

# Effects of aligned $\alpha$ -helix peptide dipoles on experimental electrostatic potentials

Jimin Wang <sup>1\*</sup>, Pablo E. Videla,<sup>2</sup> and Victor S. Batista<sup>2</sup>

<sup>1</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114

<sup>2</sup>Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107

Received 24 March 2017; Accepted 24 May 2017

DOI: 10.1002/pro.3204

Published online 27 May 2017 proteinscience.org

**Abstract:** Aligned protein  $\alpha$ -helix dipoles have been implicated in protein function and structure. The recent breakthroughs in high-resolution electron microscopy (EM) of macromolecules makes it possible to explore fundamental aspects of structural biology at the detailed molecular level. The electrostatic potential (ESP) generated by aligned protein  $\alpha$ -helix dipole should be observable in high-resolution EM maps despite the fact that the effect may be partially screened by induced electric fields. Here, we show that aligned backbone dipoles in protein  $\alpha$ -helices account for long-range features in the protein ESP functions. Our results are consistent with experimental EM maps and density functional theory calculations, including direct Fourier summation for proper calculation of the ESP due to the nonlocal nature of the ESP function from aligned dipoles and other partial atomic charges.

**Keywords:** electrostatic potential (ESP); electron scattering; electron microscopy (EM); electron diffraction; partial atomic charge; density functional theory (DFT)

## Introduction

The electric field of a protein  $\alpha$ -helix can be approximated by a continuous line dipole with a dipole density of 3.5 D per 1.5 Å, due to the dipoles of aligned backbone peptide units.<sup>1</sup> The resulting field is equivalent to the electric field generated by a positive charge at the N-terminus of the helix and a negative charge at the C-terminus, each of magnitude of about half an elementary charge unit. Aligned protein  $\alpha$ -helix dipoles have been shown to be important for protein function and structure. They can modulate  $pK_a$  of amino acid residues involved in proton/electron transfer, binding affinities of various cofactors, substrates

and ions, and stability of folded proteins.<sup>2–6</sup> The electric field of aligned  $\alpha$ -helix dipoles in a folded soluble protein is often screened by the counteractive field induced in the surrounding environment.<sup>2,7</sup> These counteractive electrostatic interactions include: (i) capping residues such as the preferential presence of negatively charged residues at the N-terminus of the helix, and positively charged residues at the C-terminus, (ii) other peptide dipoles within many basic folding motifs such as antiparallel  $\alpha$ -helix motif and in the basic  $\beta\alpha\beta$  motif, (iii) external counter-ions at the termini of the helix, and (iv) highly polarized solvent molecules, to name a few. Nevertheless, aligned dipoles of protein backbones along transmembrane  $\alpha$ -helices could play important structural/functional roles since the hydrophobic membrane environment has only limited capability of screening the overall electrostatic potential (ESP).<sup>7</sup>

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institutes of Health; Grant numbers: Grant P01 GM022778 (J. W.) and NIH GM106121 (V. S. B).

