Automated electron-density sampling reveals widespread conformational polymorphism in proteins

P. Therese Lang,[1] Ho-Leung Ng,[1] James S. Fraser,[1] Jacob E. Corn,[1] Nathaniel Echols,[1] Mark Sales,[2] James M. Holton,[3,4] and Tom Alber[1]*

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California 97420-3220
[2]Department of Physics, University of California, Berkeley, California 97420
[3]Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158-2330
[4]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720

Received 29 March 2010; Revised 4 May 2010; Accepted 5 May 2010
DOI: 10.1002/pro.423
Published online 24 May 2010 proteinscience.org

Abstract: Although proteins populate large structural ensembles, X-ray diffraction data are traditionally interpreted using a single model. To search for evidence of alternate conformers, we developed a program, Ringer, which systematically samples electron density around the dihedral angles of protein side chains. In a diverse set of 402 structures, Ringer identified weak, nonrandom electron-density features that suggest of the presence of hidden, lowly populated conformations for >18% of uniquely modeled residues. Although these peaks occur at electron-density levels traditionally regarded as noise, statistically significant ($P < 10^{-5}$) enrichment of peaks at successive rotameric χ angles validates the assignment of these features as unmodeled conformations. Weak electron density corresponding to alternate rotamers also was detected in an accurate electron density map free of model bias. Ringer analysis of the high-resolution structures of free and peptide-bound calmodulin identified shifts in ensembles and connected the alternate conformations to ligand recognition. These results show that the signal in high-resolution electron density maps extends below the traditional 1 σ cutoff, and crystalline proteins are more polymorphic than current crystallographic models. Ringer provides an objective, systematic method to identify previously undiscovered alternate conformations that can mediate protein folding and function.

Keywords: X-ray crystallography; structural polymorphism; electron-density sampling; side chain ensembles; rotamers

Abbreviations: CaM, calmodulin; smMLCK, smooth muscle myosin light chain kinase; PDB, protein data bank.

P. Therese Lang and Ho-Leung Ng contributed equally to this work.

Ho-Leung Ng's current address is ConfometRx. 3070 Kenneth St. Santa Clara, California 95054.
Jacob E. Corn's current address is Department of Biochemistry, University of Washington, Seattle, Washington 98195.
Nathaniel Echols's current address is Physical Biosciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720.
Mark Sales's current address is Stormweaver Programming, 2930 Shattuck Ave Ste 305, Berkeley, California 94705.

Grant sponsor: NIH; Grant numbers: F32 GM069165, R01 GM48598; Grant sponsors: NSF; Canadian NSERC.

*Correspondence to: Tom Alber, Department of Molecular and Cell Biology, 374B Stanley Hall, Berkeley, CA 94720. E-mail: tom@ucxray.berkeley.edu

Introduction

Proteins populate structural ensembles that mediate folding,[1] ligand binding,[2] catalysis,[3,4] signaling,[5] and evolution.[6] Recent developments in NMR spectroscopy have provided new insights into the relationship between structure, function, and dynamics.[7] X-ray crystallography complements these techniques by providing detailed structural data on the nature of conformational polymorphism. The interpretation of protein dynamics from crystal structures is traditionally limited to direct observation of conformational differences in independent molecules within the same asymmetric unit[8] or independent crystal structures.[9,10] In addition, proteins in crystals exchange

hydrogens with solvent,[11] undergo reversible conformational changes,[12] and catalyze chemical reactions as complex as DNA polymerization.[13] These and many other examples indicate that many proteins in crystals can undergo small- and large-scale motions and populate diverse conformations.[14,15]

In contrast to this evidence for conformational polymorphism, ~95% of residues in the crystal structures in the Protein Data Bank (PDB) are modeled by a single conformation. Crystal structures are typically built by interpreting electron density maps at 1 standard deviation (σ) greater than the mean electron density of the unit cell. This 1 σ level is often considered to approximate the boundary between the signal from the molecule and the noise in the electron density map.[16] At high resolution, disorder can be modeled explicitly with the placement of alternate conformers into the electron density. To describe localized motions, atoms are assigned *B* factors (also known as atomic displacement parameters, Debye-Waller factors, or thermal factors) that represent a Gaussian decay of the electron density of the modeled atoms. To represent large-scale and collective motions, refinement may also include grouped anisotropic *B* factors.[17] Because they combine contributions from motions, static disorder and model errors, however, *B* factors can disguise discretely disordered alternate side-chain and main-chain conformations.[18–20]

To account for broader structural diversity, recently developed methods leverage unsupervised automatic refinement programs to build many models simultaneously into the electron density.[21,22] A computational method that explores large numbers of side-chain conformations to more fully explain electron density has also been described.[23] These methods, however, are subject to model bias and local errors in individual models.[24] In addition, despite the addition of numerous atoms to the structural model, these methods generally afford only limited improvements in the R_{cryst} and R_{free} values. Concerns that the reduced ratio of experimental observations to refined parameters can lead to models that are over-fitted when compared to individually refined structures, especially at moderate or low resolution, have historically limited the adoption of multicopy refinement methods.[25] Moreover, these broadened ensembles still do not capture the full range of structures accessible to proteins as evidenced by independent crystal structures or solution NMR experiments.

To determine if electron density maps contain signals for small populations of alternate side-chain conformations, we developed a computational method, Ringer. Ringer automatically samples electron density around the side-chain dihedral angles (χ) of a structural model and identifies peaks that correlate with structural features. Using Ringer, we uncovered evidence suggesting unmodeled, alternate conformations in over 18% of side chains in a diverse set of 402 high-resolution structures. Evidence for small populations of alternate rotamers was also detected in an unusually accurate, experimentally phased electron density map, indicating that the signals from these low-occupancy conformers do not arise from phasing artifacts or model bias. Ringer also identified changes in side-chain ensembles in high-resolution structures of calmodulin (CaM) in both an open state and bound to a peptide, suggesting these motions are related to folding and substrate recognition. These alternate structures provide a more dynamic picture of protein conformation and ligand binding. In general, these results indicate that systematic sampling of weak electron density features reveals conformational diversity in electron density maps that has been overlooked by traditional model building and refinement practices. These newly characterized conformations can be used to provide insights linking side-chain dynamics with protein function.

Results

The Ringer procedure

We developed the program Ringer to systematically sample electron density maps in rings around each side-chain dihedral angle [Fig. 1(A)]. Sampling positions for C, O, N, and S were defined using idealized bond lengths and angles that extend between the third and fourth atoms of the dihedral angle (e.g., between the β and the γ atoms for χ^1). Preliminary analysis of high-resolution structures revealed three general trends in the Ringer plots of electron density versus χ angle: residues with a single peak in the electron density [Fig. 1(B.I)], residues with multiple peaks that are already modeled as multiple conformations [Fig. 1(B.II)] and residues with multiple peaks that are modeled as a single conformation [Fig. 1(B.III)]. The second and third highest peaks of uniquely modeled residues generally occur at electron density levels below 1 σ. We sought to establish whether this range of electron density, often disregarded as noise, contained meaningful information about alternate conformations.

