# Protein structural ensembles are revealed by redefining X-ray electron density noise

P. Therese Lang[a], James M. Holton[b,c], James S. Fraser[a,1], and Tom Alber[a,2]

[a]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; [b]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158; and [c]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

To increase the power of X-ray crystallography to determine not only the structures but also the motions of biomolecules, we developed methods to address two classic crystallographic problems: putting electron density maps on the absolute scale of $e^-/\text{Å}^3$ and calculating the noise at every point in the map. We find that noise varies with position and is often six to eight times lower than thresholds currently used in model building. Analyzing the rescaled electron density maps from 485 representative proteins revealed unmodeled conformations above the estimated noise for 45% of side chains and a previously hidden, low-occupancy inhibitor of HIV capsid protein. Comparing the electron density maps in the free and nucleotide-bound structures of three human protein kinases suggested that substrate binding perturbs distinct intrinsic allosteric networks that link the active site to surfaces that recognize regulatory proteins. These results illustrate general approaches to identify and analyze alternative conformations, low-occupancy small molecules, solvent distributions, communication pathways, and protein motions.

electron number density | refinement against perturbed input data | protein dynamics | molecular motions | Ringer

For the last half-century, X-ray crystallography has played a critical role in elucidating the 3D structures of biological molecules. Although crystalline enzymes are often active and crystalline proteins show many dynamic features, X-ray diffraction data are generally interpreted in terms of a single dominant model. Efforts to characterize the full range of motions accessible in protein crystals have been hampered by uncertainty about whether weak electron density represents small populations of alternative conformations or noise from experimental and model errors (1–3). The standard practice of calculating electron density on a relative scale compounds this problem, because different maps cannot be compared directly to identify potentially meaningful features missed by the simplifications of structural models.

Electron density maps are contoured on a relative scale, because X-ray crystallographic diffraction experiments cannot measure a key, forward-scattered reflection that is swamped by the transmitted beam. The structure factor of this reflection, $F_{000}$, is equal to the total number of electrons in the unit cell, including the contribution from disordered solvent (4). Because crystals differ in composition, the absence of $F_{000}$ puts each map on a different scale. The standard practice to circumvent this limitation is to represent electron density in relative units of the rms deviation of map values from the mean density (1). These "σ-scaled" maps are sufficient for structural modeling, but it is difficult to determine which density features are signal vs. noise because the σ unit has little to do with the uncertainty in the electron density. It is also impossible to quantitatively compare features in different maps, because the scale and offset relating σ to the absolute electron density varies among crystals of different molecules or even of the same molecular species with different symmetries or crystallization solvents.

Here we introduce computational methods to place electron density maps on a common absolute scale and to calculate the noise at each position in the map. By applying these methods to a diverse set of 685 structures from the Protein Data Bank (PDB), we find that noise varies with position and is substantially lower than the currently accepted threshold for modeling. Above the noise, in a range of electron density that is generally ignored, high-resolution electron density maps contain evidence for unmodeled, low-occupancy ligands, side-chain rotamers, and ensemble shifts. These results illustrate the utility of defining the absolute scale of electron density for characterizing protein conformational distributions.

## Results

**Converting Electron Density to the Absolute Scale.** To enable quantitative comparisons among electron density maps, we developed a computational method to calculate $F_{000}$ and render maps on the absolute scale in units of $e^-/\text{Å}^3$ (*Materials and Methods*). These electron number density (END) maps were calculated by scaling the experimental structure factors ($F_{obs}$) to structure factors calculated from the model ($F_{calc}$, which are intrinsically on an absolute scale) and adding the average electron density of the crystal (including bulk solvent and the ordered model) to each map voxel. In contrast to σ-scaled maps, where zero corresponds to the average electron density, in END maps, zero corresponds to vacuum (Fig. S1). This automated computational method closely reproduced independently estimated $F_{000}$ values (5, 6) (Table 1), verifying the accuracy of the approach.

---

**Significance**

This work presents computational solutions to two long-standing problems in protein structure determination using X-ray crystallography. Together, these methods reveal that the electron density threshold for discovering alternative protein and ligand conformations is much lower than the standard cutoff for structural modeling. Three broad applications illustrate that the features present in weak electron density can reveal important, unanticipated conformational heterogeneity in proteins. The methods introduced here help convert X-ray crystallography from the principal technique to obtain "snapshots" of biological molecules to an approach that also can reveal the signatures of molecular motions that are potentially important for function. These advances have broad implications for developing drugs and understanding protein mechanisms.

---

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

**Table 1. Comparison of $F_{000}$ calculated for END maps to $F_{000}$ determined by other methods**

| System | Method | $F_{000}$ ($e^- \times 10^4$) | Difference (%) |
|---|---|---|---|
| T4 lysozyme* | END map | 13.1 | 7.2 |
| | Analytical[†] | 12.2 | |
| IL-1β[‡] | END map | 9.57 | 3.1 |
| | Analytical[§] | 9.28 | |
| Scorpion Toxin II[¶] | END map | 2.44 | 5.6 |
| | Theoretical[‖] | 2.31 | |

*PDB ID 3DKE.

[†]Bulk solvent contribution to $F_{000}$ calculated using the measured density of the crystallization conditions (1.20 g/cc or 0.391 $e^-$/Å$^3$) and the percent of unit cell occupied by bulk solvent (39.1%) (5).

[‡]PDB ID 2NVH.