\*Correspondence to: J. Wang. E-mail: jimmin.wang@yale.edu

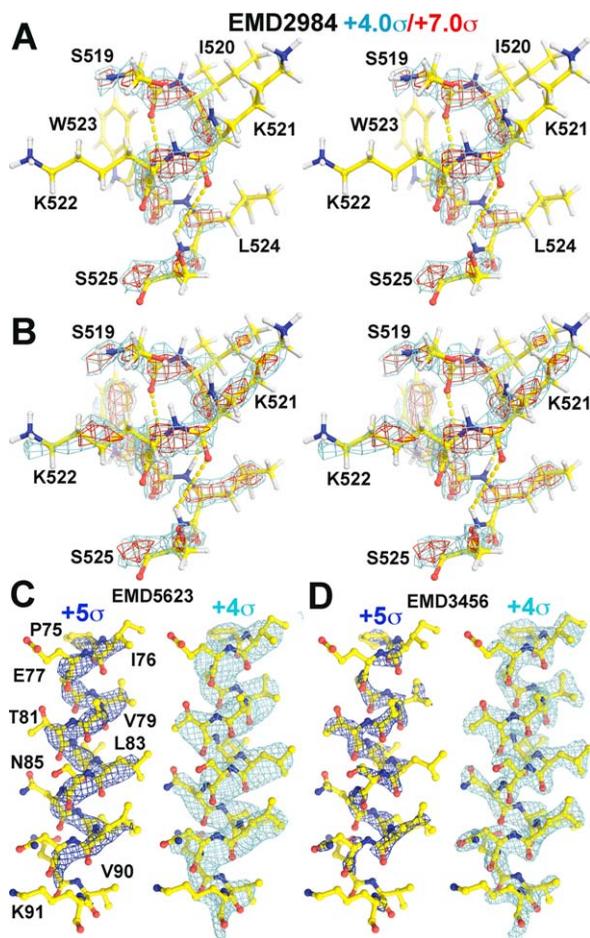
Recent advances in cryo-electron microscopy (EM) have enabled experimental determination of high-resolution ESP maps for a variety of large macromolecules.<sup>8</sup> An outstanding question is whether the EM maps could provide insights on the contributions from aligned dipoles of protein  $\alpha$ -helices. The ESP maps involve a neutral ‘atomic’ component that distributes within the van der Waals (vdW) radii of the constituent atoms, and a long-range ‘Coulomb’ component beyond the vdW radii. The Coulomb component makes experimental ESP features for atoms carrying negative (or positive) partial charges less (or more) prominent than those of neutral atoms.<sup>9–11</sup> Partial atomic charges can be further divided into residue-specific partial charges,<sup>12,13</sup> which are well documented, and environment-specific partial charges,<sup>14,15</sup> which are still under active investigation. In this paper, we study the ESP map of a short  $\alpha$ -helix found in  $\beta$ -galactosidase (BGal), computed at the density functional theory (DFT) level and compared to the corresponding experimental ESP map obtained by EM.<sup>16</sup> We focus here on the Coulomb component for the system in an aqueous phase whereas a similar early study was carried out for much smaller systems in the solid state.<sup>17</sup>

## Results and Discussion

### Gradation in the ESP map of $\alpha$ -helices

The experimental ESP map<sup>16</sup> for BGal, reported at 2.2-Å resolution, shows a large-to-small gradation along the N-to-C direction of the protein backbone for the  $\alpha$ -helix segment comprising residues S519–S525 contoured at  $+4\sigma$  to  $+7\sigma$  levels (Fig. 1). Although there are two Lys residues in this helix, neither is at the C-terminus but rather they are in the middle of the chain so that they do not offer effective sites for capping interactions. Furthermore, no counter-ions seem to bind as capping ions to the helix termini. Thus, the observed gradation in the ESP map must be due to the aligned dipoles in the  $\alpha$ -helix. The length of the helix is close to the predicted asymptotic limit for accumulative aligned helix dipole,  $\sim 10$  Å, beyond which there should be no further apparent accumulation.<sup>1</sup> We note that this helix is relatively short so that any change in atomic mobility (measured by motion B-factors) should have a minor effect on the experimental ESP map.

Two experimental ESP maps have been reported for *T. acidophilum* proteasome<sup>18,19</sup> at 3.3-Å (EMD-5623) and 2.4-Å (EMD-3456) resolution. The maps show a similar gradation as observed in BGal, although the proteasome  $\alpha$ -helix is longer and comprises residues I76 to K91 in the  $\beta$ -subunit (Fig. 1). This helix is partially buried and has capping residues, including E77 at the N-terminus and K91 at the C-terminus that could cancel part of the electric field generated by the aligned helix dipoles. It is oriented with its C-terminus



**Figure 1.** Experimental ESP maps showing gradation of ESP functions along the  $\alpha$ -helix segment S519–S525 of BGal [EMD-2984 with Ref. (16)] (stereodiagrams in panels A and B) contoured at  $+4.0\sigma$  (cyan) and  $+7.0\sigma$  (red) and an  $\alpha$ -helix from proteasomal  $\beta$ -subunit (panels C and D) [EMD-5623 with Ref. (18) EMD-3456 with Ref. (19)] comprising residues I76 to K91 (residue P75 is also included, but not labeled) contoured at  $+4.0\sigma$  (cyan) and  $+5.0\sigma$  (blue). The presence of capping E77 residue lowers the ESP values in its immediate vicinity. For clarity, the ESP features for side chains were deleted in panel A.