Numerous unmodeled alternate conformers in a large data set

To assess the generality and significance of the weak secondary peaks of uniquely modeled residues, we ran Ringer on 402 high-resolution ($\leq$1.5 Å) crystal structures from the Protein Data Bank. Omit electron-density maps[27] were analyzed to reduce the effects of model bias. We used Ringer to sample the dihedral angles of ~67,000 side chains that are unbranched at χ^1 (i.e., Ser, Gln, Asn, Glu, Asp, Arg, Lys, Met, Cys, Leu, Phe, Tyr, Trp, and His). Because

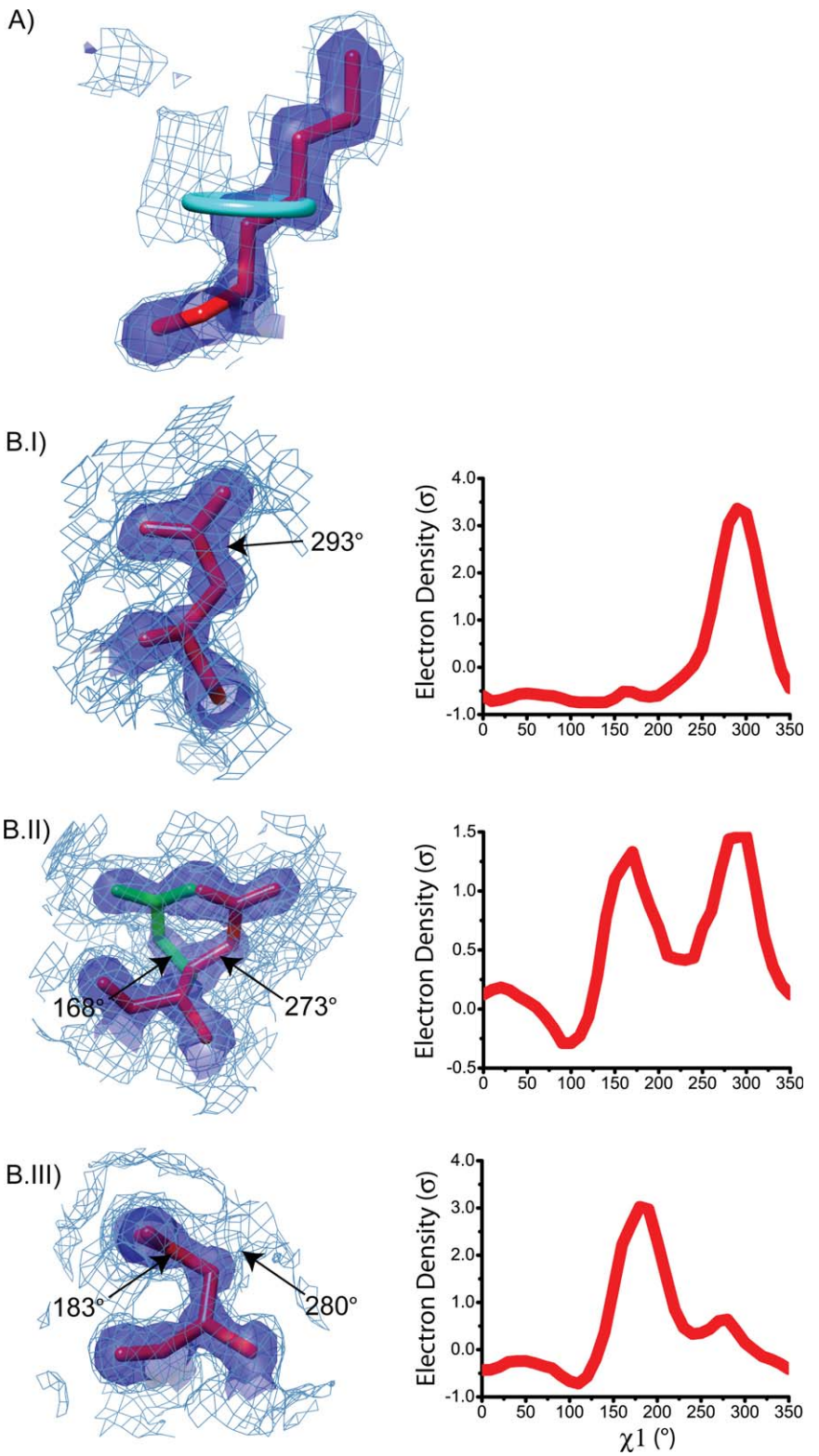

Figure 1. Ringer systematically samples electron density in real space. (A) Ringer computes electron density (1.0 σ, solid; 0.3 σ, mesh) around idealized dihedral angles (cyan ring) extending from the modeled side chain (red). Shown here is Lys86 from a putative tyrosine phosphatase from *Rhodopseudomonas palustris* (PDB ID 2HHG). (B) Ringer identifies several types of conformational polymorphism in the 1.0 Å-resolution electron density of oxy-myoglobin (PDB ID 1A6M[26]). Left: the model is superimposed on the electron density map displayed at the standard cutoff (1.0 σ, solid) and the mean electron density (0.0 σ, mesh). Right: plots of electron density (σ) as a function of χ^1 angle. (I) Asp27: a single peak in the electron density distribution around a uniquely modeled residue. (II) Glu4: multiple peaks from multiple conformations (red, major; green, minor) in the model. (III) Asn132: multiple peaks in the electron density interpreted with a unique conformation in the model.

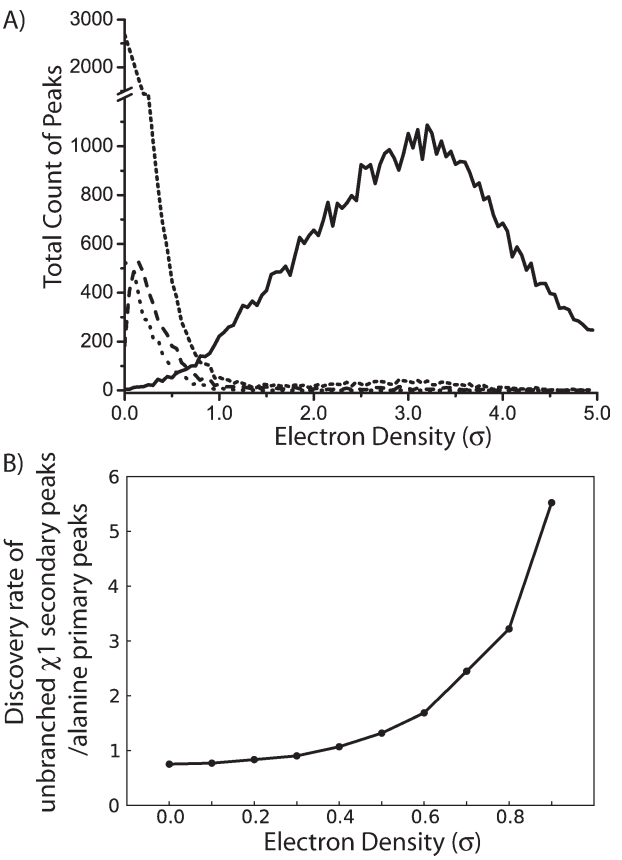

Figure 2. Analysis of a test set of 402 high-resolution structures suggests a lower threshold for defining side-chain electron density. (A) Distribution of Ringer peak heights at the C—C bond length for primary (solid), secondary (short dash), and tertiary peaks (dash) for residues that are unbranched at χ^1. In comparison to the alanine primary electron density peak heights (dot), also sampled at C—C bond length, the secondary and tertiary peaks of longer side chains are enriched above 0.3 σ. (B) Lower electron density threshold that enriches for heavy-atom over hydrogen contributions at the C—C bond length. Discovery rate, calculated as the ratio of the number χ^1 secondary electron density peaks (normalized by the total number of χ^1 side chains) to alanine primary electron density peaks (normalized by the total number of alanines), is plotted versus the lower electron density cutoff (σ).