[§]Bulk solvent contribution to $F_{000}$ calculated from the measured density of crystallization conditions (0.383 $e^-$/Å$^3$) and the percent of unit cell occupied by bulk solvent (64.7%) (6).

[¶]PDB ID 1AHO.

[‖]Structures were obtained from a molecular dynamics simulation in which 12 copies of the unit cell were preserved (25). The theoretical $F_{000}$ was calculated by dividing the sum of all atomic numbers by 12.

**Calculating Noise at Each Map Position.** With the electron density maps on the absolute scale, we searched for an approximate electron density threshold to distinguish signal from noise. In contrast to recent theoretical treatments (7–9), the noise level at every position in the unit cell was determined by empirically propagating errors from the structure factors into the electron density map. In this general analytical approach, which we call refinement against perturbed input data (RAPID), errors in the experimental measurements [$\sigma(F_{obs})$] or the model ($|F_{obs} - F_{calc}|$) (10) were used to add simulated noise to $F_{obs}$ before rerefining the structure. Over several trials using different random number seeds, the RMS change in electron density observed at each point in the map in response to the changes in $F_{obs}$ was used to calculate a RAPID map of the spatial distribution of errors in the electron density (Fig. S2A). To evaluate the typical levels of noise, we calculated END and RAPID maps for 685 representative protein structures at 1.0- to 3.5-Å resolution. Across these structures, the average value for the RAPID map based on model error (0.12 $e^-$/Å$^3$) is higher than the corresponding average experimental error (0.037 $e^-$/Å$^3$) in all but two cases (Fig. 1A and Fig. S2B). RAPID maps reveal that noise is variable throughout the unit cell and highest under molecular features (Fig. S2C).

Comparing the contour in END maps equivalent to the 1 σ contour typically used for model building revealed that the 1 σ contour varies by over twofold between 0.4 and 1.0 $e^-$/Å$^3$ (Fig. 1A). Thus, 1 σ in standard maps of different molecules can represent substantially different numbers of electrons, and the same contours in different σ-scaled maps generally are not comparable. END maps overcome this problem. Importantly, the maximum RMS errors estimated by the RAPID procedure were generally six- to eightfold lower than σ (Fig. 1B), indicating that the standard 1 σ threshold for model building overestimates the noise in electron density.

**Modeling a Partially Occupied Inhibitor.** To exemplify the utility of the newly defined noise distributions to search for molecular features, we explored the 1.50-Å resolution END and RAPID maps of the HIV capsid protein bound to a drug-like molecule (PDB ID 2PXR). In standard maps on a relative scale, the ligand
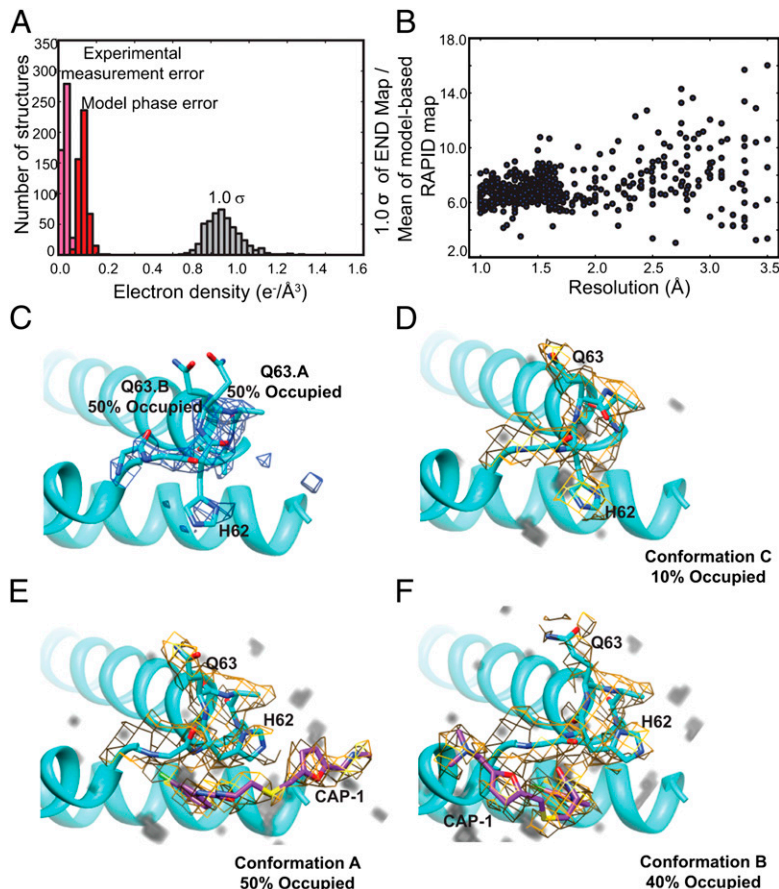


Fig. 1. END and RAPID maps define the absolute scale and noise level of electron density and expose a hidden ligand. (A) Histogram of END values corresponding to standard modeling threshold (1 σ above the mean; gray) in 485 high-resolution (≤1.7 Å) structures. Mean values for |Fo-Fc| RAPID maps represent errors from the model (red), and mean values for σ($F_{obs}$) RAPID maps represent experimental error (rose). (B) The 1 σ threshold overestimates noise by six-to eightfold. Ratio of the 1 σ value to the average value of the noise due to model errors determined by the RAPID procedure for 685 representative structures in the Protein Data Bank. (C) Standard σ-weighted map contoured at 1 σ (blue mesh) shows weak, uninterpretable electron density features in the CAP-1 binding site of HIV capsid protein. (D–F) Electron density for END (orange mesh; 0.5 $e^-$/Å$^3$) and RAPID (gray solid; 0.5 $e^-$/Å$^3$) maps shows where CAP-1 binds to the HIV capsid protein. The lowest occupancy protein conformation (10%) resembles the original model. The ligand was built in two conformations (at 50% and 40% occupancy). Shifts in His62 and Gln63 accommodate ligand binding.