towards the center of the proteasome so that any angular errors during image reconstruction would make the motion B-factors for the N-terminus residues (i.e., residues away from the center of the proteasome) larger than those of the C-terminus, and therefore the ESP at the N-terminus smaller than at the C-terminus. However, the two experimental ESP maps for this proteasomal helix consistently show otherwise that there is a large-to-small gradation from N-to-C direction in spite of the presence of the two strong counteractive factors. This observation indicates that the gradation is likely due to the intrinsic property of an aligned helix dipole in this helix. This example extends beyond the predicted accumulative limit of  $\alpha$ -helix dipoles mentioned before.<sup>1</sup>

It should be emphasized that other helices in the proteasome or BGal proteins do not exhibit the

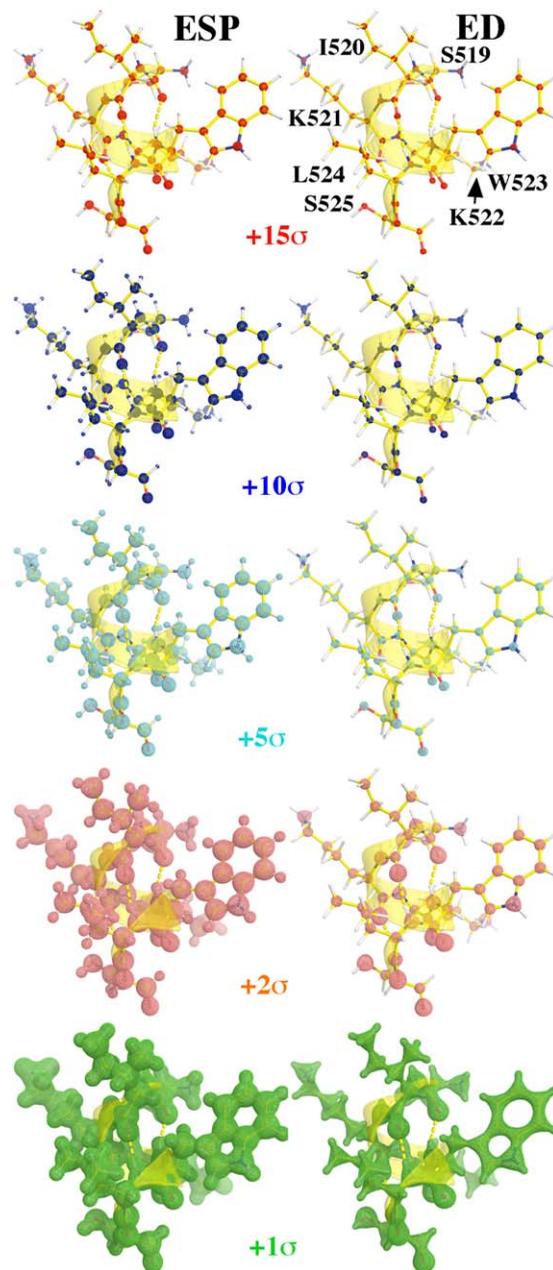
same gradation features mentioned above. Occasionally, an opposite ESP gradation is observed due to specific capping interactions<sup>2,7</sup> or due to limited angular resolution in image reconstruction. Moreover, we note that the ESP map reported at 3.2-Å resolution for the photosystem II-light harvesting complex II supercomplex (EMD-6717 and PDB-3JCU) does not yield a unified view of ESP gradation along  $\alpha$ -helices, either.<sup>20</sup> This is rather unexpected since one might expect some insights from experimental ESP data into the ESP basis for the positive-inside rule in the topology of transmembrane helices.<sup>21–23</sup> It remains to be analyzed whether an overly sharpening of the experimental ESP maps could be distorting the ESP function.

