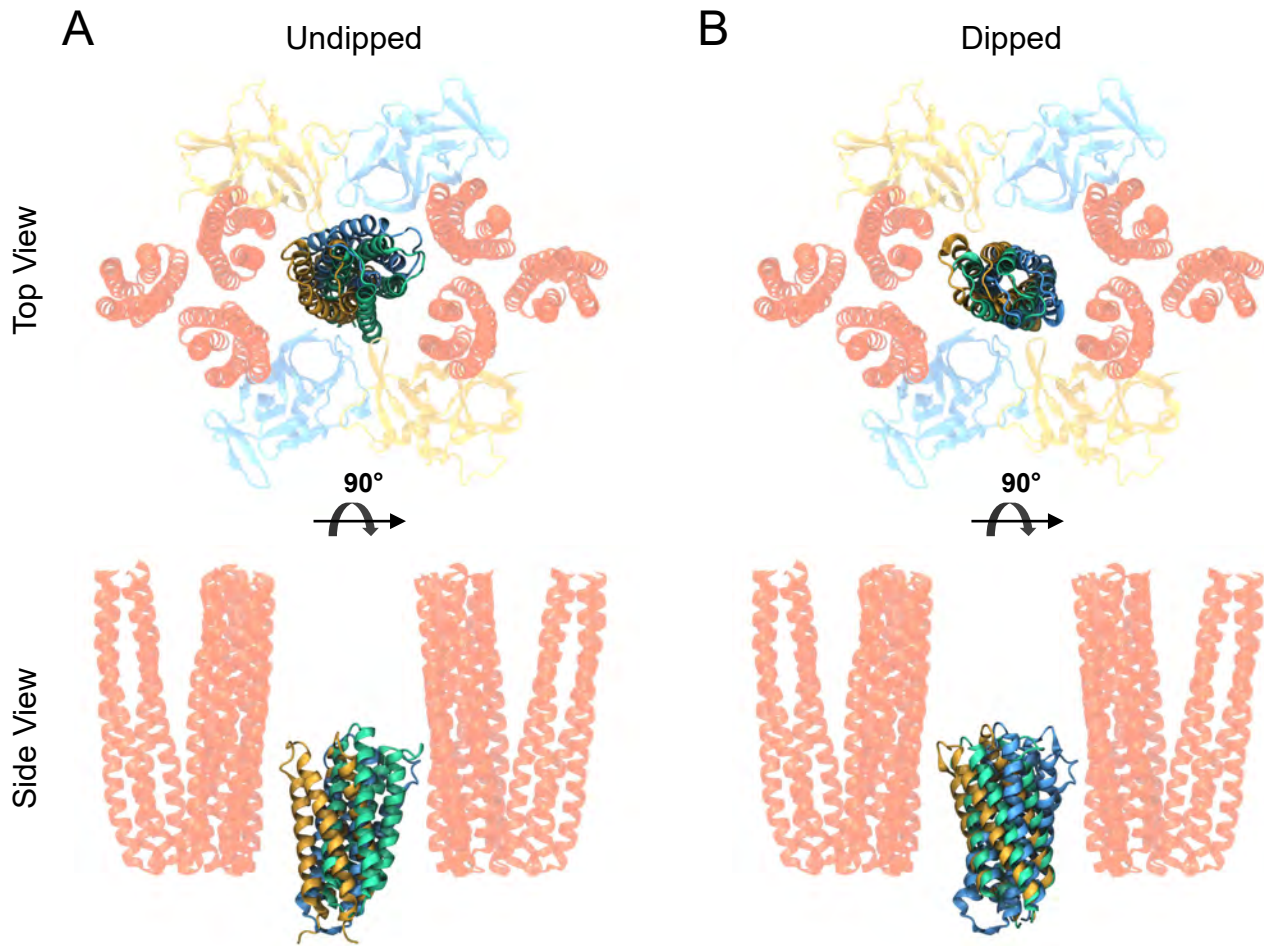




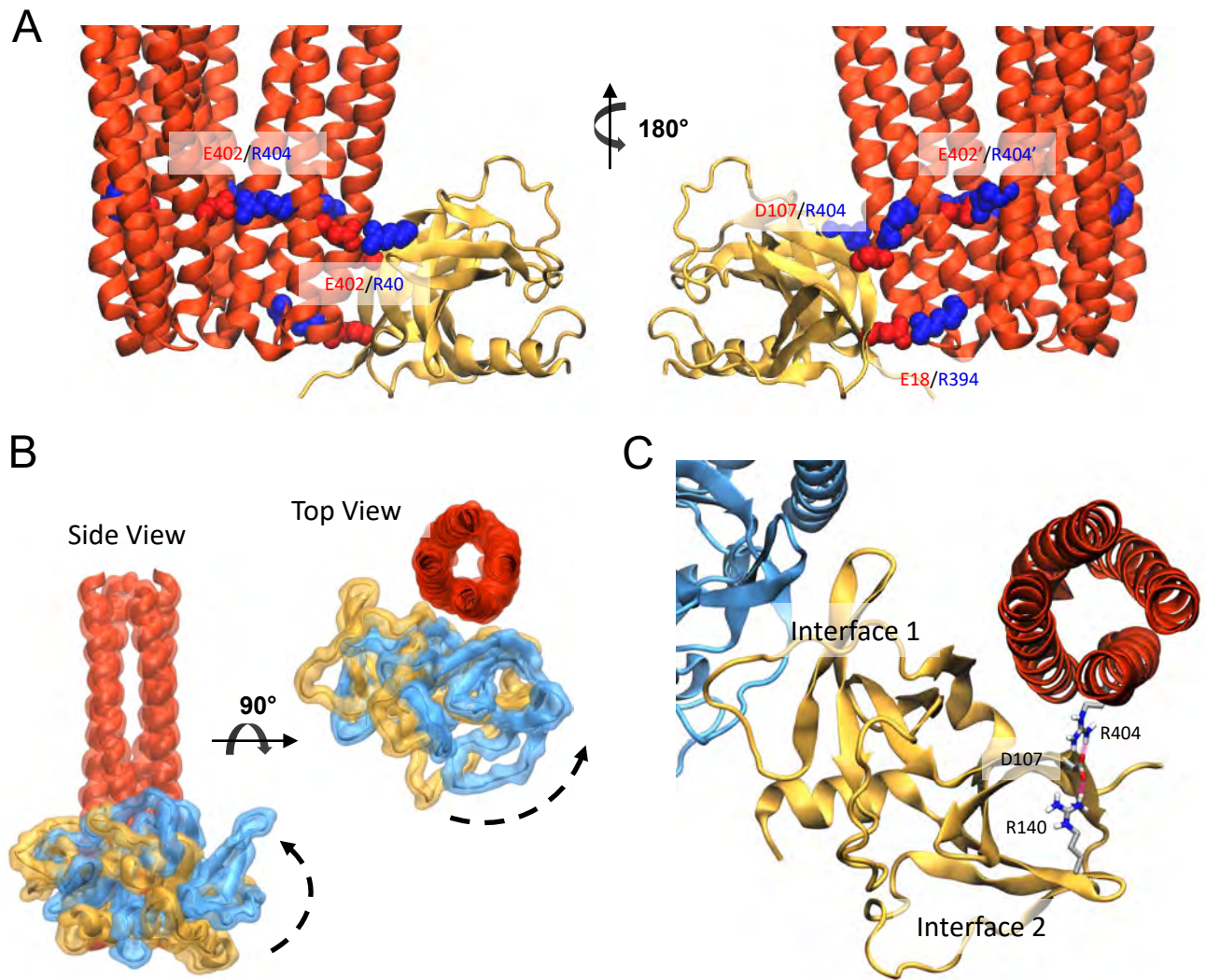
**Supplementary Figure 7.** 4Q density corresponds to multiple, distinct CheA.P4 conformations. (A) Timeseries of the local cross-correlation coefficient (see Methods) showing the quality-of-fit between each of the 12 GSA-derived CheA.P4 conformations and the 4Q density map plotted over the course of a 20-ns MDFF refinement simulation. Individual traces are colored according to the CheA conformational class to which their final conformation was assigned via hierarchical clustering. (B) Dendrogram resulting from hierarchical clustering of the 12 MDFF-refined CheA conformations. A horizontal dashed line denotes the distance cutoff, chosen to be 40% of the maximum distance, resulting in 3 clusters: “undipped” (three conformations), “intermediate” (three conformations), and “dipped” (six conformations). (C-D) Representative core signaling unit model from each CheA conformational class shown from the front (C) and side (D). The “undipped” and “intermediate” classes are qualitatively similar, but the latter displays an increased P4-P5 separation due to a  $\sim 30^\circ$  increase in the P3-P4 hinge angle.