Ringer identifies alternate conformations based on electron density peaks other than the modeled the primary peak, we excluded from this analysis β-branched residues (Ile, Val, and Thr) that show two intense χ^1 Ringer peaks. We also excluded residues that are currently modeled in multiple conformations (5.1% of the side chains). Of the uniquely modeled residues analyzed, 31.6% had more than one peak at greater than the mean electron density for each map (0 σ). These secondary and tertiary peaks of uniquely modeled residues generally occur at electron density levels below 1 σ [Fig. 2(A)].

Although Ringer samples the electron density at distances corresponding to the heavy-atom (nonhydrogen) bond lengths (e.g., 1.53 Å for a sp^3 C—C bond), the weak electron-density features may arise from noise, alternate conformers and/or hydrogen atoms (with a C—H bond distance of ~1 Å). To distinguish these possibilities, we established a characteristic level of electron density that maximized detection of heavy atoms and minimized the signal from hydrogen atoms. To estimate the signals from hydrogen atoms and noise, we analyzed Ringer peaks of the 8565 alanines in the test set. The peaks from alanine residues generally occurred at low σ values when sampled by Ringer at ideal carbon-carbon bond lengths and angles [Fig. 2(A)]. We reasoned that heavy atom contributions predominate when the discovery rate of unmodeled secondary peaks for unbranched χ^1 residues exceeds the discovery rate of primary alanine peaks. The ratio of alanine peaks to χ^1 secondary peaks is consistent from 0.1 to 0.2 σ. However, at electron density levels ≥0.3 σ, the frequency of χ^1 peaks is enriched over alanine [Fig. 2(B)] and continues to rise as the electron density level increases. These results suggest that secondary Ringer peaks ≥0.3 σ are likely the result of the heavier atoms of alternate side-chain conformations.

Applying this new threshold to the uniquely built side chains in the test set of 402 structures, we found 11.8% of residues display χ^1 secondary peaks ≥0.3 σ, suggesting the widespread presence of small populations of alternate conformers. Sampling around χ^2, χ^3, and χ^4 reveals secondary peaks for 12.5, 18.1, and 26.0% of side chains, respectively (Table I). The increasing frequency of secondary peaks at successive χ angles suggests that the peaks correspond to alternate conformations rather than hydrogen atoms, as longer side chains are less constrained away from the backbone. In total, Ringer discovers at least one secondary peak ≥0.3 σ for 18.5% (12,476/67,487) of all uniquely modeled side chains analyzed in the test set.

If these peaks represent minor conformers and not noise, then the distribution of peaks in dihedral space should be nonrandom and biased towards low-energy rotational isomers (rotamers).[28–30] Indeed, the distribution of secondary χ^1 peaks ≥0.3 σ is strongly enriched at the rotameric angles of 60, 180, and 300° [Fig. 3(A)]. The tertiary peaks show a

Table I. *The Fraction of Uniquely Modeled Side Chains with Secondary Electron-Density Peaks ≥0.3* σ

Chi angle	Total no of side chains	2° peaks (%)	Rotameric 2° peaks (%)
χ^1	58,608	11.8	9.7
χ^2	26,712	12.5	9.6
χ^3	11,554	18.1	13.0
χ^4	10,051	26.0	17.3

Rotameric peaks are within 30° of χ values of 60, 180, or 300°.

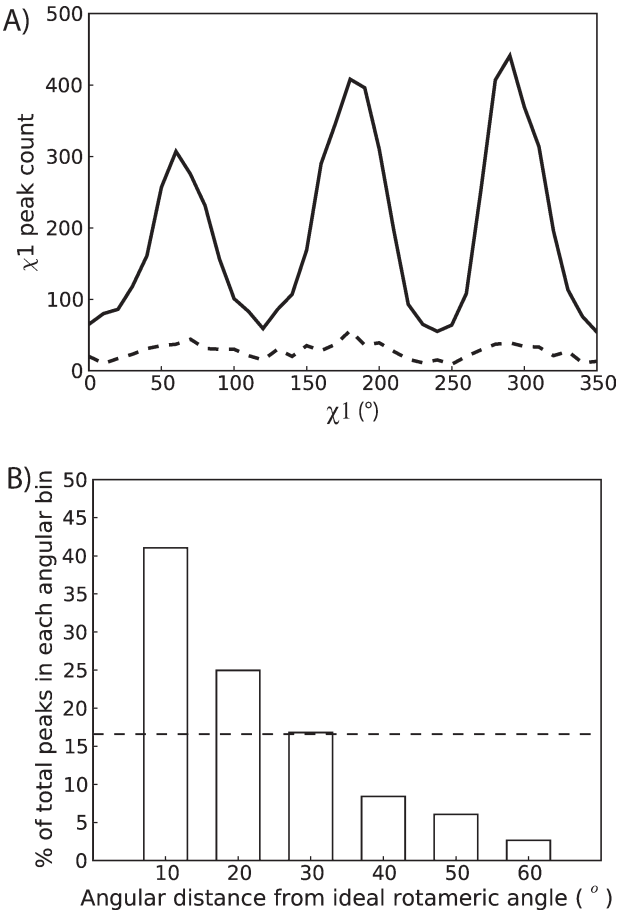

Figure 3. Weak electron-density features are enriched in rotameric positions. (A) Distribution of secondary peaks (solid line) $\geq$0.3 σ versus χ^1 angle shows a trimodal distribution strongly enriched for preferred rotameric positions. The distribution of tertiary peaks (dotted line) $\geq$0.3σ shows a similar tri-modal distribution. In contrast, the distribution of quaternary peaks, which cannot physically correspond to chemical features of the side chains, is random (data not shown). (B) Percent of the total secondary peaks within the indicated angular difference from ideal rotameric values. Peaks are enriched over a random distribution (dashed line) up to 30° from the rotameric values.

similar distribution but at a lower discovery rate [Fig. 3(A)]. To assess the significance of this distribution, we compared the angular difference from of each secondary peak to the nearest ideal rotameric angle to a random distribution [Fig. 3(B)]. This analysis showed that secondary χ^1 Ringer peaks were enriched at χ angles within 30° of ideal rotameric angles. Similar results were found for peaks of the remaining three side-chain dihedral angles ($\chi^2 - \chi^4$; data not shown). Thus, both the comparison to alanine residues and the rotameric angular distributions suggest that Ringer peaks $\geq$0.3 σ may reflect contributions from alternate conformations of side chains. These results suggest that these previously undetected, low-occupancy conformations are widespread in protein crystals.