### DFT-derived ESP and electron density (ED) maps

ESP and ED functions for a model  $\alpha$ -helix, based on the BGal X-ray crystal structure,<sup>16,24</sup> were calculated at the DFT level at  $\sim 0.27$  Å resolution as described in “Methods”. The model comprises residues S519–S525 in a continuum model aqueous-phase environment ( $\epsilon = 78$ ), which should exhibit more screening than inside the folded BGal protein ( $\langle \epsilon \rangle = 4$ ) (Figs. 1, 2). Figure 2 shows the side-by-side comparison of ESP and ED functions at various contour levels, normalized to their respective standard deviations. Figure 2 also shows, for example, that H atoms are visible in the ESP function at the  $\sim +15\sigma$  contour level, clearly resolved from their partnering bonding atoms. In the ED functions (Fig. 2), however, the H atoms are visible only at  $\sim +1\sigma$ , but remain spatially unresolved from the peaks of partnering bonding atoms.

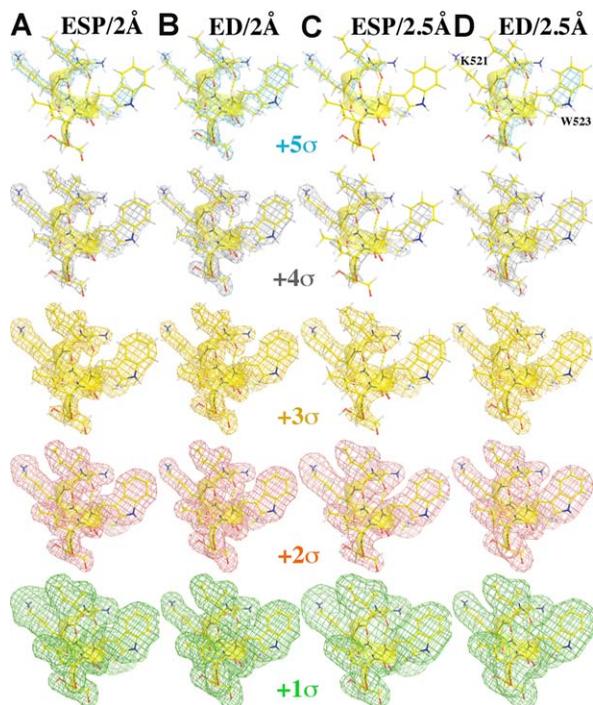
The results shown in Figure 2 are consistent with known scattering factors of neutral atoms for both X-ray diffraction and electron imaging. In X-ray crystallography, the optimum resolution for resolving H atoms is  $\sim 0.8$ -Å, with a maximum of the signal-to-noise ratio in experimental ED maps.<sup>25</sup> The H signal decreases with decreasing resolution, due to a loss of spatial resolution, and also decreases at increasing resolution due to the rapid loss of H scattering power relative to non-H atoms. Therefore, crystallographers often truncate the resolution for revealing the presence of H atoms and their locations.<sup>26</sup> We note in the ESP functions that H peaks are clearly resolved along covalent bonds but not in the ED functions (i.e., below the noise level of Fourier series termination) (Supporting Information Figs. S1, S2). The electron loss on H reduces the ED peaks but increases the ESP peaks due to a net increase of positive charge.

A Gaussian-smoothing function applied in reciprocal space to structure factors of Fourier transforms of maps dampens the contributions of high-resolution terms of both DFT-derived ESP and ED maps. With proper correction of motion B-factors for a given resolution (see “Methods”), the resolution of



**Figure 2.** DFT-derived ESP (left) and ED (right) functions for a stationary helix contoured at  $+15\sigma$  (red, top),  $+10\sigma$  (blue),  $+5\sigma$  (cyan),  $+2\sigma$  (salmon), and  $+1\sigma$  (green, bottom).

DFT-derived ESP and ED maps can be reduced to 2.0 and 2.5 Å, respectively (Fig. 3). Interestingly, the ESP map shows no visible feature for the side chain of W523 at 2.5-Å resolution contoured at  $+5\sigma$  level whereas there are some discontinuous features at 2.0-Å resolution. At either resolution, the side chain of K521 exhibits strong, continuous features in the DFT-derived ESP map (Fig. 3). W523 has lower ESP because it is near the C-terminus of the helix, away from the two positively charged side-chains of K521 and K522 (Fig. 3). K521 has higher ESP because it is near the N-terminus of the helix, next to K522. For comparison, the side chains of K521, K522, and



**Figure 3.** DFT-derived ESP (A,C) and ED (B,D) functions are truncated to 2.0-Å resolution (A,B) and 2.5-Å resolution (C,D) contoured at  $+5\sigma$  (cyan, top),  $+4\sigma$  (silver),  $+3\sigma$  (yellow),  $+2\sigma$  (salmon), and  $+1\sigma$  (green, bottom).