**Supplementary Figure 8.** Sequence conservation in modeled region of the *E. coli* core signaling unit. (A-C) WebLogos derived from multi-sequence alignments of CheA.P3.P4.P5 (A), CheW (B), and Tsr (C) using <https://weblogo.berkeley.edu>. Residues discussed in the main text are labeled explicitly. Sequences from a broad range of organisms were obtained for each protein from ‘reviewed’ datasets on UniProtKB and aligned using Clustal-Omega. Sequences included in the alignments are as follows: CheA [9 sequences/organisms: *ECOLI*, *BACSU*, *SALTY*, *THEMA*, *KLEAK*, *RHOSH*, *LISIN*, *LISMO*, *HALS3*], CheW [10 sequences/organisms: *ECOLI*, *BACSU*, *SALTY*, *THEMA*, *KLEAK*, *RHOSH*, *SHIFL*, *CAUVC*, *RHIME*, *RHIEC*], and Tsr [30 sequences/8 organisms: *ECOLI*, *BACSU*, *SALTY*, *THEMA*, *VIBCH*, *PSEPK*, *PSEAE*, *PSEPU*].



**Supplementary Figure 9.** CheA.P3 can adopt asymmetric positions within the core signaling unit. (A-B) Final frames taken from three "undipped" (A) and three "dipped" (B) core signaling unit simulations, showing the position of the CheA.P3 domain from the top and side. Individual P3 four-helix bundles are shown in blue, teal, and gold. For clarity, CheA.P4 is not shown and CheA.P5-CheW are only in the top view.



**Supplementary Figure 10.** Characterization of CheW-receptor and CheA.P5-receptor interfaces. (A) Key salt bridge interactions, involving Tsr-E402 and Tsr-R404, at CheW-receptor and receptor-receptor interfaces. (B) Superposition of the pseudosymmetric CheA.P5-receptor and CheW-receptor interfaces from the side and top. Dashed arrows denote the relative rotation between CheW (gold) and CheA.P5 (blue). (D) Possible allosteric pathway between Tsr-R404 and interface 2, involving salt bridge interactions with CheW residues D107 and R140.

**Supplementary Table 1.** Summary of cryo-electron tomography data acquisition and image processing.

Acquisition setup			
	Microscope	FEI Polara	
	Voltage (keV)	300	
	Detector	Gatan 4Kx4K CCD	
	Energy-filter	No	
	Å/pixel	2.98	
	Acquisition scheme	Bi-directional	
Processing	4Q	QEQE	4E
# of tomograms	24	22	19
Defocus range	-5.0 - 7.0	-4.0 - 6.0	-3.0 - 5.0
Total dose (e <sup>-</sup> /Å <sup>2</sup> )	60	60	60
Tilt range	-70° to 70°	-70° to 70°	-70° to 70°
# of CSUs	91,636	56,726	33,194
Final resolution in Å	8.38	10.07	14.47

**Supplementary Table 2.** Summary of protein-protein interfaces observed in MD simulations of the “undipped” and “dipped” CSUs.

Interface	Residues†	Undipped	Dipped	Occupancy
P3-P4	R325-D272	x	x	High in undipped Intermediate in dipped
	R325-D363	x		High
	R325-E361		x	High
	K270-E361	x		Intermediate
	K364-D272	x	x	Low in undipped High in dipped
P4-P5	Cluster 1* (K479,R480,K484)- D636	x		High
	Cluster 2* (D447,D448,E449)- (R590,R591)	x		High
P3-P5	E280-K608	x	x	-
	R315-D513	x	x	-
	R291-D628	x	x	-
P3-R	(D316,E319)-R394	x	x	-
	D302-R409	x	x	-
W-R	R40-E402	x	x	High
	D107-R404	x	x	High
	E18-R394	x	x	Intermediate

†Residue numbers are listed in the order denoted in the “Interface” column. Interactions involving multiple potential partners within the same class are denoted with parentheses. Occupancies computed for the two contact “clusters” at the P4-P5 interface, denoted with a “\*”, consider that at least one of the possible contacts within the cluster is present in a given frame. Due to their asymmetric nature, occupancies for the P3-P5 and P3-R interface contacts are not given.