## Title
E pluribus unum, no more: from one crystal, many conformations



## Journal
Current Opinion in Structural Biology, 28(1)

## ISSN
0959-440X

## Authors
Woldeyes, Rahel A
Sivak, David A
Fraser, James S

## Publication Date
2014-10-01

## DOI
10.1016/j.sbi.2014.07.005

Peer reviewed

Published in final edited form as:
*Curr Opin Struct Biol*. 2014 October ; 0: 56–62. doi:10.1016/j.sbi.2014.07.005.

# *E pluribus unum,* no more: From one crystal, many conformations

**Rahel A Woldeyes**[1], **David A Sivak**[2], and **James S Fraser**[3]

[1]Chemistry and Chemical Biology Graduate Program, University of California San Francisco, San Francisco, CA 94158

[2]Center for Systems and Synthetic Biology, University of California San Francisco, San Francisco, CA 94158

[3]Department of Bioengineering and Therapeutic Sciences and California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, CA 94158

## Abstract

Several distinct computational approaches have recently been implemented to represent conformational heterogeneity from time- and space-averaged X-ray crystallography datasets. As these modeling methods mature, newly discovered alternative conformations are being used to derive functional protein mechanisms. Room temperature X-ray data collection is emerging as a key variable for sampling functionally relevant conformations also observed in solution studies. Although concerns about radiation damage are warranted with higher temperature data collection, “diffract and destroy” strategies on X-ray free electron lasers may permit radiation damage-free data collection. X-ray crystallography need not be confined to “static unique snapshots”; these experimental and computational advances are revealing how the many conformations populated within a single crystal are used in biological mechanisms.

## Introduction

Macromolecular X-ray crystallography measures averaged intensities diffracted from $\sim 10^{13}$ molecules in the crystal lattice. The resulting electron density map, which is used to locate the positions of atoms in the unit cell, is therefore an ensemble-averaged probability distribution. Traditionally, structural models are built into the highest peaks of the electron density distribution. The lower electron density values that surround these high signals are fit by the B-factor (the temperature factor, thermal factor, Debye-Waller factor, or atomic displacement parameter) [1] (Figure 1a), which models the isotropic fall-off of the density from the mean position as a Gaussian. Although the electron density distribution around each atom is often anisotropic, the additional parameters needed for anisotropic B-factors



Correspondence to: JSF - james.fraser@ucsf.edu, tel: 415-493-8421, fax: 415-514-0169.

require a higher number of observations that is only afforded by very high resolution data [2].

Additionally, even in the crystal lattice, proteins can adopt multiple discrete conformations [3]. Both anharmonic motion and static disorder can result in multiple relatively weak peaks in electron density maps [4]. Even individual anisotropic or grouped Translation Libration Screw (TLS) B-factors are an insufficient description of these multiple minima in the probability distribution [5-7] (Figure 1b). This review focuses on the emerging strategies for modeling conformational heterogeneity from X-ray data and the potential for modeled alternative conformations to generate new mechanistic insights into the function of macromolecules.

## Electron density maps: More than meets the eye

Local maxima in electron density maps can be difficult to identify visually when electron density maps are rendered at a single threshold, as is common in the isosurface wire-frame representation used by Coot [8] and other graphics programs. Multiple contours or colour maps may be preferable for identifying conformations in weak, irregular electron density [9]. Maps modified by applying local feature enhancement [10], maximum entropy principles [11], and B-factor sharpening [12] may also aid in visually identifying important alternative conformations.

An alternative approach to visual inspection of electron density maps involves plotting electron density distributions as a function of dihedral angle (Figure 1c) [13]. Most side chain alternative conformations are confined to preferred rotameric torsion angles. The program Ringer identifies peaks originating from discrete alternative conformations sampled by protein side chains. Recent work demonstrates that these secondary peaks preferentially occur at low-energy rotameric positions [14]. Placing maps on an absolute electron-density scale and using new estimates of the noise throughout the map increases the power of Ringer to identify alternative conformations of protein and ligand atoms [15].

## Transforming electron density into models reflecting heterogeneity

Performing multiple independent refinements from a slightly perturbed starting model can reveal conformational heterogeneity (Figure 2a). The starting diversity can be generated by multiple simulated annealing trajectories [16-18], alternative parameter sets [19], Monte Carlo sampling [20], or randomly-seeded automated rebuilding (using ARP/wARP [21], RAPPER [22], or Phenix Autobuild [23]). Blundell and colleagues interpreted the variability across multiple RAPPER rebuilding runs as yielding a “potentially more accurate representation of the true underlying structure than does a single model” [23,24]. However, Terwilliger later showed, using synthetic data and Phenix Autobuild, that repetitive rebuilding often does not accurately sample known conformations, but rather gives an estimate of the imprecision of the coordinates of the main conformation [25].