W523 exhibit the same level of ED features at 2.0-Å resolution, independently of their local environments [Fig. 4(A)].

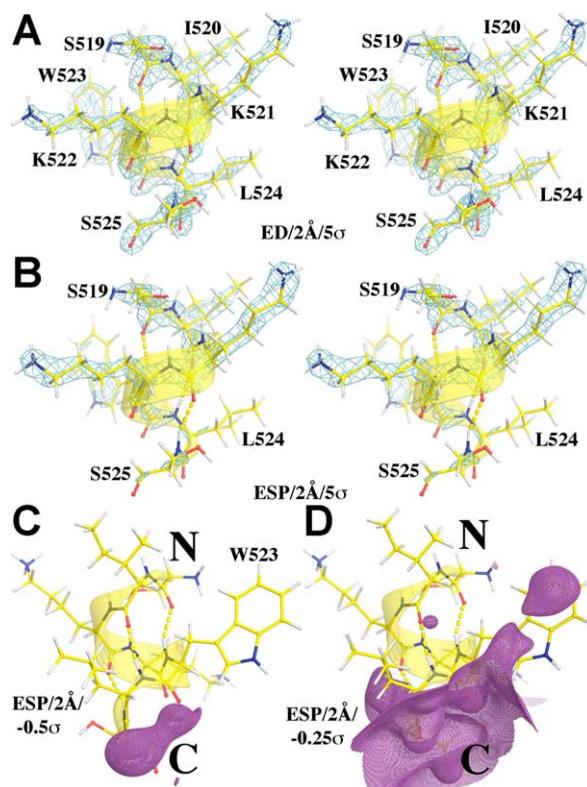
Figure 4 shows that the overall ESP value gradually decreases from the N-to-C terminus. The first residue of the helix (i.e., S519) has strong, continuous ESP features, clearly visible at  $+5\sigma$ . However, the last residue (i.e., S525) has virtually no significant ESP feature, nothing in this level [Fig. 4(B)]. Similar effects are observed when comparing I520 near the N-terminus and L524 near the C-terminus [Fig. 4(B)]. The features in the DFT-derived ESP function of this helix are largely consistent with what has been observed in the experimental ESP map in BGal (Fig. 1).<sup>16</sup> It is clear that negative features are observed at the C-terminus of the helix extending toward the W523 side chain [Fig. 4(C,D)], as shown by the negative ESP features when the DFT-derived ESP map is examined at  $-0.5\sigma$  or  $-0.25\sigma$  (it should be noted that the reliability of any negative values in the experimental ESP maps needs further investigation). In the absence of any negative side-chain atomic charge, the negative ESP generated by aligned dipole moments should be cylindrically symmetric along the helical axis. The two positive charges of K521 and K522 side chains are responsible for its altered appearance so that extra negative ESP features generated by the aligned dipoles extend only towards the W523 side-chain away from the two positive side chains.

In contrast in the ED function, the residues S519 and S525 have exactly the same features and so have residues I520 and L524, largely independent of their local environment [Fig. 4(A)].

### Modeling long-range electrostatic interactions and electron structure factors

Charges inside the low dielectric environment of a folded protein can produce long-range effects, in contrast to effectively shielded charges when exposed to aqueous environment as typically modeled by the Poisson–Boltzmann equation.<sup>27</sup> Therefore, the long-range effects of electrostatic interactions must be properly modeled because they are always present.<sup>4–6</sup> They often play essential roles in biological systems by altering the distribution of counter-ions, or polarizable  $\pi$ -conjugated systems, or changing the  $pK_a$  values of acids/bases or redox potentials of redox cofactors.<sup>4–6</sup> In the past, the effects of cumulative dipoles of the protein backbone were not sufficiently investigated, partially due to the lack of suitable experimental methods that could directly probe the ESP function.