Ringer peaks are correlated at χ1 and χ2

The conclusion that the alternate χ^1 Ringer peaks represent alternate rotamers predicts that sampling around successive χ angles of an unmodeled conformer should show peaks for the remainder of the side chain. Accordingly, peaks at subsequent dihedral angles should also be enriched at rotameric angles. To test these predictions, we analyzed ~6,500 nonaromatic side chains that are unbranched at χ^1 and have a heavy atom at the δ position (i.e., Arg, Asn, Asp, Gln, Glu, Leu, Lys, and Met). We placed a simulated Cγ atom at the secondary χ^1 peak and used this simulated Cγ atom to define a new χ^2 sampling radius [Fig. 4(A)]. Using this strategy, 83% of secondary χ^1 peaks $\geq$0.3 σ had a peak at the derived χ^2 angle and bond length. The $\chi^1 - \chi^2$ correlation forms a "checkerboard" pattern consistent with rotameric distributions of bonded side-chain heavy atoms [Fig. 2(D), *P*-value $<10^{-5}$]. A similar test for alanine peaks revealed no significant association between χ^1 primary peaks and χ^2 peaks extending from simulated Cγ atoms (data not shown). The enrichment of weak electron density features at these specific angular positions suggested that the Ringer plots detect populations of unmodeled alternate conformations.

Ringer peaks reflect signal rather than Fourier errors or model bias

To confirm that secondary peaks detected by Ringer derive from alternate conformers and are not the consequence modeling and refinement procedures, we analyzed an unusually accurate experimental electron density map of the 33-residue, designed coiled coil, RH4B (PDB ID 2O6N[31]). The electron density map was calculated at 1.1 Å resolution using multiple wavelength anomalous diffraction (MAD) data collected from a Yb3+ derivative with an average phasing power >7. Because this map is accurately phased, it can be used to model the structure directly, avoiding model bias. In addition to the five alternate conformations already modeled into this map, Ringer found rotameric peaks $\geq$0.3 σ for an additional 11 residues (Fig. 5). This systematic identification of weak electron density features in an accurate, experimentally phased map supports the conclusion that the peaks are due to signal from low populations of alternate conformations rather than errors from model bias.

To explore the influence of truncating the Fourier transform, we calculated an electron density map ($F_{calc}\phi_{calc}$ truncated at 1.1 Å) directly from the RH4B model coordinates. The Ringer plots showed no evidence for alternate conformations in this calculated map (data not shown). These results indicated that the weak density signaled alternate side-chain conformations rather than truncation effects.

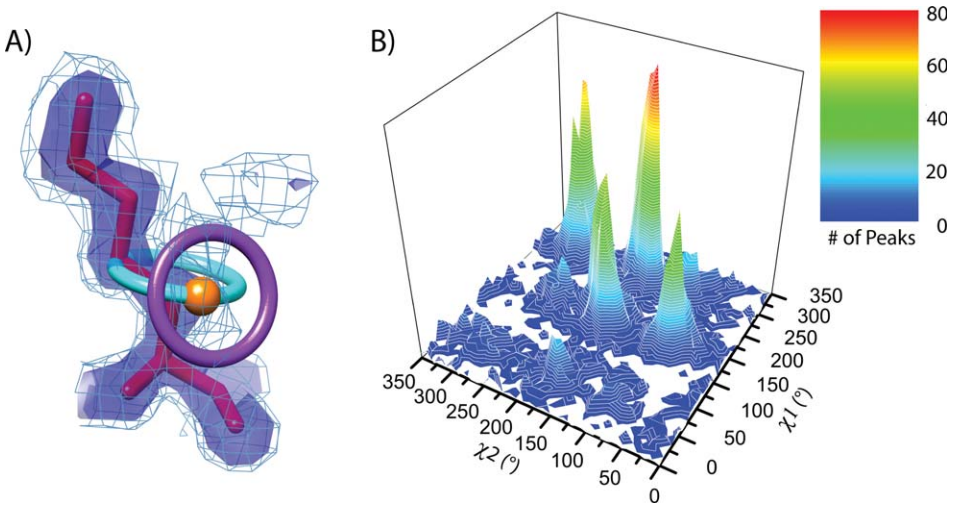

Figure 4. Ringer detects peaks at correlated χ^1 and χ^2 angles. (A) Peaks in electron density (1.0 σ, solid; 0.3 σ, mesh) are identified by sampling χ^1 (cyan ring) and then sampling χ^2 (purple ring) at idealized bond lengths from the χ^1 secondary peak (orange sphere). (B) A histogram of secondary χ^1 peaks and primary χ^2 peaks built from the unmodeled χ^1 peaks. Cross-peaks are significantly enriched in rotameric positions (*P*-value $< 10^{-5}$).

Furthermore, the Ringer peaks are not the consequence of refinement procedures but instead correspond to alternate conformations present in the crystal lattice.

Discovery rate of alternate conformers depends on resolution

To explore the resolution-dependence of the Ringer peak discovery rate, we designed a test set of structures from 0.6 to 3.0 Å resolution. The discovery rate of unmodeled electron density features remains high throughout this data set [Fig. 6(A)]. However, as resolution decreases, the percentage of identified peaks that occur in rotameric positions also decreases [Fig. 6(B)]. These results suggest that Ringer can be applied to identify unmodeled electron density features in a wide range of resolutions but, as expected, the correlation of these features with alternate conformations gradually decreases beyond ~2.0 Å.

To further evaluate the sensitivity of Ringer, we analyzed a test set of perfectly phased, noise-free, simulated electron density maps in which we set the occupancy, *B* factor and resolution. In this set, a side chain in the structure of ubiquitin (PDB ID 1UBQ) was computationally mutated to an ensemble containing primary and secondary conformations. The occupancy for the conformations was explored in 10% increments between 10 and 50% over a range of resolution from 0.8 to 2.8 Å. The *B* factors of the altered residue were set to 15 or 30 Å^2. This procedure was repeated for all side chains sampled by Ringer. For both *B* factors tested, Ringer identified peaks ≥0.3 σ for all residues with 10% occupancy at 2.0 Å or higher resolution (data not shown). At 2.8 Å resolution, Ringer detected all side chains at 50% occupancy. These results coincide with the expectation that higher resolution data are needed to detect increasingly smaller populations of conformers.