Averaging the structure factors from independently refined structures often results in a lower $R_{free}$ than that from any individual refinement [25]. This result suggests a logical extension of the multiple independent copies approach: multiple conformations that are simultaneously

refined together as an ensemble (Figure 2b). In an early application, Kuriyan showed that "twin" conformations, moving independently while contributing equally to the refinement, could improve R-factors relative to a single model [26]. But why stop at "twin" refinement? Many subsequent ensemble-modeling efforts scanned 2-20 copies of the protein [27-30]. In a recent, more comprehensive evaluation of ensemble refinement, Phillips and colleagues scanned ensembles of 2, 4, 8, and 16 copies across 50 structural genomics targets and selected a final ensemble based on the lowest $R_{free}$ [31]. Their ensembles averaged 10.6 copies and improved $R_{free}$ by 1.9%, suggesting that the conformations sampled across the ensemble agree well with the X-ray data. Although the most obvious improvements generally occurred in going from 1 to 2 or from 2 to 4 copies, the diminishing returns do not preclude moving beyond 16 copies. Could having even more structures further improve $R_{free}$?

Time-averaging can slow the increase in effective free parameters as the number of copies increases: the samples in the ensemble are not truly independent as they are generated by a single molecular dynamics simulation restrained by (time-averaged) agreement with X-ray structure factors (Figure 2c) [32]. Although any individual snapshot is generally a poor fit to the observed X-ray data, the agreement improves when averaged over many simulation snapshots. In practice, this creates attraction towards relatively undersampled regions (those with positive $F_{obs}$-$F_{calc}$ difference density) and repulsion away from relatively oversampled local energy minima (those with negative $F_{obs}$-$F_{calc}$ difference density). Although the original implementation, which did not include any B-factors, was susceptible to overfitting [33], including a reasonable B-factor model allowed for parallel reductions in R and $R_{free}$ during the simulation [34-36].

Burnley and Gros have recently contributed a dramatically improved time-averaged ensemble refinement method [37], which is incorporated into the Phenix software suite [38]. To account for lattice disorder, an underlying TLS [39] B-factor model is first fit to the core of the molecule and then applied to the entire modeled structure. In addition to a bulk solvent model, explicitly-modeled solvent atoms are added and deleted throughout the simulation based on conventional real-space map criteria. In principle the restrained simulation could be extended to produce an arbitrarily large number of snapshots. Burnley and Gros keep only those simulation blocks of 250 consecutive snapshots with low $R_{work}$ values. The still ungainly number of structures is further reduced to the final output ensemble, the smallest evenly-distributed set of structures that reproduces the time-averaged $R_{free}$ to within 0.1%. This procedure resulted in ensembles containing between 50-800 structures across a wide variety of proteins [37]. In addition, the Gros group has recently applied this exciting method for an in-depth study of conformational heterogeneity of proteases [41].

A distinct model type, the multiconformer model, represents conformational diversity without creating multiple copies of the entire protein (Figure 2d). In multiconformer models, if the electron density distribution for a continuous segment is well fit by a single conformation with an appropriate B-factor model, then only a single conformation is used. However, if the electron density distribution suggests discrete conformations, the heterogeneous atoms are copied and given an "ALTLOC" identifier in the PDB record. This

second (or third or fourth) conformation is allowed to diverge and a new parameter "q" is refined reflecting the occupancies of the primary and secondary conformations. This tedious manual process can be subject to the whim and thoroughness of the model builder. Despite the small gains in $R_{free}$, modeling alternative conformations can reduce geometric distortions and rotamer-outlier side chain conformations. Fortunately, van den Bedem developed an automated approach to identify, build, and refine multiple conformations: qFit [42]. qFit enumerates many potential main and side chain conformations for each residue. Next, the program selects the optimal combination of conformations and associated initial occupancies based on combined fit to the density; often, only one conformation is selected. Fragments of neighboring residues are assembled using computational approaches borrowed from robotics to build the final model. Although correlated movement along directly adjacent residues sharing the same number of conformations is assumed, global correlated movements through tertiary contacts cannot be inferred directly from experimental data without reference to diffuse scattering features [43,44]. To address this problem, a companion approach, CONTACT, identifies networks of residues that can move between experimentally observed alternative conformations in a coupled manner [45]. In DHFR, residues with functionally relevant concerted motion (originally revealed using solution NMR experiments) were independently identified using the combined qFit - CONTACT approach. Both multiconformer and ensemble models present additional complications for validation. In particular, the use of $R_{free}$ in the parameter optimization and validation steps highlights a potential need for a new generation of model selection criteria [40].