Lang et al.

was not apparent at the conventional 1 σ threshold in the electron density of the complex (Fig. 1C and Fig. S3B). The inhibitor binding mode was identified in solution using NMR (PDB ID 2JPR) (11), which provides a check on the potential influence of model bias (12). In contrast to the discontinuous 1 σ density, the END map contoured at 0.5 $e^-$/Å$^3$, more than twice the mean RAPID noise level of 0.24 $e^-$/Å$^3$, revealed continuous electron density for the inhibitor, as well as shifts in the HIV capsid protein that accompany binding (Fig. 2 D–F and Fig. S3B). Including the low-occupancy ligands and alternate loop conformations reduced the average B-factors for the protein atoms in the binding site and slightly improved the $R$ and $R_{free}$ values of the model (Fig. S3). One of the conformations in the refined crystallographic model of the inhibitor superimposes well on the NMR model (Fig. S3C). This example supports the conclusion that the gap between the standard 1 σ modeling threshold and the RAPID map noise level contains information about low-occupancy structures.

**Alternative Side-Chain Conformations.** To gauge the amount of protein side-chain structural polymorphism detectable above the noise, we calculated END and RAPID maps for a set of 485 structures at 1.7-Å resolution or better. This resolution cutoff was chosen to ensure that discrete alternative side-chain conformations could be resolved. To automatically identify small populations of unmodeled conformations, we used the program Ringer (13) to systematically sample the electron density around side-chain dihedral angles (χ angles; Fig. S4 A and B). Peaks in the END maps falling below the noise level defined by the RAPID maps were ignored. We applied Ringer to 113,285 side chains unbranched at χ1 and found evidence for peaks at rotameric positions for 98.7% of residues, suggesting that sampling down to the noise level detects both alternate conformations and hydrogens (Fig. 2A and Fig. S4C). Above 0.4 $e^-$/Å$^3$, the discovery rate of unmodeled peaks was higher for the unbranched side-chains than for alanine residues, which lack a Cγ, suggesting that these features reflect alternative side-chain conformations (Fig. 2B).

To test this idea, we built an additional χ angle from the secondary peaks identified by Ringer (Fig. 2C). A 3D histogram of secondary (unmodeled) χ1 and added-χ2 peaks identified by building from the unmodeled χ1 peak center produced a checkerboard pattern that is enriched in rotameric positions (Fig. 2D; P < 0.0001). In contrast, a similar analysis of alanine peaks above the noise showed a tripartite χ1 distribution expected for staggered hydrogens and a random pseudo-χ2 distribution, corresponding to noise (Fig. 2E). Strikingly, eliminating alanine peaks below 0.4 $e^-$/Å$^3$ suppressed nearly all of the features on the χ1:pseudo-χ2 plot (Fig. 2F). These results suggest that the END method placed the maps on a common scale (cf. Fig. 2 E and F), that RAPID maps capture the major sources of noise (hydrogen signature is visible in Fig. 2E), and that features above 0.4 $e^-$/Å$^3$ are enriched for heavy atoms over hydrogens (cf. Fig. 2 D and F) and reflect small populations of alternative conformations. Using this approach, which involves directly sampling the electron density rather than building and refining alternative structures, we detected unmodeled, rotameric structural heterogeneity at 45% of side-chains.

**Allosteric Communication Networks in Protein Kinases.** Recent studies have emphasized that specific regions of enzymes undergo coupled motions that define functional transitions (14–17). Protein kinases, for example, are crucial signaling enzymes that catalyze the transfer of the ATP γ-phosphate to specific substrates, switching the activities of thousands of proteins in the cell. As such, a critical problem is to determine how allosteric regulators influence protein-kinase active sites. To address this question, we explored the consequences of ATP binding on the conformational ensembles of protein kinases visualized by high-resolution X-ray crystallography.