Low-occupancy conformations provide functional insights: Calmodulin binding

To test whether the low-occupancy conformations in high-resolution structures provide biological insights in addition to structural information, we analyzed structures of the calcium sensor, calmodulin (CaM). Because the side-chain conformational dynamics of CaM underlie ligand binding through target-specific deformation of the binding pockets[32] and entropic contributions,[2] the exhaustive annotation of conformational diversity is essential to understand recognition specificity. We first analyzed the high-quality, 1.0 Å-resolution structure of the unbound form, which has one of the highest percentages (24.1%) of multiple side-chain conformations of all crystallographic models[32] (R/R-free 0.134/0.163; PDB ID 1EXR[32]). Excluding these previously modeled conformers, Ringer identified evidence for unmodeled alternate rotamers of 22 additional residues. These minor conformations are distributed throughout the protein and enriched in some of the most compactly folded portions of the structure [Fig. 7(A)]. These results indicate that polymorphism is not confined to the surface of CaM but also is observed in areas that are critical for target peptide binding.

To investigate the functional relevance of these features, we compared the Ringer analyses for the unbound form of CaM and the high-resolution structure of CaM bound to smooth muscle myosin light chain kinase peptide (smMLCK; 1.08 Å; R/R-free 0.144/0.165; PDB ID 2O5G). This peptide-bound structure has a similar resolution and noise level as

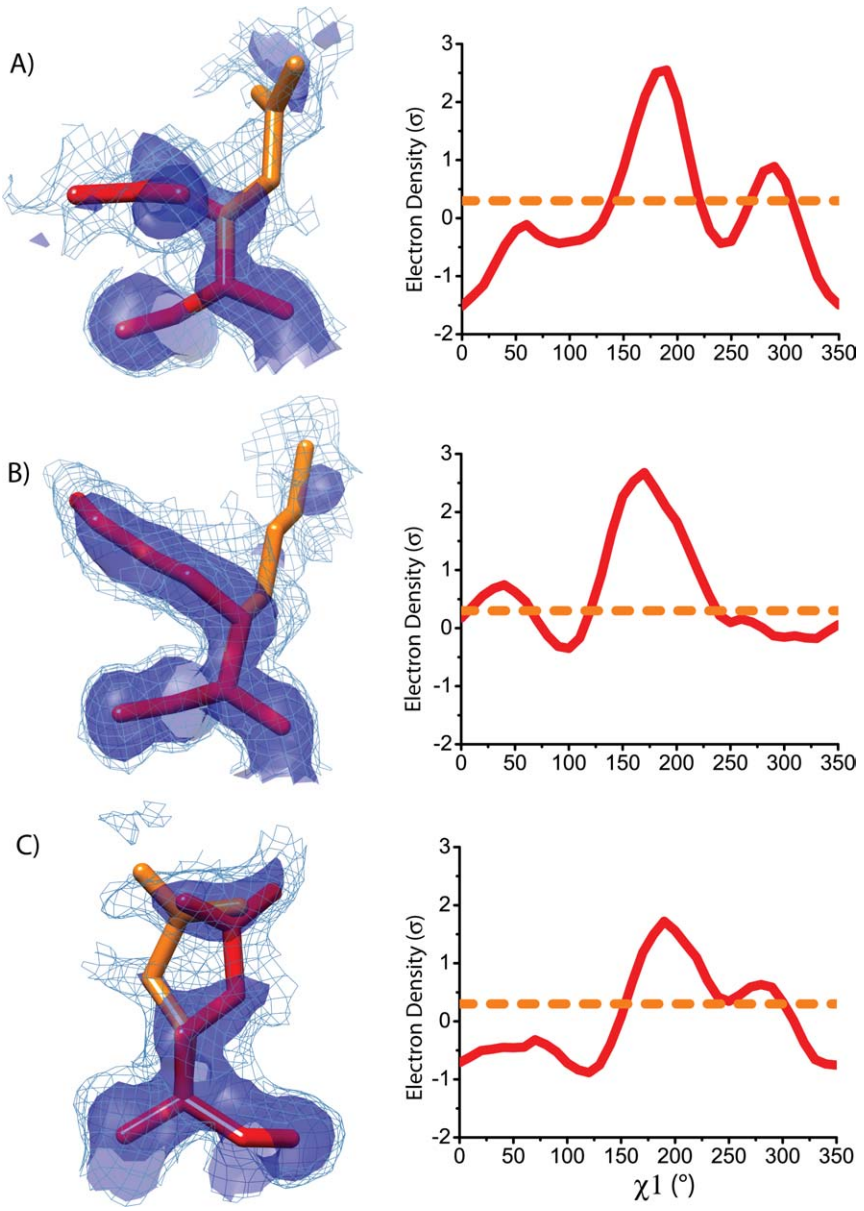

Figure 5. Ringer detects evidence for minor unmodeled conformations in an electron density map calculated with accurate experimental phases. Ringer plots of electron density (σ) versus χ^1 angle for three representative residues in the RH4B coiled coil (PDB ID 2O6N[31]). The orange line indicates 0.3 σ. The experimental, MAD-phased electron density at 1 σ (solid) and 0.3 σ (mesh) is shown with the modeled conformations (red) for residues (A) Gln5, (B) Lys7, and (C) Glu20. Structures of the minor rotamers (orange) have been added to guide the eye and are not the result of refinement.

the free CaM structure, making the electron density more amenable to direct comparisons. In the smMLCK peptide complex structure, only 8% of residues have alternate conformations included in the model. Excluding these residues, Ringer found evidence for unmodeled alternate conformations of an additional 34 residues. These results indicate that, even in the closed, bound form, CaM side chains access diverse structural ensembles.

When comparing the free and peptide-bound CaM structures, many residues [i.e., Arg126; Fig. 7(B.I)] showed only a single Ringer peak ≥0.3 σ indicative of the transition between unique conformations in the free and peptide-bound forms. Other residues showed evidence for minor alternate conformations in only one structure. For example, the modeled χ^1 angle of Ser38 changes from 80° in free CaM to 295° in the smMLCK complex. This shift would traditionally suggest that Ser38 undergoes local repacking upon smMLCK peptide binding. Ringer identified the primary peak at 80° in the unbound CaM structure, but also identified a secondary peak (0.56 σ) at 290°. While the change in population of the side chain conformations cannot be determined without fully refining the structures, the pattern identified in the Ringer plots provides

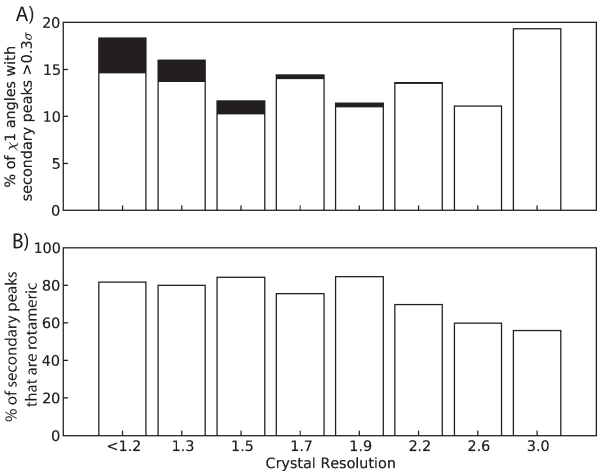

Figure 6. Resolution dependence of Ringer peak detection at the 0.3 σ threshold. (A) Percent of unique residues unbranched at χ^1 with secondary peaks ≥0.3 σ over a range of resolutions (white). The percent of residues with modeled alternate conformations is displayed in black. (B) The percentage of secondary Ringer peaks within 30° of rotameric angles over a range of resolutions.

evidence that smMLCK peptide binding selects for a minor conformation of the unbound form rather than inducing local repacking [Fig. 7(B.II)].