## Warming up to different data collection strategies

As these modeling methods mature (Table 1), the major question is changing from "is there conformational heterogeneity?" to "what functions can conformational heterogeneity mediate?" Several recent studies, building on classic work by Petsko and others on the protein "glass transition" [46,47], highlight that the common practice of cryo-cooling can complicate the process of relating heterogeneity to function: cryo-cooling has obvious advantages in reducing radiation damage [48], but elegant theoretical studies by Halle suggest that the annealing that occurs during cryo-cooling may redistribute conformational heterogeneity [49]. The idea that conformational heterogeneity within the crystal can connect to biological function is dramatically exemplified by CypA [50]: data collection at room temperature allowed sampling of higher-energy conformations essential for catalysis also observed in solution by NMR experiments; in contrast, high-resolution cryogenic data revealed only a single conformational state. Similar changes to conformational ensembles were also observed across a larger sample of 30 proteins [51].

However, concerns about radiation damage have historically reduced the widespread use of higher temperatures during data collection. Cryo-cooling makes complete datasets obtainable from a single crystal for many systems that otherwise could not have structures determined. Several strategies have been applied to outrun or reduce the damage at "room temperature," where the temperature is generally maintained at 0-15°C. While it is always best to evenly expose the entire diffracting volume, the beam size is often limited by the physical setup of the beam line. In these cases, strategic translation can enable crystals to be exposed for longer and yield higher-quality data [52]. The ability to use higher dose rates

and free-radical scavengers to outrun damage remains controversial, and there are many aspects of room temperature diffraction still available to optimize [53].

X-ray free electron lasers (XFELs), which can deliver short pulses of extremely large doses of X-rays [54], decouple the relationship between diffraction, radiation damage, and temperature. In “diffract and destroy” data collection strategies, diffraction occurs on a faster timescale than the radiation damage, affording an approximately radiation damage-free view of the molecule at any temperature. Indeed, differences are already being observed between the same molecules imaged at cryogenic temperatures at synchrotrons and room temperature at XFELs [55]. While XFELs provide obvious applications for viewing conformational dynamics of proteins within a crystal, there remain several roadblocks before “molecular movies” can be routinely recorded [56]. In particular, improved crystal delivery methods [57], synchronized triggering of conformational changes [58], and data processing schemes [59] are on the horizon.

## Conclusions

What do the next 5-10 years hold? As the focus of our modeling efforts shift to representing the conformational ensemble, opportunities for integrative refinement and cross validation with solution experiments [60,61] will undoubtedly play a larger role. The major challenges will shift from describing conformational ensembles to understanding which of the populated conformations are important for biochemical functions. Time-resolved studies will likely be critical in this endeavor, but synchrotron-based Laue diffraction studies have previously been applied only to a limited set of systems [62-64]. The large changes in conformational ensembles often observed in the same crystal form [65,66] (Figure 3) coupled with the capabilities of XFELs for circumventing radiation damage at ambient temperatures suggest that the future will be dominated by teasing apart not only how the many conformations populated in a crystal relate to each other in space, but also how they relate to each other in time.

## Acknowledgments

We thank Gira Bhabha, Nathaniel Echols, R. Bryn Fenwick, James Holton, Daniel Keedy, Timothy Springer, Andrew VanBenschoten, Henry van den Bedem, and Peter Wright for helpful comments. RAW is supported by an NSF Graduate Research Fellowship. DAS is supported by NIH GM081879. JSF is supported by NIH OD009180, GM110580 and NSF STC-1231306.

## References

1. Jensen LH. Protein model refinement based on x-ray data. Annu Rev Biophys Bioeng. 1974; 3:81–93. [PubMed: 4370908]
2. Merritt EA. Expanding the model: anisotropic displacement parameters in protein structure refinement. Acta Crystallogr D Biol Crystallogr. 1999; 55:1109–1117. [PubMed: 10329772]
3. Phillips SE. Structure and refinement of oxymyoglobin at 1.6 A resolution. J Mol Biol. 1980; 142:531–554. [PubMed: 7463482]
4. Smith JL, Hendrickson WA, Honzatko RB, Sheriff S. Structural heterogeneity in protein crystals. Biochemistry. 1986; 25:5018–5027. [PubMed: 3768328]