The effect of Mg$^{2+}$-ATP binding on kinase motions was identified and analyzed using END and RAPID maps of calmodulin-
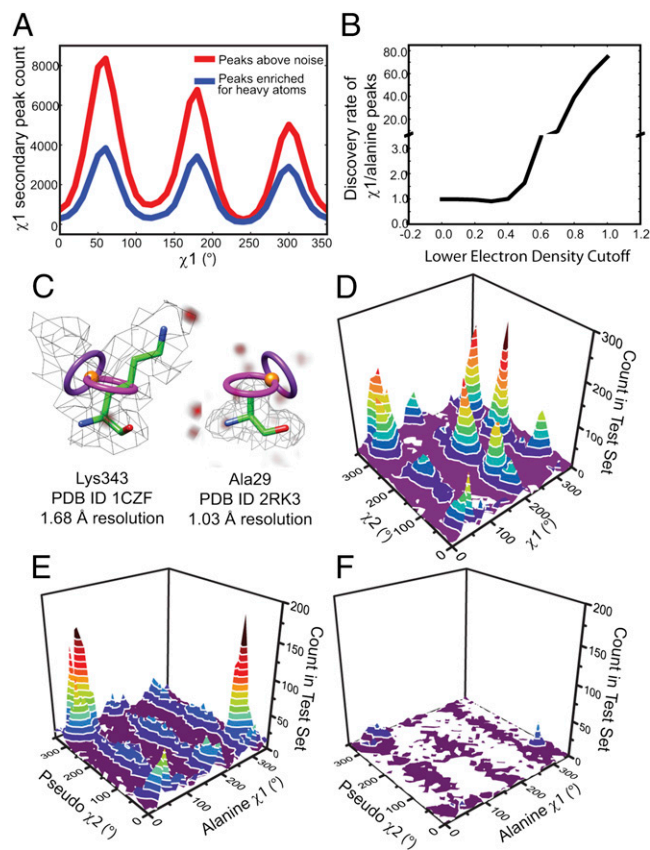


**Fig. 2.** END and RAPID maps reveal signal for unmodeled, low-population side-chain conformations. (*A*) Histogram of unmodeled, secondary χ1 electron density peaks from Ringer plots above the noise (red) and above 0.4 $e^-$/Å$^3$ (blue). (*B*) The discovery rate, calculated as the ratio of the number χ1 secondary electron density peaks (normalized by the total number of χ1 side-chains) to alanine primary electron density peaks (normalized by the total number of alanines), plotted vs. the lower electron density cutoff ($e^-$/Å$^3$) in END maps of 485 1.0- to 1.7-Å resolution structures. Values above 0.4 $e^-$/Å$^3$ enrich for alternative, low-occupancy structural features. (*C*) Unmodeled conformations were detected by identifying electron density correlations in dihedral space. Peaks in the electron density (gray mesh; 0.4 $e^-$/Å$^3$) above the noise (red solid; 0.4 $e^-$/Å$^3$) were identified by sampling χ1 (pink ring) and χ2 (purple ring) at idealized heavy-atom bond lengths from the χ1 secondary peak (orange sphere). (*D*) Correlated unmodeled χ1 and χ2 peaks for side-chains unbranched at χ1 in 485 high-resolution structures. 3D histogram of correlated secondary χ1 Ringer peaks and primary χ2 Ringer peaks built from the unmodeled secondary χ1 peaks for 31,086 side-chains unbranched at χ1. The nonrandom, low-energy, checkerboard distribution suggests that unmodeled side-chain conformations are common. (*E*) For alanine residues in 485 high-resolution structures, histogram of χ1 and pseudo-χ2 Ringer peaks above the RAPID noise. The columns of peaks at χ1 = 60°, 180°, and 240° suggest that the hydrogens are staggered, the noise peaks are distributed randomly around the pseudo-χ2, and there is no missing source of noise that swamps the hydrogen signals. The strong cross-peaks in the left, right, and front corners come from the backbone amide hydrogen. (*F*) For alanines in the set of 485 structures, the histogram of χ1 and pseudo-χ2 Ringer peaks above 0.4 $e^-$/Å$^3$. The suppression of features compared with *E* provides confidence that the END maps were on a similar scale and the 0.4 $e^-$/Å$^3$ threshold effectively enriches for alternative conformations of longer side-chains.

activated human death associated protein kinase (DAPK) (18), cyclin-dependent kinase 2 (CDK2) (19), casein kinase 2α (CK2α) (20), and ephrin type-A receptor 3 (EphA3) (21) (Table S1). Changes in side-chain ensembles between maps of free and ATP-bound forms were calculated using X-ray data from the PDB. Side-chain ensemble shifts were not detected in the published

molecular models (Fig. S5), but rather using the correlation coefficients (ccs) of the Ringer (13, 22) plots of the END maps (Fig. 3 and Figs. S6 and S7). Shifts were confirmed in the electron density by visual inspection. ATP binding coupled residues across the active site cleft of all four enzymes, as observed in NMR studies of cAMP-dependent protein kinase (PKA) in which $Mg^{2+}$-ATP alters the distribution and transitions of alternative kinase structures before phosphotransfer (14, 16). In DAPK, CDK2, and CK2α, however, the changes associated with ATP binding propagated in distinct directions and resulted in connected cascades across more than 20 Å.

Although these kinases share 24–29% sequence identity and adopt the same fold, the structural perturbation pathways differ. The majority of the side-chain ensemble rearrangements in DAPK are communicated from the ATP binding site toward the calmodulin sensing surface in the C-lobe (23) (Fig. 3 A and E and Figs. S6A, S7A, and S8A), whereas the majority of the CDK2 perturbations reach toward the cyclin binding surface on the back of the N-lobe (Fig. 3 B and F and Figs. S6B, S7B, and S8B) (24). Similarly, ATP binding to CK2α results in coupled ensemble shifts that extend through the N-lobe to the distinct recognition site for the CK2β regulatory subunit (Fig. 3C and Figs. S7C and S8C). Assembly of the heterotetrameric CK2 holoenzyme stabilizes the active conformation of CK2α (20). In contrast, ensemble shifts upon ATP binding are localized to the nucleotide-binding site in EphA3 (Fig. 3D and Figs. S7D and S8D), which is activated not by "remote control," but rather by release of the juxtamembrane segment from the active site, which causes localized intramolecular conformational changes (21).