Ringer also detects shifts in ensembles of minor conformations, even when the primary conformation stays constant. In both the free and smMLCKp-bound structures, for example, Asp22 is built with a χ^1 angle of ~70°. However, Ringer also detects evidence for side-chain structural polymorphism at $\chi^1 = 210°$ in the unbound form, which disappears upon peptide binding. Such changes have been proposed to dominate the overall entropy of peptide binding and underlie the compensation of entropic and enthalpic contributions to peptide recognition.[2] In general, the unmodeled alternate conformations identified by Ringer help explain the observed conformational plasticity of the CaM binding site and provide testable hypotheses about where residual side-chain entropy is distributed in the protein.

Discussion

In contrast to traditional crystallographic models, which generally represent proteins in "unique" conformations, direct sampling of the electron density using Ringer reveals evidence for numerous small populations of alternate conformers. When applied after refinement is considered complete, Ringer discovers polymorphism at over 3.5 times the frequency that is currently modeled in the PDB. Multiple conformers are found for >18% of unbranched residues in a test set of 402 high-resolution structures, in addition to the 5.1% that are already modeled. It should be emphasized that the 0.3 σ cutoff applied here to identify alternate conformations is an operational threshold rather than an absolute lower electron density limit. Among the reasons for this characterization are that the noise level in electron density maps varies with phase accuracy and that 0.3 σ represents a different absolute electron density in different structures. Additionally, backbone shifts can accommodate additional alternate conformations that fall outside the Ringer sampling radius. As a result, the degree of polymorphism detected by Ringer above 0.3 σ may represent a lower estimate of the alternate structures that have escaped detection in high-resolution X-ray crystal structures.

The alternate conformers revealed by Ringer have remained hidden because they occur at electron density levels (0.3–1.0 σ) that are generally ignored during model building. Moreover, complex combinations of overlapping multiple conformations may escape visual recognition. Because new unsupervised modeling programs generally fit to higher values of electron density, even these methods may miss conformations detected by Ringer. From a statistical perspective, adding multiple conformers produces only small reductions in R factors,[22,33] leading to concerns that the increased ratio of parameters to observations may not be justified and raising the question of how to validate each specific polymorphic model. However, the statistical enrichment of rotameric nonhydrogen Ringer peaks, the $\chi^1 - \chi^2$ correlation of minor peaks around long side chains, and the discovery of rotameric peaks in a wide range of maps provide a consistent interpretation of these weak features in terms of alternate rotamers missed by traditional, subjective model-building practices. Thus, unlike automated refinement methods, Ringer provides direct evidence that there is additional signal in high-resolution electron density maps not being captured by current modeling practices.

Ringer enables automated, systematic identification of regions of conformational polymorphism. In contrast to manual methods, the program rapidly analyzes every residue in a model and applies consistent, objective criteria throughout the map to discover potentially polymorphic side chains. Peaks detected by Ringer currently require manual inspection and building of alternate conformers. Corroboration of individual alternate conformers should be obtained from F_o-F_c difference density,[4] coupled backbone shifts,[19] and connected electron density corresponding to the entire side chain (Fig. 5). Nonrotameric peaks, which can arise from coupled main-chain shifts[19] or high-energy conformations, can be difficult to build by visual inspection. We suggest using Ringer iteratively with model building and occupancy refinement. Side-chain conformations added to the model to fit Ringer peaks should be validated by small improvements in map quality and interpretability, real space correlation coefficients between F_c and F_{obs}, the consistency of B values or