5. Kuriyan J, Petsko GA, Levy RM, Karplus M. Effect of anisotropy and anharmonicity on protein crystallographic refinement. An evaluation by molecular dynamics. J Mol Biol. 1986; 190:227–254. [PubMed: 3795269]
6. Vitkup D, Ringe D, Karplus M, Petsko GA. Why protein R-factors are so large: a self-consistent analysis. Proteins. 2002; 46:345–354. [PubMed: 11835510]
7. Kuzmanic A, Pannu NS, Zagrovic B. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. Nat Commun. 2014; 5:3220. [PubMed: 24504120]
8. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
9. Maia FRNC, DeLano W, Szöke A, Van Der Spoel D. Interactive visualization of electron density slices. Journal of Applied Crystallography. 2005; 38:563–565.
10. Afonine PV. FEM: Feature Enhanced Maps. Computational Crystallography Newsletter. 2013; 4(2):28–29.
11. Glykos NM, Kokkinidis M. GraphEnt: a maximum-entropy program with graphics capabilities. Journal of Applied Crystallography. 2000; 33:982–985.
12. Liu C, Xiong Y. Electron density sharpening as a general technique in crystallographic studies. J Mol Biol. 2014; 426:980–993. [PubMed: 24269527]
13. Shapovalov MV, Dunbrack RL Jr. Statistical and conformational analysis of the electron density of protein side chains. Proteins. 2007; 66:279–303. [PubMed: 17080462]
14. Lang PT, Ng HL, Fraser JS, Corn JE, Echols N, Sales M, Holton JM, Alber T. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. Protein Sci. 2010; 19:1420–1431. [PubMed: 20499387]
15. Lang PT, Holton JM, Fraser JS, Alber T. Protein structural ensembles are revealed by redefining X-ray electron density noise. Proc Natl Acad Sci U S A. 2014; 111:237–242. [PubMed: 24363322]
16. Rice LM, Shamoo Y, Brünger AT. Phase Improvement by Multi-Start Simulated Annealing Refinement and Structure-Factor Averaging. Journal of Applied Crystallography. 1998; 31:798–805.
17. Korostelev A, Laurberg M, Noller HF. Multistart simulated annealing refinement of the crystal structure of the 70S ribosome. Proc Natl Acad Sci U S A. 2009; 106:18195–18200. [PubMed: 19822758]
18. Kohn JE, Afonine PV, Ruscio JZ, Adams PD, Head-Gordon T. Evidence of functional protein dynamics from X-ray crystallographic ensembles. PLoS Comput Biol. 2010; 6
19. Nwachukwu JC, Southern MR, Kiefer JR, Afonine PV, Adams PD, Terwilliger TC, Nettles KW. Improved crystallographic structures using extensive combinatorial refinement. Structure. 2013; 21:1923–1930. [PubMed: 24076406]
20. Knight JL, Zhou Z, Gallicchio E, Himmel DM, Friesner RA, Arnold E, Levy RM. Exploring structural variability in X-ray crystallographic models using protein local optimization by torsion-angle sampling. Acta Crystallogr D Biol Crystallogr. 2008; 64:383–396. [PubMed: 18391405]
21. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol. 1999; 6:458–463. [PubMed: 10331874]
22. Depristo MA, de Bakker PI, Johnson RJ, Blundell TL. Crystallographic refinement by knowledge-based exploration of complex energy landscapes. Structure. 2005; 13:1311–1319. [PubMed: 16154088]
23. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr D Biol Crystallogr. 2008; 64:61–69. [PubMed: 18094468]
24. Furnham N, Blundell TL, DePristo MA, Terwilliger TC. Is one solution good enough? Nat Struct Mol Biol. 2006; 13:184–185. discussion 185. [PubMed: 16518382]
25. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Adams PD, Moriarty NW, Zwart P, Read RJ, Turk D, Hung LW. Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models. Acta Crystallogr D Biol Crystallogr. 2007; 63:597–610. [PubMed: 17452785]

26. Kuriyan J, Osapay K, Burley SK, Brunger AT, Hendrickson WA, Karplus M. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins. 1991; 10:340–358. [PubMed: 1946343]
27. Rader SD, Agard DA. Conformational substates in enzyme mechanism: the 120 K structure of alpha-lytic protease at 1.5 A resolution. Protein Sci. 1997; 6:1375–1386. [PubMed: 9232638]
28. Burling FT, Weis WI, Flaherty KM, Brunger AT. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. Science. 1996; 271:72–77. [PubMed: 8539602]
29. Wilson MA, Brunger AT. The 1.0 A crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol. 2000; 301:1237–1256. [PubMed: 10966818]
30. Pellegrini M, Gronbech-Jensen N, Kelly JA, Pfluegl GM, Yeates TO. Highly constrained multiple-copy refinement of protein crystal structures. Proteins. 1997; 29:426–432. [PubMed: 9408940]
31. Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN Jr. Ensemble refinement of protein crystal structures: validation and application. Structure. 2007; 15:1040–1052. [PubMed: 17850744]
32. Gros P, van Gunsteren WF, Hol WG. Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. Science. 1990; 249:1149–1152. [PubMed: 2396108]
33. Clarage JB, Phillips GN Jr. Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50:24–36. [PubMed: 15299473]
34. Schiffer CA, van Gunsteren WF. Accessibility and order of water sites in and around proteins: A crystallographic time-averaging study. Proteins. 1999; 36:501–511. [PubMed: 10450092]
35. Schiffer CA, Gros P, van Gunsteren WF. Time-averaging crystallographic refinement: possibilities and limitations using alpha-cyclodextrin as a test system. Acta Crystallogr D Biol Crystallogr. 1995; 51:85–92. [PubMed: 15299339]
36. Romo TD, Clarage JB, Sorensen DC, Phillips GN Jr. Automatic identification of discrete substates in proteins: singular value decomposition analysis of time-averaged crystallographic refinements. Proteins. 1995; 22:311–321. [PubMed: 7479706]
37. Burnley BT, Afonine PV, Adams PD, Gros P. Modelling dynamics in protein crystal structures by ensemble refinement. Elife. 2012; 1:e00311. [PubMed: 23251785]
38. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
39. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr. 2006; 62:439–450. [PubMed: 16552146]
40. Kleywegt GJ. Separating model optimization and model validation in statistical cross-validation as applied to crystallography. Acta Crystallogr D Biol Crystallogr. 2007; 63:939–940. [PubMed: 17704561]
41. Forneris F, Burnley BT, Gros P. Ensemble refinement shows conformational flexibility in crystal structures of human complement factor D. Acta Crystallogr D Biol Crystallogr. 2014; 70:733–743. [PubMed: 24598742]
42. van den Bedem H, Dhanik A, Latombe JC, Deacon AM. Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. Acta Crystallogr D Biol Crystallogr. 2009; 65:1107–1117. [PubMed: 19770508]
43. Wall ME, Adams PD, Fraser JS, Sauter NK. Diffuse X-ray scattering to model protein motions. Structure. 2014; 22:182–184. [PubMed: 24507780]
44. Kuzmanic A, Kruschel D, van Gunsteren WF, Pannu NS, Zagrovic B. Dynamics may significantly influence the estimation of interatomic distances in biomolecular X-ray structures. J Mol Biol. 2011; 411:286–297. [PubMed: 21645520]
45. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS. Automated identification of functional dynamic contact networks from X-ray crystallography. Nat Methods. 2013; 10:896–902. [PubMed: 23913260]