Given the distinct nature of X-ray diffraction and electron scattering processes, significant differences in structure scattering factors should be considered for the interpretation of ED and ESP maps.



**Figure 4.** Close-up of ED (A) and ESP (B–D) functions at 2.0 Å resolution contoured at  $+5\sigma$  (stereodiagrams in A, B),  $-0.5\sigma$  (C), and  $-0.25\sigma$  (D). An N-to-C dipole of  $\alpha$ -helix is oriented from top to bottom.

As shown here, long-range interactions due to partial atomic charges inside proteins can clearly extend throughout an entire helix, far beyond 10 Å in spite of screening effects, consistent with an early suggestion.<sup>1</sup> Such a range is not covered by conventional fast Fourier transform (FFT) procedures applied for calculations of structure factors from models or during model refinement in X-ray crystallography. Typically, atomic ED functions are considered to be highly local with the total electrons of each atom estimated to be >99% inside the 2.5 Å Shannon radius of the atom when the atom is either stationary or has an atomic motion B-factor at sub-Å resolution. In this case, Nyquist–Shannon theory<sup>28,29</sup> is applicable in FFT where the contributions to the ED function from any atom outside the Shannon radius can completely be ignored. However, such an omission in the FFT calculation will certainly lead to incorrect ESP results from atomic models after neglecting the long-range ESP contributions as observed and discussed in this study. In the FFT procedure, structure factors are calculated for all reflections at once from Fourier inversion of an intermediate step of a density map, which is calculated point-by-point for each real space point from only a few atoms within Shannon radius of that point. To preserve any long-range ESP feature, we always apply direct Fourier summation in which structure factors are calculated one reflection at a time through summation of all the atoms in atomic model rather than through an intermediate step of a map.<sup>9–11</sup>

## Methods

Atomic positions for the  $\alpha$ -helix formed by amino acid residues S519–S525 of BGal were extracted from the X-ray crystal model 4TTG after rigid-body translation/rotation into the ESP map.<sup>16,24</sup> The positions of H atoms were determined by geometry optimization at the DFT level using the B3LYP/6–31G(d) functional with an implicit solvent model (SMD).<sup>30,31</sup> Electronic structure calculations were performed using Gaussian 09.<sup>32</sup> ESP maps were computed following the Merz–Singh–Kollman scheme on a rectangular grid with a density of 6 points per Bohr radius,<sup>33,34</sup> and ED maps were generated using the same grid. We have verified that the use of a finer grid (12 points/Bohr) does not affect the results. Application of the atomic spreading function using Gaussian function at reduced nominal resolution has been carried out using the CCP4 program for Fourier inversion and Fourier synthesis afterwards.<sup>35</sup> During Fourier synthesis, uniform atomic B-factors were applied as described in earlier work ( $B = 27.50 \text{ \AA}^2$  at 2.0-Å resolution, and  $B = 41.25 \text{ \AA}^2$  at 2.5-Å resolution).<sup>9–11</sup> Figures were prepared with Pymol.<sup>36</sup>

## ACKNOWLEDGMENTS

VSB acknowledges NERSC for an allocation of high performance computing time.

## CONFLICT OF INTEREST

The authors declare no conflict of interest in publishing results of this study.