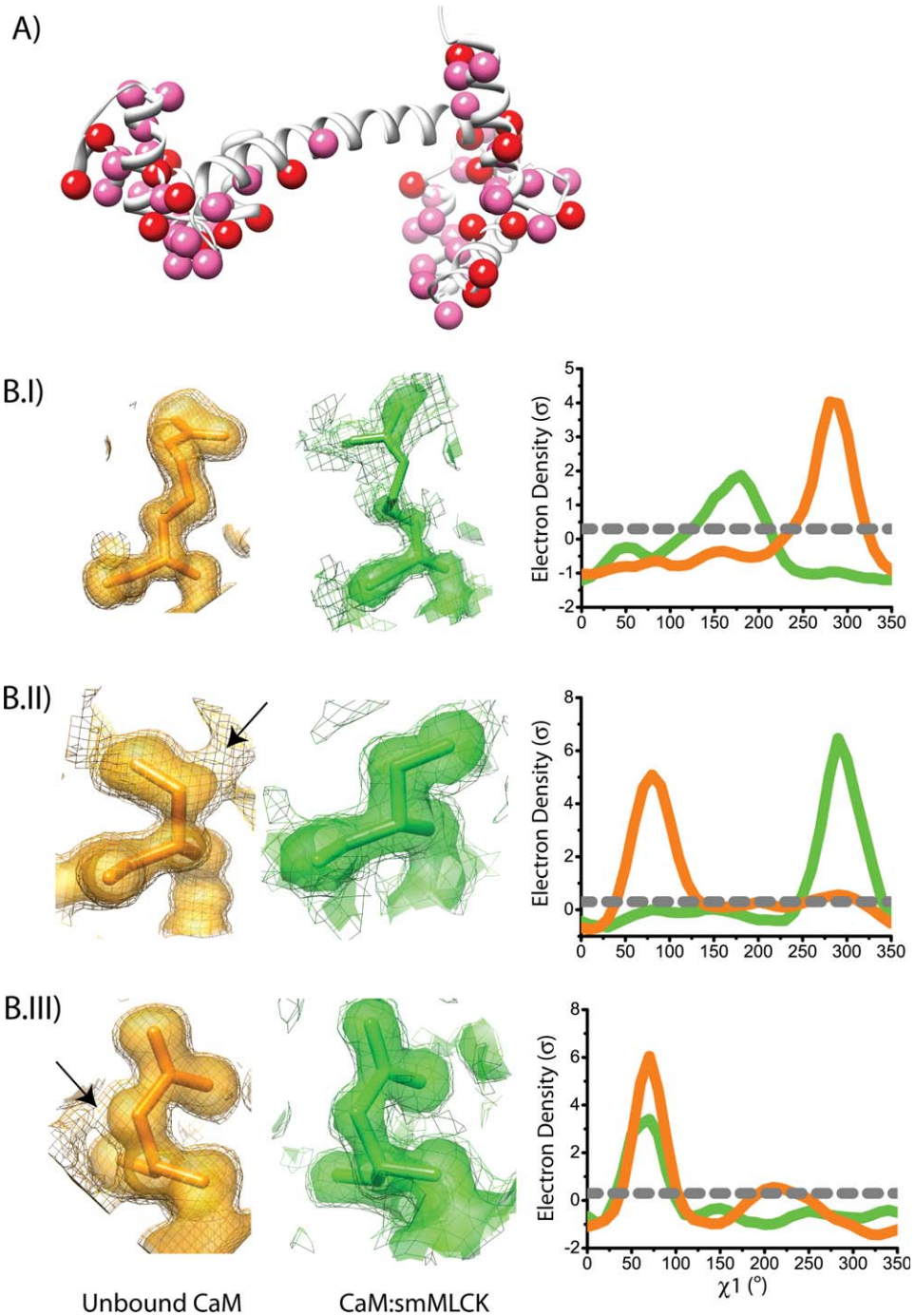

Figure 7. Ringer finds evidence for functionally relevant, unmodeled minor conformers in CaM. (A) Distribution in unbound CaM of unmodeled (red spheres) and modeled alternate conformations (pink spheres). (B) Electron density features of unbound CaM (orange; 1.0 Å resolution) and CaM complexed with the smMLCK peptide (green; 1.08 Å resolution). Modeled conformations are shown superimposed on the electron density at 1 σ (solid) and 0.3 σ (mesh). Arrows indicate areas of interest in electron density. Ringer plots of electron density (σ) versus χ^1 angle for three representative residues. Gray dashed line indicates 0.3 σ. (I) Arg126: single peaks ≥0.3 σ in the electron density distribution around each uniquely modeled residue. (II) Ser38: the unbound structure has two electron-density peaks ≥0.3 σ, and the minor peak is strongly enriched in the peptide complex. Selection of the secondary peak supports a population shift upon binding. (III) Asp22: the unbound structure shows two peaks ≥0.3 σ. The secondary peak is absent in the smMLCK peptide complex, suggesting rigidification of this residue accompanies binding.

R/R_{free} values. New methods of displaying weak electron density and refining occupancies may be necessary for both manual and automatic building of accurate ensembles. The high prevalence of side-chain polymorphism also may require new methods of displaying protein models to adequately visualize

the alternate structures in intuitive and informative ways.

More than a method for enhancing crystallographic refinement, Ringer is best used as a tool for systematically detecting low-occupancy structural features. The hidden conformational substates identified using Ringer provide clues to the functional roles of protein structural polymorphism and to assess the response of protein side-chain distributions to perturbations including ligand binding, temperature changes and mutations. In CaM, for example, Ringer identifies side chains that undergo conformational population inversions and side-chain rigidification upon peptide binding, linking the structure to dynamic properties. Similarly, in human proline isomerase, Ringer was used to define the nature of a coupled conformational switch in the free-enzyme that defines motions that occur during turnover.[4] In both cases, the alternate conformations identified by Ringer provided structural insights not available from any other experimental technique.

The energetic consequences of shifts in the populations of alternate conformations depend critically on coupling. Independent conformers, for example, would contribute significantly to the residual entropy of the folded state.[34,35] In contrast, motional contributions to turnover, allosteric regulation and signal transmission in proteins are contingent on the coupling of side-chain conformations over significant distances.[36] Ringer does not by itself define the interdependence of alternate conformations. Additional criteria such as steric compatibility, similar occupancies across a network or information from other methods are required to assess the coupling of substates. Nonetheless, by providing evidence for the structures of alternate conformations, the residue-specific identification of alternate rotamers derived from Ringer analysis of electron-density maps has the potential to illuminate the atomic motions that govern macromolecular functions.

Materials and Methods

Data sets

Unless otherwise stated, PDB files and structure factors were downloaded from the PDB web-site (http://www.pdb.org) based on (a) the availability of the experimental X-ray amplitudes, (b) resolution $\leq$1.5 Å, (c) R factor $\leq$0.22, (d) lack of homology to other proteins in the test set, and (e) absence of nucleic acids. We were able to convert reflections and calculate omit maps[27] for a total of 402 structures. For test structures >1.5 Å resolution, PDB files and structure factors were filtered to remove metalloproteins. The remaining structures were clustered by homology and selected for uniqueness. Representatives were chosen randomly to give an average of 44 structures for each resolution bin.

To calculate electron density maps, the downloaded structure factors were converted into MTZ format. If structure amplitudes (F_o) were deposited, they were used directly. Otherwise, the structure intensities were converted into structure amplitudes. The deposited structures were used to calculate the structure factors (F_c) and phases to generate a $2F_o$-F_c Bhat omit map with a grid spacing of one-third the resolution or finer.[27] Calculations were performed using the CCP4 suite of programs.[37]

Ringer

Ringer analyzes crystallographic electron density using atomic coordinate files in standard PDB format. The electron density maps can be computed from structure factors deposited in the PDB (as in this study), downloaded from the Uppsala Electron Density Server,[38] or calculated in the course of crystallographic refinement. Ringer reads electron density maps in CCP4, X-PLOR, and CNS file formats. The maps are expanded to ensure they cover the entire model. The σ values in the map are scaled to ensure a mean electron density of 0 σ and a standard deviation of 1 σ. This scaling includes all voxels in the electron density, and it emulates the standard scaling in the commonly used modeling program, Coot.[39]

Each χ angle is defined based on the modeled side chains. χ^1 angles are built starting at the backbone nitrogen. For residues with alternate conformations included in the model, the user has the option either to sample using the higher-occupancy conformation or to exclude the side chain from the analysis (skip_multi_conf). The bond length and angle used for moving the terminal atom (atom_sample_type) are controlled by one of three options defined by the user to be constant, experimental, or dynamic. For "constant" sampling, χ angles are sampled with a bond length of 1.53 Å and angle of 111.1°. For "experimental" sampling, the heavy atom used to sample the map is placed at the bond length and angle found in the model. For "dynamic" sampling, the terminal heavy atom is placed at the bond length and angle defined in the AMBER parm99 parameter set.[40] With all sampling options, dihedral angles that terminate in hydrogens are sampled with a standard sp^3 carbon bond length of 1.53 Å and angle of 111.1°. We used "dynamic" sampling for our studies except for Figure 4(B), in which the χ^2 angles were sampled using "constant" sampling.

The dihedral angles are systematically sampled by absolute torsion angle in user-defined increments (chi_sample_degree). The electron density at each point is extrapolated from the map electron density in Cartesian space by trilinear interpolation. The electron density (in units of σ) is plotted versus the sampled χ angles for peak identification. For this study, we used 10° increments as this was the

coarsest sampling that still captured the critical features of the electron density plots (data not shown). In Figure 4(B), we sampled in 1° increments to ensure most accurate placement of subsequent sampling rings.

Peaks are identified as the maxima in the plots of σ versus χ angle above the user-defined lower cutoff (lower_sigma_cutoff). Unless otherwise stated, we used 0.3 σ. The peaks in order of height are called primary, secondary, tertiary, and so forth. The user has the option to write out the results to generate the Ringer plots for specific residues (plot_residue and write_sigma_plot) and the list of the identified peaks (write_peak_list).