46. Tilton RF Jr, Dewan JC, Petsko GA. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. Biochemistry. 1992; 31:2469–2481. [PubMed: 1547232]
47. Frauenfelder H, Hartmann H, Karplus M, Kuntz ID Jr, Kuriyan J, Parak F, Petsko GA, Ringe D, Tilton RF Jr, Connolly ML, et al. Thermal expansion of a protein. Biochemistry. 1987; 26:254–261. [PubMed: 3828301]
48. Hope H. Cryocrystallography of biological macromolecules: a generally applicable method. Acta Crystallographica Section B. 1988; 44:22–26.
49. Halle B. Biomolecular cryocrystallography: structural changes during flash-cooling. Proc Natl Acad Sci U S A. 2004; 101:4793–4798. [PubMed: 15051877]
50. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T. Hidden alternative structures of proline isomerase essential for catalysis. Nature. 2009; 462:669–673. [PubMed: 19956261]
51. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T. Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc Natl Acad Sci U S A. 2011; 108:16247–16252. [PubMed: 21918110]
52. Zeldin OB, Brockhauser S, Bremridge J, Holton JM, Garman EF. Predicting the X-ray lifetime of protein crystals. Proc Natl Acad Sci U S A. 2013; 110:20551–20556. [PubMed: 24297937]
53. Warkentin M, Hopkins JB, Badeau R, Mulichak AM, Keefe LJ, Thorne RE. Global radiation damage: temperature dependence, time dependence and how to outrun it. J Synchrotron Radiat. 2013; 20:7–13. [PubMed: 23254651]
54. Spence JC, Weierstall U, Chapman HN. X-ray lasers for structural and dynamic biology. Rep Prog Phys. 2012; 75:102601. [PubMed: 22975810]
55. Liu W, Wacker D, Gati C, Han GW, James D, Wang D, Nelson G, Weierstall U, Katritch V, Barty A, et al. Serial femtosecond crystallography of G protein-coupled receptors. Science. 2013; 342:1521–1524. [PubMed: 24357322]
56. Neutze R, Moffat K. Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges. Curr Opin Struct Biol. 2012; 22:651–659. [PubMed: 23021004]
57. Weierstall U, James D, Wang C, White TA, Wang D, Liu W, Spence JC, Bruce Doak R, Nelson G, Fromme P, et al. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. Nat Commun. 2014; 5:3309. [PubMed: 24525480]
58. Aquila A, Hunter MS, Doak RB, Kirian RA, Fromme P, White TA, Andreasson J, Arnlund D, Bajt S, Barends TR, et al. Time-resolved protein nanocrystallography using an X-ray free-electron laser. Opt Express. 2012; 20:2706–2716. [PubMed: 22330507]
59. Hattne J, Echols N, Tran R, Kern J, Gildea RJ, Brewster AS, Alonso-Mori R, Glockner C, Hellmich J, Laksmono H, et al. Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. Nat Methods. 2014
60. Fenwick RB, van den Bedem H, Fraser JS, Wright PE. Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. Proc Natl Acad Sci U S A. 2014; 111:E445–454. [PubMed: 24474795]
61. Rinaldelli M, Ravera E, Calderone V, Parigi G, Murshudov GN, Luchinat C. Simultaneous use of solution NMR and X-ray data in REFMAC5 for joint refinement/detection of structural differences. Acta Crystallogr D Biol Crystallogr. 2014; 70:958–967. [PubMed: 24699641]
62. Aranda, Rt; Levin, EJ.; Schotte, F.; Anfinrud, PA.; Phillips, GN, Jr. Time-dependent atomic coordinates for the dissociation of carbon monoxide from myoglobin. Acta Crystallogr D Biol Crystallogr. 2006; 62:776–783. [PubMed: 16790933]
63. Ihee H, Rajagopal S, Srajer V, Pahl R, Anderson S, Schmidt M, Schotte F, Anfinrud PA, Wulff M, Moffat K. Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. Proc Natl Acad Sci U S A. 2005; 102:7145–7150. [PubMed: 15870207]
64. Schlichting I, Almo SC, Rapp G, Wilson K, Petratos K, Lentfer A, Wittinghofer A, Kabsch W, Pai EF, Petsko GA, et al. Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. Nature. 1990; 345:309–315. [PubMed: 2111463]