## Discussion

In contrast to the view that X-ray crystal structures provide static snapshots of proteins, a variety of studies have emphasized the signatures of dynamics in crystallographic images. Weak electron density features visualized after the end of standard refinement in the high-resolution crystal structures of proline isomerase, Ras and dihydrofolate reductase, for example, have recently revealed functional ensembles that matched expectations from NMR data and mutagenesis (15, 22, 25). Modeling these additional conformations has a minor effect on conventional parameters such
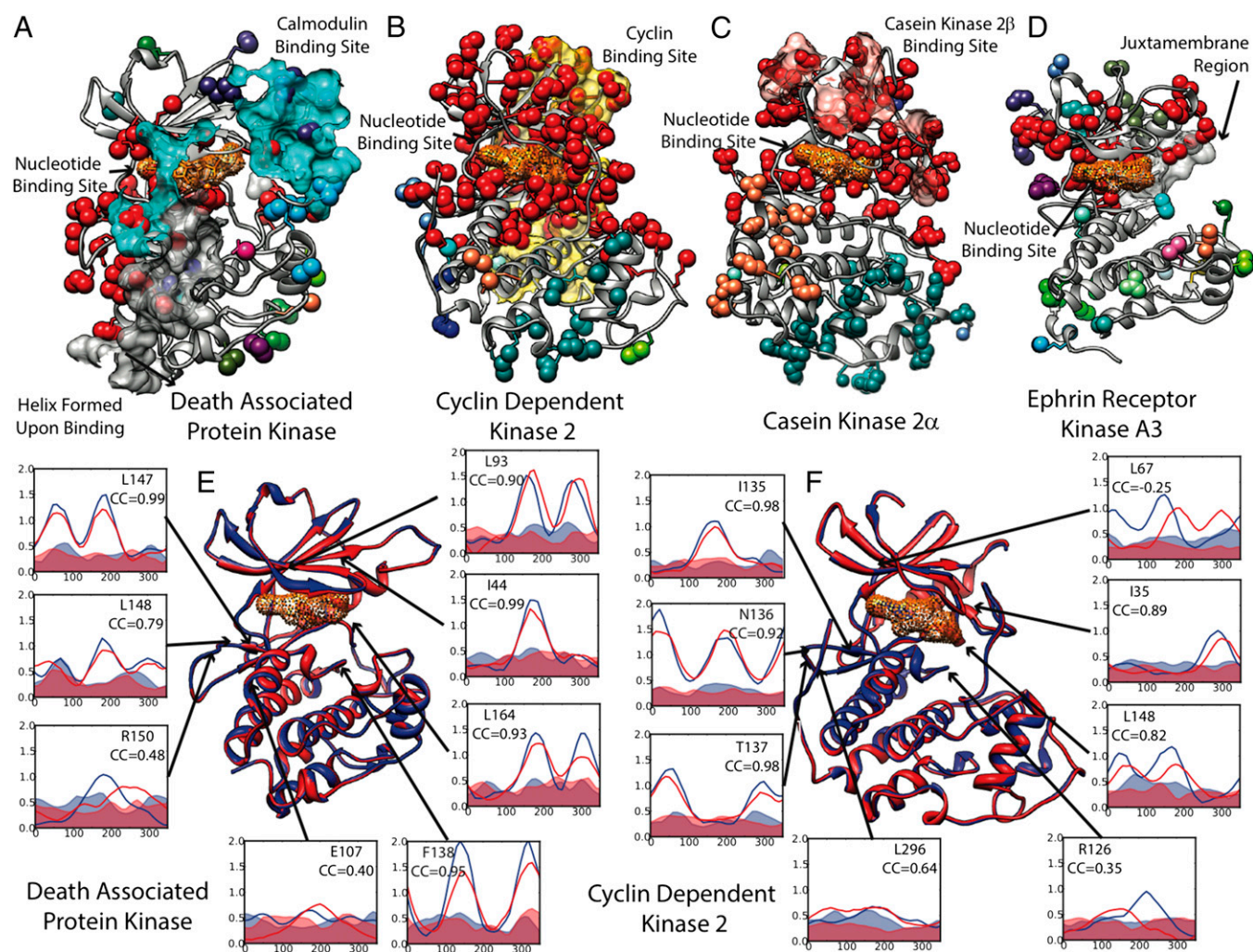


**Fig. 3.** ATP perturbs different allosteric networks in human protein kinases (A) DAPK, (B) CDK2, and (C) CK2α, but causes localized ensemble shifts in (D) EphA3. Residues with Ringer ccs < 0.85 (spheres) between free and ATP-bound (orange surface) electron density maps indicate shifts in the side-chain ensembles or rotamer flips. Side-chains that cluster within 4 Å of the nucleotide (red) show allosteric networks connected to the active site. Other clusters are shown in different colors. The surfaces show atoms in the regulatory protein that are within 4 Å of the kinase. Representative Ringer plots for key residues from the nucleotide-free (blue line) and bound (red line) END maps for (E) DAPK and (F) CDK2. Ringer plots for the apo (blue fill) and bound (red fill) RAPID maps are shown, as well, to indicate the distribution of noise. Residues illustrate the coupled ensemble shifts in the allosteric pathways for each protein in response to $Mg^{2+}$-ATP binding, as well as the corresponding residue in the other protein as a control. The perturbations connect the kinase active site to distinct regulatory surfaces.

as $R_{\text{free}}$, suggesting the need for alternate metrics and methods of detecting and representing conformational heterogeneity (26, 27). To better evaluate the extent of polymorphism in the PDB and develop tools to assess structural heterogeneity, we developed the END and RAPID methods to place the electron density on an absolute scale and calculate the noise at each position in a map. In a set of 685 representative structures, we found that noise varies with position, is dominated by errors in the model, and is generally six to eight times lower than the current threshold for modeling. In the HIV capsid–capsid inhibitor 1 (CAP-1) complex, weak electron density above the noise could be modeled with multiple ligand conformations.