The identified peaks are divided into χ angle-based categories: χ^1 = Ser, Gln, Asn, Glu, Asp, Arg, Lys, Met, Cys, Leu, Phe, Tyr, Trp, and His; χ^2 = Gln, Glu, Arg, Lys, Met, and Ile; χ^3 = Lys, Arg, and Met; χ^4 = Lys and Arg. Because Ringer identifies alternate conformations based on density peaks beyond the primary peak, only unbranched torsion angles are chosen for each χ angle category. Arg, Asn, Asp, Gln, Glu, Leu, Lys, and Met residues were used for χ^1/χ^2 "checkerboard" correlation analysis. Unless otherwise stated, rotameric conformations have χ angles of 60°, 180°, or 300° ± 30°.

Ringer is implemented in Python and depends on the freely available software, Chimera.[41] Data were generated using Chimera version 1.3. The code base is freely available for download for academic users (http://ucxray.berkeley.edu/ringer.htm).

Acknowledgments

The authors thank G. A. Petsko, J. Kuriyan, A. J. Wand, P. Barth, and D. Kern for stimulating discussions.

References

1. Vendruscolo M (2007) Determination of conformationally heterogeneous states of proteins. Curr Opin Struct Biol 17:15–20.
2. Frederick KK, Marlow MS, Valentine KG, Wand AJ (2007) Conformational entropy in molecular recognition by proteins. Nature 448:325–U323.
3. Boehr DD, McElheny D, Dyson HJ, Wright PE (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. Science 313:1638–1642.
4. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T (2009) Hidden alternative structures of proline isomerase essential for catalysis. Nature 462:669–673.
5. Li PL, Martins IRS, Amarasinghe GK, Rosen MK (2008) Internal dynamics control activation and activity of the autoinhibited Vav DH domain. Nat Struct Mol Biol 15:613–618.
6. Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. Science 324:203–207.
7. Mittermaier A, Kay LE (2006) Review—new tools provide new insights in NMR studies of protein dynamics. Science 312:224–228.
8. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hubner CG, Kern D (2007) Intrinsic motions along an enzymatic reaction trajectory. Nature 450: 838–U813.
9. Zhang XJ, Baase WA, Shoichet BK, Wilson KP, Matthews BW (1995) Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. Protein Eng 8:1017–1022.
10. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M (2006) Relation between native ensembles and experimental structures of proteins. Proc Natl Acad Sci USA 103:10901–10906.
11. Niimura N, Arai S, Kurihara K, Chatake T, Tanaka I, Bau R (2006) Recent results on hydrogen and hydration in biology studied by neutron macromolecular crystallography. Cell Mol Life Sci 63:285–300.
12. Rasmussen BF, Stock AM, Ringe D, Petsko GA (1992) Crystalline ribonuclease A loses function below the dynamical transition at 220 K. Nature 357:423–424.
13. Johnson SJ, Beese LS (2004) Structures of mismatch replication errors observed in a DNA polymerase. Cell 116:803–816.
14. Frauenfelder H, Petsko GA, Tsernoglou D (1979) Temperature-dependent X-ray-diffraction as a probe of protein structural dynamics. Nature 280:558–563.
15. Wang JW, Dauter M, Alkire R, Joachimiak A, Dauter Z (2007) Triclinic lysozyme at 0.65 angstrom resolution. Acta Cryst D63s:1254–1268.
16. Jensen LH (1997) Refinement and reliability of macromolecular models based on X-ray diffraction data. Methods in enzymology: Macromolecular crystallography, Part B, Vol. 277. San Diego: Academic Press, pp 353–366.
17. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods in enzymology: Macromolecular crystallography, Part D, Vol. 374. San Diego: Academic Press, pp 300–321.
18. Kuriyan J, Petsko GA, Levy RM, Karplus M (1986) Effect of anisotropy and anharmonicity on protein crystallographic refinement—an evaluation by molecular-dynamics. J Mol Biol 190:227–254.
19. Davis IW, Arendall WB, Richardson DC, Richardson JS (2006) The backrub motion: how protein backbone shrugs when a sidechain dances. Structure 14:265–274.
20. Shapovalov MV, Dunbrack RL (2007) Statistical and conformational analysis of the electron density of protein side chains. Proteins 66:279–303.
21. Furnham N, Blundell TL, DePristo MA, Terwilliger TC (2006) Is one solution good enough?Nat Struct Mol Biol 13:184–185.
22. Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN (2007) Ensemble refinement of protein crystal structures: validation and application. Structure 15:1040–1052.
23. van den Bedem H, Dhanik A, Latombe JC, Deacon AM (2009) Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. Acta Cryst D65: 1107–1117.
24. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Adams PD, Moriarty NW, Zwart P, Read RJ, Turk D, Hung LW (2007) Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models. Acta Cryst D63:597–610.
25. Chen Z, Chapman MS (2001) Conformational disorder of proteins assessed by real-space molecular dynamics refinement. Biophys J 80:1466–1472.
26. Vojtechovsky J, Chu K, Berendzen J, Sweet RM, Schlichting I (1999) Crystal structures of myoglobin-

ligand complexes at near-atomic resolution. Biophys J 77:2153–2174.
27. Vellieux FMD, Dijkstra BW (1997) Computation of Bhat's OMIT maps with different coefficients. J Appl Cryst 30:396–399.
28. Janin J, Wodak S, Levitt M, Maigret B (1978) Conformation of amino-acid side-chains in proteins. J Mol Biol 125:357–386.
29. Ponder JW, Richards FM (1987) Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 193:775–791.
30. Lovell SC, Word JM, Richardson JS, Richardson DC (2000) The penultimate rotamer library. Proteins 40: 389–408.
31. Sales M, Plecs JJ, Holton JM, Alber T (2007) Structure of a designed, right-handed coiled-coil tetramer containing all biological amino acids. Protein Sci 16: 2224–2232.
32. Wilson MA, Brunger AT (2000) The 1.0 angstrom crystal structure of Ca2+-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol 301:1237–1256.
33. Kuriyan J, Osapay K, Burley SK, Brunger AT, Hendrickson WA, Karplus M (1991) Exploration of disorder in protein structures by X-ray restrained molecular-dynamics. Proteins 10:340–358.
34. Sturtevant JM (1977) Heat-capacity and entropy changes in processes involving proteins. Proc Natl Acad Sci USA 74:2236–2240.
35. DuBay KH, Geissler PL (2009) Calculation of proteins' total side-chain torsional entropy and its influence on protein-ligand interactions. J Mol Biol 391:484–497.
36. Hardy JA, Wells JA (2004) Searching for new allosteric sites in enzymes. Curr Opin Struct Biol 14:706–715.
37. Bailey S (1994) The Ccp4 suite—programs for protein crystallography. Acta Cryst D50:760–763.
38. Kleywegt GJ, Harris MR, Zou JY, Taylor TC, Wahlby A, Jones TA (2004) The Uppsala electron-density server. Acta Cryst D60:2240–2249.
39. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Cryst D60:2126–2132.
40. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197.
41. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.