65. Johnson SJ, Taylor JS, Beese LS. Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. Proc Natl Acad Sci U S A. 2003; 100:3895–3900. [PubMed: 12649320]

66. Zhu J, Zhu J, Springer TA. Complete integrin headpiece opening in eight steps. J Cell Biol. 2013; 201:1053–1068. [PubMed: 23798730]

## Highlights

- X-ray crystallography contains information about multiple conformations
- Multiple strategies are being tested to model conformational heterogeneity
- Free electron lasers present new room temperature data collection opportunities
- The dream of routinely creating molecular movies of proteins now seems attainable

## Highlighted References

**** - of outstanding interest**

*Fenwick 2014* – Order parameters, calculated from X-ray data that integrate displacements from B-factors and alternative conformations, agree with NMR solution measurement. The combination of room temperature data collection and qFit multiconformer modeling is key to integrative studies by NMR and X-ray crystallography.

*Burnley 2012* – The updated time-averaged ensemble refinement method has been integrated into Phenix. A TLS model to capture lattice disorder and a dual bulk/discrete solvent model are key new features that lead to improved $R_{free}$ values.

*Liu 2013* – Serial femtosecond crystallography of GPCR microcrystals reveals conformational differences compared to a previously determined cryogenic structure. Although flexibility increases overall, functionally important regions are even more flexible than expected. Additionally, several key enthalpic interactions are strengthened by cryo-cooling.

*Zhu 2013* – This work is a striking example of the large conformational changes that can occur in the context of a crystal lattice. Soaking RGD ligand peptides into integrin headpiece crystals reveals, across 8 independent datasets, how allosteric changes propagate over 40 Å.

*** - of interest**

*Kuzmanic 2014* – Molecular dynamics simulations in the context of the crystal lattice provide a control dataset for showing that the RMSDs calculated from harmonic B-factors are a poor model for heterogeneity calculated over snapshots of the MD simulation. Outlier snapshot conformations conformations significantly influence the RMSD without greatly increasing electron density spread.

*Terwilliger 2008* – This work is a masterful control experiment using synthetic datasets to examine how well multiple independent refinements represent conformational heterogeneity.

*van den Bedem 2013* – This paper introduces a new approach for identifying networks of residues that collectively sample alternative conformations from multiconformer refinement of experimental X-ray data (CONTACT: COntact Networks Through Alternate Conformation Transitions).

*Zeldin 2013* – The authors contribute software to optimize data collection strategies to avoid radiation damage. This advance is especially important for experimental phasing and room temperature data collection.