In END and RAPID maps of high-resolution structures, 45% of side-chains showed electron density peaks with the stereochemical signatures of small populations of alternative, low-energy conformations. These signals were apparent above $0.4\ e^-/\text{Å}^3$ in the electron density and were not the result of explicit modeling of additional conformations. The estimate of 45% of side chains with unmodeled alternative conformations is ninefold higher than the ~5% of structurally polymorphic residues in current crystallographic models and also higher than the 18% of polymorphic residues detected using σ-scaled electron density (13). Because this procedure ignores σ backbone shifts and conformations with signals below $0.4\ e^-/\text{Å}^3$, these results represent a lower bound on the amount of structural polymorphism in protein crystals. The increase in the number of alternative conformations detected reflects access to signal that is gained using END maps. The signals for these conformations not only correspond mostly to side-chains populating rotameric angles, but they also occur above the local noise levels in RAPID maps, indicating that they reflect dynamic structural features rather than noise.

To explore how regulatory signals propagate to the active site of protein kinases, we probed how a perturbation at the active site, the binding of the ATP substrate, alters conformational distributions in DAPK, CDK2, CK2α, and Eph3a. As a measure of ensemble shifts that is more sensitive than available models, we calculated the correlation coefficient of the electron density distribution around each side-chain dihedral angle between the free and ATP-bound structures. Strikingly in these four kinases, these calculations revealed ensemble shifts coupled to the ATP that link the active site to distinct functional regulatory surfaces. These distinct shifts are candidates for intrinsic allosteric communication networks between the active and regulatory sites. Although DAPK, CDK2, and CK2α were not in the active complex with the cognate regulatory protein, these patterns suggest how substrate binding can perturb the potential intrinsic communication networks before interactions with regulators. The distinct structural responses of DAPK, CDK2, and CK2α to ATP binding suggest that networks and motional boundaries can differ in homologous proteins along their functional trajectories. Recent studies have emphasized that residues that covary in evolution define regions in protein families that correspond to functional sectors, raising the possibility that physical communication networks also are conserved (28). In contrast, analysis of END maps reveals that nucleotide binding to the protein kinases DAPK, CDK2, and CK2α generates distinct ensemble shifts that couple to different regulatory surfaces.

By enabling direct access to structural ensembles, END and RAPID maps provide tools to explore the roles of structural heterogeneity in macromolecular function and evolution. END and RAPID maps enable a unified quantitative interpretation of electron density that reveals not only low-occupancy ligands but also dynamic structural features and alternative solvent constellations. By analogy to the Beer–Lambert–Bouguer Law in spectroscopy, which defines the relationship between molecular concentration and optical absorbance, END maps report the concentration of scattering electrons at each point in space. A current challenge

remains to automatically model alternative conformations (3, 26, 27). This information about structural distributions in crystals, when critically analyzed, offers increased power to X-ray crystallography to facilitate inhibitor development, visualize structural ensembles, and connect macromolecular motions to functions. These capabilities open windows into biologically relevant information not included in current X-ray structural models.

## Materials and Methods

**Estimation of $F_{000}$ for END Maps.** $F_{000}$ was obtained by summing the total number of electrons in the coordinate model and the bulk solvent. An absolute scale-and-offset map for the coordinate model was obtained using the ATMMAP mode of SFALL from the CCP4 Suite11. The mean value of this map is $<\rho_{\text{atoms}}>$. The structure factors of the bulk solvent mask from phenix. refine12 were used to estimate $<\rho_{\text{bulk}}>$. The histogram of the density from these structure factors has a mean value of zero and two peaks: one above the mean, corresponding to solvent density, and one below the mean, corresponding to vacuum. The shift required to move the negative peak to 0.0, the true vacuum level, is $<\rho_{\text{bulk}}>$ (Fig. S1A). To obtain the END map, volume-scale map coefficients were specified from phenix.refine. Once the map was calculated from absolute-scale coefficients, adding the quantity $<\rho_{\text{atoms}}> + <\rho_{\text{bulk}}>$ to each map voxel converts the map to absolute electron number density $(e^-/\text{Å}^3)$.

The robustness of this method was evaluated by examining $F_{000}$ obtained over the course of automated building and refinement (Fig. S1C). Most of the deviations in $F_{000}$ arise from changes in the bulk solvent mask, but $F_{000}$ converges with increasing phase accuracy, as indicated by the likelihood weight (figure of merit). Previously, $F_{000}$ was measured analytically for crystals of IL-1β and T4 lysozyme (10, 11). Our END method reproduced these values within 3% and 7%, respectively (Table 1). In addition, we refined a model of scorpion protein toxin (sPT) against structure factors calculated from the average electron density of an all-atom molecular dynamics simulation of the crystal (29, 30). The sum of the total number of electrons for the structure and all solvent molecules in this simulation was within 5% of the END method $F_{000}$. The reproducibility of the END method was evaluated with a test set of hen egg white lysozyme crystallized over a range of resolutions (Fig. S1D). $F_{000}$ remained fairly consistent across this set (average of $9.9 \times 10^4 \pm 0.69 \times 10^4\ e^-/\text{Å}^3$ or 14% rmsd). These comparisons show that the END method for estimating $F_{000}$ is accurate and generally applicable.