## REFERENCES

1. Hol WG, van Duijnen PT, Berendsen HJ (1978) The alpha-helix dipole and the properties of proteins. *Nature* 273:443–446.
2. Hol WG (1985) The role of the alpha-helix dipole in protein function and structure. *Prog Biophys Mol Biol* 45:149–195.
3. Gunner MR, Nicholls A, Honig B (1996) Electrostatic potentials in *Rhodospseudomonas viridis* reaction centers: implications for the driving force and directionality of electron transfer. *J Phys Chem* 100:4277–4291.
4. Gascon JA, Leung SSF, Batista ER, Batista VS (2006) A self-consistent space–domain decomposition method for QM/MM computations of protein electrostatic potentials. *J Chem Theory Comp* 2:175–186.
5. Sproviero EM, Newcomer MB, Gascón JA, Batista ER, Brudvig GW, Batista VS (2009) The MoD-QM/MM methodology for structural refinement of photosystem II and other biological macromolecules. *Photosynth Res* 102:455–470.
6. Ho J, Newcomer MB, Ragain CM, Gascon JA, Batista ER, Loria JP, Batista VS (2014) MoD-QM/MM structural refinement method: characterization of hydrogen bonding in the *Oxytricha nova* G-quadruplex. *J Chem Theory Comp* 10:5125–5135.
7. Sengupta D, Behera RN, Smith JC, Ullmann GM (2005) The alpha helix dipole: screened out? *Structure* 13:849–855.
8. Kuhlbrandt W (2014) Biochemistry. The resolution revolution. *Science* 343:1443–1444.
9. Wang J, Moore PB (2017) On the interpretation of electron microscopic maps of biological macromolecules. *Protein Sci* 26:122–129.
10. Wang J (2017) On the appearance of carboxylates in electrostatic potential maps. *Protein Sci* 26:396–402.
11. Wang J (2017) On contribution of known atomic partial charges of protein backbone in electrostatic potential density maps. *Protein Sci* 26:1098–1104.
12. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the Resp model. *J Phys Chem* 97:10269–10280.
13. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KMJ, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman P (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197.
14. Bader RFW (1990) *Atoms in molecules: a quantum theory*. Oxford; New York: Clarendon Press.
15. Cole DJ, Vilseck JZ, Tirado-Rives J, Payne MC, Jorgensen WL (2016) Biomolecular force field parameterization via atoms-in-molecule electron density partitioning. *J Chem Theory Comput* 12:2312–2323.
16. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S (2015) 2 Å resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science* 348:1147–1151.

17. Chang S, Head-Gordon T, Glaeser RM, Downing KH (1999) Chemical bonding effects in the determination of protein structures by electron crystallography. *Acta Crystallogr A* 55:305–313.
18. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10:584–590.
19. Danev R, Tegunov D, Baumeister W (2017) Using the Volta phase plate with defocus for cryo-EM single particle analysis. *Elife* 6:E23006.
20. Wei X, Su X, Cao P, Liu X, Chang W, Li M, Zhang X, Liu Z (2016) Structure of spinach photosystem II-LHCII supercomplex at 3.2 Å resolution. *Nature* 534:69–74.
21. Heijne G (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J* 5:3021–3027.
22. Wallin E, von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* 7:1029–1038.
23. Nilsson J, Persson B, von Heijne G (2005) Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins* 60:606–616.
24. Wheatley RW, Juers DH, Lev BB, Huber RE, Noskov SY (2015) Elucidating factors important for monovalent cation selectivity in enzymes: *E. coli* beta-galactosidase as a model. *Phys Chem Chem Phys* 17:10899–10909.
25. Laplaca SJ, Ibers JA (1965) Crystal and molecular structure of trisphenylphosphine rhodium carbonyl hydride. *Acta Crystallogr* 18:511–519.
26. Farrugia LJ (2014) Accurate H-atom parameters from X-ray diffraction data. *IUCrJ* 1:265–266.
27. Debye P, Huckel E (1923) Zur theorie der elektrolyte. *Phys Zeitschr* 24:185–186.
28. Nyquist H (2002) Certain topics in telegraph transmission theory (reprinted from *Transactions of the A. I. E. E.*, February, pg 617–644, 1928). *IEEE* 90:280–305.
29. Shannon CE (1998) Communication in the presence of noise (reprinted from the *Proceedings of the IRE*, vol 37, pg 10–21, 1949). *IEEE* 86:447–457.
30. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652.
31. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* 113:6378–6396.
32. Gaussian 09 Revision D.01 F. M. J.; Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ. Gaussian, Inc., Wallingford CT, 2009.
33. Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. *J Comput Chem* 5:129–145.
34. Besler BH, Merz KM, Kollman PA (1990) Atomic charges derived from semiempirical methods. *J Comput Chem* 11:431–439.
35. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D* 67:235–242.
36. DeLano WL (2002) The PyMOL Molecular Graphics System. Version 1.8. Schrödinger, LLC.