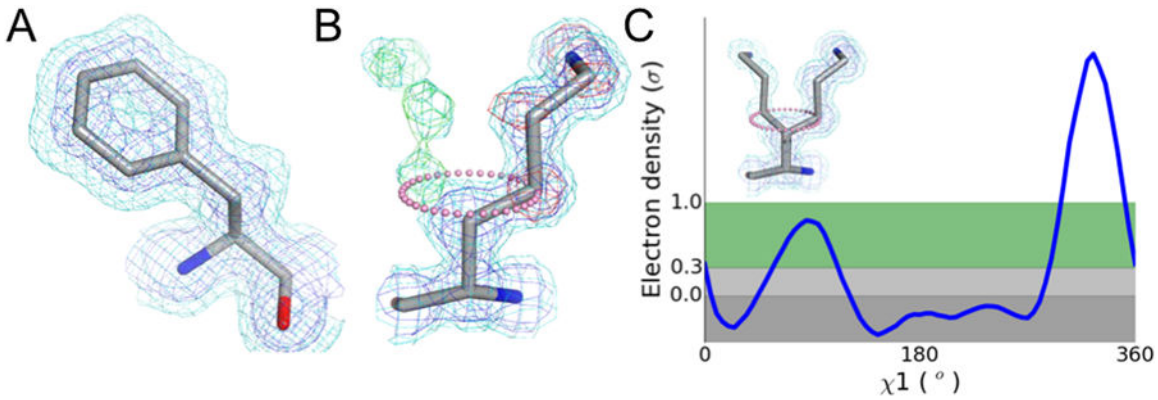


**Figure 1. Electron density maps contain ensemble-averaged information about multiple protein conformations**

**a)** The spread in electron density (blue mesh, 1σ high contour; cyan mesh, 0.5 σ low contour) around each atom is approximated by a B-factor, which models the thermal motion as a Gaussian displacement about the mean position. **b)** An electron density map with multiple maxima (blue mesh, high contour; cyan mesh, low contour) is inadequately modeled by B-factors, resulting in difference map peaks (red mesh, -1.5σ Fo-Fc peak underlying the model; green mesh, +1.5σ Fo-Fc peak indicating potential placement of alternative conformations). Because the alternative conformation partially overlaps with the primary conformation and is at lower occupancy, it is not visible at high contour. **c)** Sampling the electron density around the $\chi_1$ dihedral angle of the map shown in (b) (pink dots) reveals the presence of a distinct peak at the rotameric angle of -60°, providing an anchor point for manual model building of an alternative conformation. Automated model building is further complicated by the potential for backbone motions that can shift these peaks out of rotameric angles.

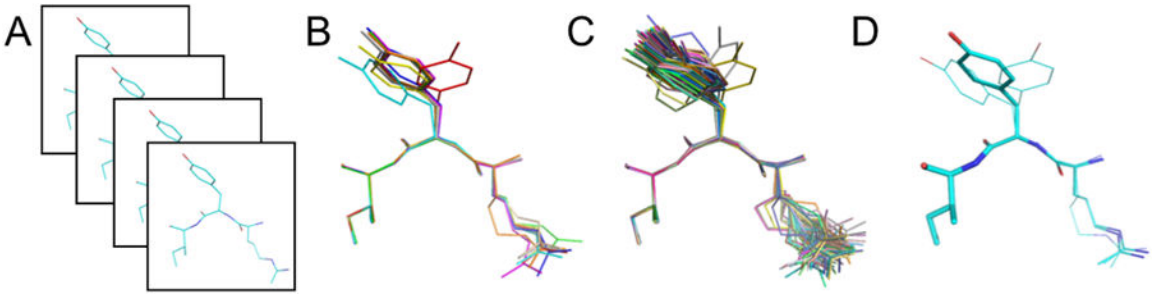


**Figure 2. Different model types are being used to interrogate conformational heterogeneity**

**a)** In multiple independent refinements, each copy contributes to a distinct set of model structure factors. The distinct structures, separated here by boxes, yield an estimate of the relative precision of the refinement method. **b)** In multi-copy ensemble refinement, a set number of copies of the protein, represented here by different colors, are refined together. **c)** Similarly, in time-averaged ensemble refinement, multiple copies of the protein are selected from an MD simulation where the structure factors are averaged over a defined time window. **d)** In multiconformer approaches, an optimal combination of between 1 and 4 conformations with associated occupancies (represented here by the thickness of the sticks) is constrained to sum to unit occupancy for each residue.