**RAPID Maps.** The theoretical foundation for RAPID maps arises from the fundamental property of Fourier tranforms that the errors in the coefficients propagate into the errors in the function. The errors in electron density, $\sigma(\rho)$, arise from measurement errors $\sigma(F_{\text{obs}})$ and modeling errors ($F_{\text{obs}}$ vs. $F_{\text{calc}}$) (10). To account for the effects of measurement errors, $F_{\text{obs}}$ was perturbed by $\sigma(F_{\text{obs}})$, and a model was refined against these data. This process revealed the magnitude of the phase error due to $\sigma(F_{\text{obs}})$ and the corresponding change in ρ at each map voxel. The RMS change in ρ from this procedure yielded the contribution of $\sigma(F_{\text{obs}})$ to $\sigma(\rho)$. $F_{\text{obs}}$ may also be changed by an amount proportional to $\sigma(F_{\text{obs}})$ by simply adding noise and measuring the response induced in the map.

The contribution from phase error was defined as the change in the phase from a refined model in response to a change in target amplitude ($F_{\text{obs}}$). Absolute-scale values of $F_{\text{obs}}$, $\sigma(F_{\text{obs}})$, and $F_{\text{calc}}$ were obtained from the phenix.refine run used to generate the END map. $F_{\text{obs}}$ was perturbed using SFTOOLS in the CCP4 suite using the following formula:

$$F'_{\text{obs}} = F_{\text{obs}} + r \bullet \delta,$$

where r is a random deviate chosen from a Gaussian distribution with mean = 0 and SD = 1, and δ is either $\sigma(F_{\text{obs}})$ or $|F_{\text{obs}} - F_{\text{calc}}|$. Negative values of $F_{\text{obs}}$ were set to zero. The new set of $F'_{\text{obs}}$ was used to refine the atomic coordinates in phenix.refine, generating a new $2mF'_{\text{obs}}$-D$F_{\text{calc}}$ map (ρ'). This process was repeated five times, using different random seeds for r. The original map (ρ) was subtracted from the five new maps (ρ'), and the RAPID map value $\sigma(\rho)$ was defined as the RMS of all five ρ' − ρ values at each voxel (Fig. S2A). Five replicates were found sufficient to obtain $\sigma(\rho)$ to within 35% of the $\sigma(\rho)$ value from 500 replicates in the test case of PDB ID 2I4A. We note that this represents the uncertainty in the value of $\sigma(\rho)$ and not a 35% error in ρ. For situations where $\sigma(\rho)$ must be more precise than ρ itself, 50 or more replicates may be desirable.

To ensure that $\sigma(\rho)$ was of appropriate magnitude, we confirmed that the average value of each model-based RAPID map correlated well with the RMS value of the original $mF_{\text{obs}}$-D$F_{\text{calc}}$ map (Fig. S2B), as expected from Parseval's

theorem (31). In all but two examples with unusually high $\sigma(F_{obs})$, the average $\sigma(\rho)$ obtained by RAPID using $\sigma(F_{obs})$ was lower than that from $|F_{obs} - F_{calc}|$. To explore the distribution of noise throughout the RAPID map, we used Ringer (13) to sample side-chains unbranched at $\chi 1$ (Fig. 1C). Unlike the secondary peaks in the END maps, the primary peaks in RAPID maps were more randomly distributed, as indicated by high troughs between the slight rotameric peaks (Fig. S2C). Further analysis showed the rotameric peaks were predominately located under heavier atoms (e.g., oxygen and sulfur) but were still well below the signal level of the corresponding voxel in the END map. Secondary noise peaks showed no enrichment in $\chi$ angle space.

Phenix version 1.6.1 (32) and CCP4 version 6.1.3 (33) were used for calculations of both END and RAPID maps. A script to generate the maps is available at http://bl831.als.lbl.gov/END/RAPID/.

**Preparation of Test Sets.** Structures and X-ray data were obtained from the PDB (34). For the 485 X-ray crystal structures in the 1.0- to 1.7-Å resolution set, the $R$ values were less than 0.22, and mutual sequence identity was <95%. For the 1.0- to 3.5-Å resolution structures, the $R$ values were less than 0.1 times the resolution, the mutual sequence identity was <30%, and the molecular weight was <80,000 kDa. The list of PDB IDs can be found at http://ucxray.berkeley.edu/ringer/TestSets/testSets.htm.

Coordinate and structure factor files were converted and refined for five macrocycles using phenix.refine. When not available, $R_{free}$ flags were automatically generated. In addition to default parameters, automatic optimization of weights was enabled, as were anisotropic B-values for data better than 1.6-Å resolution. Hydrogens were added to models. The model for the

complex of CAP-1 with HIV capsid protein was built manually with Coot and subjected to further refinement with phenix.refine.