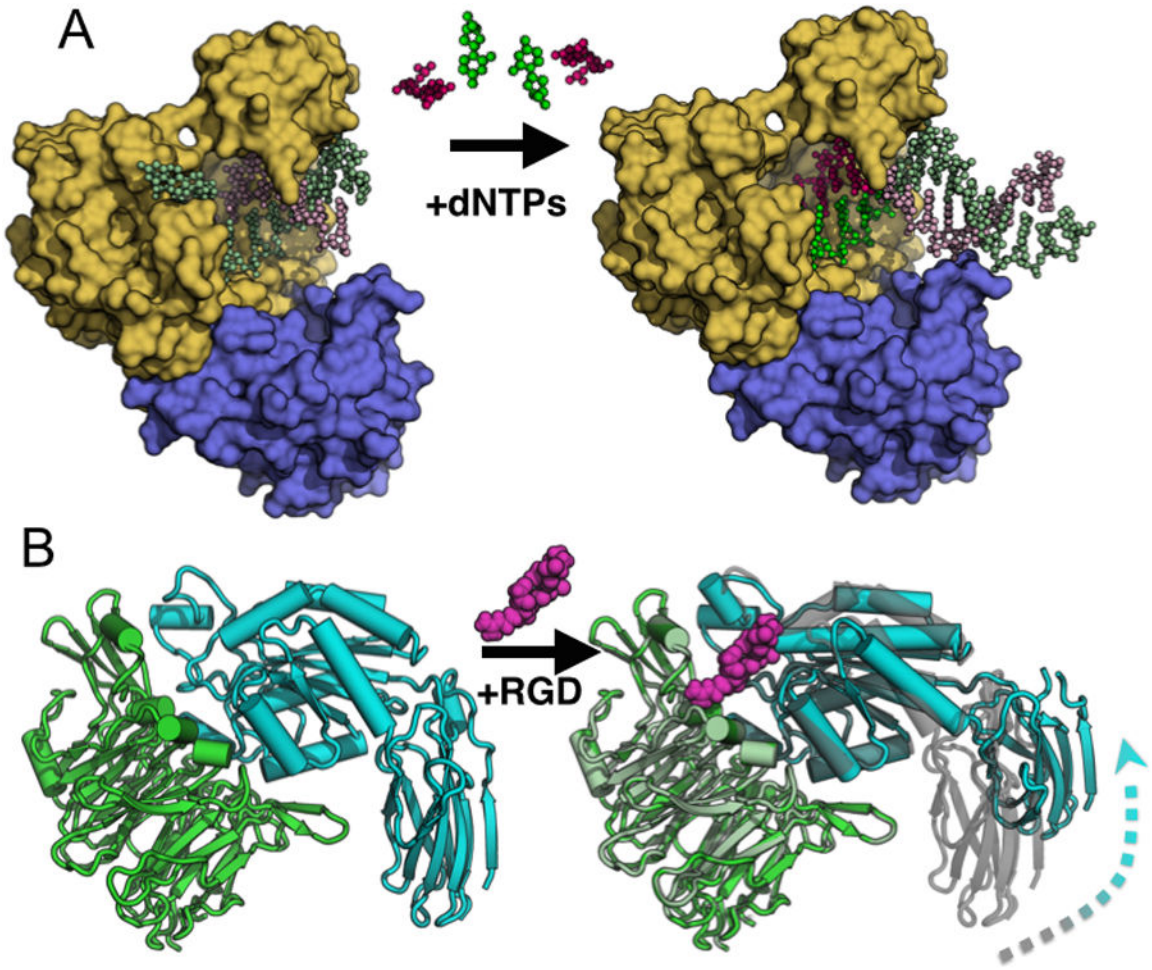


**Figure 3. Examples of functional conformational changes within a single crystal lattice**

**a)** *Bacillus* DNA polymerase I can catalyze DNA synthesis in the crystal lattice. The initial complex (PDB: 1L3S) containing the polymerase domain (yellow) and exonuclease domain (blue) is soaked with dNTPs. During catalysis the nascent strands are extended (brighter colors) and the pre-existing strands are extruded towards the solvent channels (PDB: 1L3V). **b)** The head domain of $\alpha_{IIb}\beta_3$ integrin (PDB: 3ZDX) undergoes a large allosteric conformational change when the ligand RGD peptide (magenta) is soaked into the crystal lattice (PDB: 3ZE2). Aligning the β-propeller domains (green) reveals how the $\beta_3$ inserted and hybrid domains extend and swing away (cyan) from the initial position (grey) upon peptide binding.

## Table 1

**Comparison of approaches to address the challenge of modeling conformational heterogeneity**

| | Independent | Multi-copy Ensemble | Time-averaged Ensemble | Multiconformer |
|---|---|---|---|---|
| **Overall goal** | Many independent models that each independently explain the data | Fixed number of models that collectively explain the data | Ensemble of models related by dynamical simulation and fit to the data | Single model with locally-fit 1-4 conformations per segment |
| **Diversity generation** | Multi-start simulated annealing or automated rebuilding | Simulated annealing | Molecular dynamics simulation augmented with an X-ray energy term | Rotamer library extended from backbone atoms sampled along extent of anisotropic ellipsoids |
| **Output** | User-specified number of completely independent models | 2-16 copies of the protein with equal occupancies, collectively contributing to $F_{calc}$ | 50-800 related models with equal occupancies, selected from the simulation | Single model with each residue having 0-3 alternative conformations, each with a partial occupancy |
| **User-defined parameters** | Rebuild fragment size, degree of cross-over between parallel models, and others depending on program | Ensemble size, B-factor model | TLS B-factor group selection, simulation relaxation time (default value determined by data resolution), X-ray:MD energy weight | Extent of sampling of backbone conformations from which to build rotamer library |
| **Output structure selection** | Retain all models within an $R_{free}$ threshold | Vary ensemble size to minimize $R_{free}$ | Minimize sampling frequency subject to maintaining $R_{free}$ within 0.1% of final rolling average | Optimal combination of conformations for each residue that best explains local density features |
| **Potential weaknesses** | Biased to local energy minima; yields only an estimated precision of the refinement procedure | Low observation-to-parameter ratio for larger ensembles; occupancies fixed to number of models | Validation and interpretation of ensemble models requires new tools; coupling information may be limited | Limited backbone conformational sampling |
| **Potential advantages** | Extensive sampling of starting conformations | Sampling of anharmonic distributions | Time averaging may limit number of free parameters | Refinement of intermediate occupancies; limited number of free parameters |
| **Implementation** | RAPPER[22], phenix.autobuild[23], ExCoR[19] | CNS[31], custom scripts in other refinement protocols | phenix.ensemble-refinement[37] | qFit[42] |