**Ringer Analysis.** END and RAPID maps were analyzed using Ringer 2.0 (13) and Chimera 1.4.1 (34), which has been adapted to sample absolute scaled maps and dynamically filter peaks above noise from RAPID maps. Unless otherwise stated, default parameters were used. The code for Ringer can be accessed at http://ucxray.berkeley.edu/ringer.htm. To identify connected clusters in the protein kinases, ensembles of side-chains were automatically built using qFit (26). The backbones of the qFit models for the free and ATP-bound structures were superimposed. All conformations of each residue, in both free and bound, identified by Ringer as shifting as a result of binding (cc < 0.85) were considered for clustering. Clusters were defined such that the all residues in a cluster are within 4 Å of each other using the python-cluster module version 1.1.b3 distributed by SourceForge. The 0.85 cc and 4-Å distance were chosen because they automatically identified the dynamic network characterized previously in proline isomerase (15).

1. Jensen LH (1997) *Macromolecular Crystallography, Pt B* (Academic Press, San Diego), Vol 277, pp 353–366.
2. Terwilliger TC, et al. (2007) Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models. *Acta Crystallogr D Biol Crystallogr* 63(Pt 5): 597–610.
3. Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN, Jr. (2007) Ensemble refinement of protein crystal structures: Validation and application. *Structure* 15(9):1040–1052.
4. James RW (1948) *The Optical Principles of the Diffraction of X-rays* (G. Bell and Sons Ltd, London).
5. Liu LJ, Quillin ML, Matthews BW (2008) Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc Natl Acad Sci USA* 105(38):14406–14411.
6. Quillin ML, Wingfield PT, Matthews BW (2006) Determination of solvent content in cavities in IL-1beta using experimentally phased electron density. *Proc Natl Acad Sci USA* 103(52):19749–19753.
7. Giacovazzo C, Mazzone A (2011) Variance of electron-density maps in space group P1. *Acta Crystallogr A* 67(Pt 3):210–218.
8. Giacovazzo C, Mazzone A, Comunale G (2011) Estimation of the variance in any point of an electron-density map for any space group. *Acta Crystallogr A* 67(Pt 4):368–382.
9. Tickle IJ (2012) Statistical quality indicators for electron-density maps. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):454–467.
10. Henderson R, Moffat JK (1971) The difference Fourier technique in protein crystallography: Errors and their treatment. *Acta Crystallogr D Biol Crystallogr* 27(7):1414–1420.
11. Kelly BN, et al. (2007) Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein. *J Mol Biol* 373(2):355–366.
12. Pozharski E, Weichenberger CX, Rupp B (2013) Techniques, tools and best practices for ligand electron-density analysis and results from their application to deposited crystal structures. *Acta Crystallogr D Biol Crystallogr* 69(Pt 2):150–167.
13. Lang PT, et al. (2010) Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci* 19(7):1420–1431.
14. Masterson LR, et al. (2010) Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat Chem Biol* 6(11):821–828.
15. Fraser JS, et al. (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462(7273):669–673.
16. Masterson LR, et al. (2011) Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. *Proc Natl Acad Sci USA* 108(17):6969–6974.
17. Lee HJ, Lang PT, Fortune SM, Sassetti CM, Alber T (2012) Cyclic AMP regulation of protein lysine acetylation in Mycobacterium tuberculosis. *Nat Struct Mol Biol* 19(8): 811–818.
18. Tereshko V, Teplova M, Brunzelle J, Watterson DM, Egli M (2001) Crystal structures of the catalytic domain of human protein kinase associated with apoptosis and tumor suppression. *Nat Struct Biol* 8(10):899–907.
19. Schulze-Gahmen U, De Bondt HL, Kim SH (1996) High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: Bound waters and natural ligand as guides for inhibitor design. *J Med Chem* 39(23):4540–4546.
20. Niefind K, Issinger OG (2010) Conformational plasticity of the catalytic subunit of protein kinase CK2 and its consequences for regulation and drug design. *Biochim Biophys Acta* 1804(3):484–492.
21. Davis TL, et al. (2008) Autoregulation by the juxtamembrane region of the human ephrin receptor tyrosine kinase A3 (EphA3). *Structure* 16(6):873–884.
22. Fraser JS, et al. (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc Natl Acad Sci USA* 108(39):16247–16252.
23. de Diego I, Kuper J, Bakalova N, Kursula P, Wilmanns M (2010) Molecular basis of the death-associated protein kinase-calcium/calmodulin regulator complex. *Sci Signal* 3(106):ra6.
24. Russo AA, Jeffrey PD, Pavletich NP (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* 3(8):696–700.
25. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS (2013) Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat Methods* 10(9):896–902.
26. van den Bedem H, Dhanik A, Latombe JC, Deacon AM (2009) Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr D Biol Crystallogr* 65(Pt 10):1107–1117.
27. Burnley BT, Pavel VA, Paul DA, Piet G (2012) Modelling dynamics in protein crystal structures by ensemble refinement. *Elife* 1:e00311.
28. Reynolds KA, McLaughlin RN, Ranganathan R (2011) Hot spots for allosteric regulation on protein surfaces. *Cell* 147(7):1564–1575.
29. Cerutti DS, Freddolino PL, Duke RE, Jr., Case DA (2010) Simulations of a protein crystal with a high resolution X-ray structure: Evaluation of force fields and water models. *J Phys Chem B* 114(40):12811–12824.
30. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3(7):1171–1179.
31. Read RJ (1997) [Model phases: Probabilities and bias. *Methods Enzymol* 277:110–128.
32. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
33. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760–763.
34. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.

Lang et al.