TOOLS FOR PROTEIN SCIENCE

New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics

Michael G. Prisant (ID) | **Christopher J. Williams** (ID) | **Vincent B. Chen** (ID) |
Jane S. Richardson (ID) | **David C. Richardson** (ID)

Department of Biochemistry, Duke University Medical Center, Durham, North Carolina

Correspondence
Jane S. Richardson, Department of Biochemistry, Duke University Medical Center, Durham, NC 27710.
Email: jsr@kinemage.biochem.duke.edu

Abstract

The MolProbity web service provides macromolecular model validation to help correct local errors, for the structural biology community worldwide. Here we highlight new validation features, and also describe how we are fighting back against outside developments which compromise that mission. Our new tool called UnDowser analyzes the properties and context of clashing HOH "waters" to diagnose what they might actually represent; a dozen distinct scenarios are illustrated and described. We now treat alternate conformations more thoroughly, and switching to the Neo4j database (graphical rather than relational) enables cleaner, more comprehensive, and much larger reference datasets. A problematic outside change is that refinement software now increasingly restrains traditional validation criteria (geometry, clashes, rotamers, and even Ramachandran) in order to supplement the sparser experimental data at 3–4 Å resolutions typical of modern cryoEM. But unfortunately the broad density allows model optimization without fixing underlying problems, which means these structures often score much better on validation than they really are. CaBLAM, our tool designed for evaluating peptide orientations at lower resolutions, was described in the previous Tools issue, and here we demonstrate its effectiveness in diagnosing local errors even when other validation outliers have been artificially removed. Sophisticated hacking of the MolProbity server has required continual monitoring and various security measures short of restricting user access. The deprecation of Java applets now prevents KiNG interactive online display of outliers on the 3D model during a MolProbity run, but that important functionality has now been recaptured with a modified version of the Javascript NGL Viewer.

KEYWORDS

backbone conformation, chiral volumes, ion binding, Neo4j, overfitting, Ramachandran restraints, server security, structure validation, water analysis

1 | INTRODUCTION

MolProbity does macromolecular model validation across a suite of criteria, for X-ray, neutron, NMR, computational, and now cryoEM models.[1–4] It is a current widely used system within a 30-year history of structure validation, where advances in validation methods have been prompted sometimes by advances in structural biology capabilities and sometimes by crises in distrust of the structural results.

The field of structure validation began around 1990, when crystal freezing, synchrotron data, and more powerful refinement methods such as simulated annealing in XPLOR[5] lowered R-factors, invalidating prior rules of thumb for when a structure had reached acceptable accuracy. That led to several serious mistracings and high-profile retractions such as for RuBisCo,[6] H-ras p21,[7] and HIV protease.[8] That crisis in turn prompted the development of Rfree,[9] Oops,[10] ProCheck,[11] WhatCheck,[12] and other validation systems.

By about 2000, more full-system automation had opened crystallography to nonexperts, and those tools were both exploited and advanced by Structural Genomics centers. Distrust of black-box crystallography prompted community demand for required validation in the customary "Table 1" of structure papers. For our research group, the driving factor was the failure of early protein de novo design to produce well-packed interiors and avoid molten globules.[13,14] That problem led us to develop optimized H addition & all-atom contact analysis for quantification of packing quality.[15,16] The contact analysis turned out also to be a powerful new tool for model validation, and its local and directional nature guides correction of the flagged errors. That process was extensively tested in production use[17] and has helped structural biologists improve their models ever since (see Section 2.1), up to about 2.5 Å resolution.

One investigator's set of 11 fraudulent structures[18,19] led to the worldwide PDB's Validation Task Forces for community organized standards,[20,21] validation on deposition including reports for referees,[22] and many further developments.

Just recently, the cryoEM revolution has produced an urgent need for better validation at 2.5–4 Å resolution, prompting new methods and tools including CaBLAM,[23] EMRinger,[24] Qscore[25] and the current and planned changes in MolProbity described here. Our new tools and strategies on one hand respond to excellent developments, such as the high-resolution revolution in cryoEM and expansion in the use of ensemble structures. On the other hand, we must also respond to negatives, such as targeted website attacks and the fact that at 2.5–4 Å the broad density necessitates modeling and refinement which directly or indirectly restrains most current validation criteria and makes the structures score better than they really are.

2 | RESULTS

2.1 | Indicators of progress

In the previous Tools issue[4] we described the epidemic overuse of unfavorable and very rare *cis*-nonPro peptides; only 1 in 3000 residues are genuine, usually functional occurrences.[26] We showed then that since addition of flags for *cis* and twisted peptides in MolProbity, Phenix, and Coot, there had already been some improvement in that problem. It is now clear that overuse of *cis*-nonPro has indeed gone back down nearly to pre-epidemic levels (Figure 1a).

Midresolution clashscores (number of non-H-bond steric overlaps ≥0.4 Å per thousand atoms) for worldwide PDB depositions have been steadily decreasing since the 2002 introduction of all-atom contact analysis by MolProbity. That metric has now leveled off at a reasonably low level of about four on average (Figure 1b). By definition clashscores cannot go below zero, and a clashscore less than about two is often a result of overfitting, since 2.7 is the average score in the best parts of the quality-filtered, high-resolution reference data.

Another very positive development is that in addition to increasingly thorough integration of model validation between the MolProbity web service and the validation GUIs in Phenix,[30] MolProbity validation is now also available in the CCP4 system.[31]

2.2 | New features: UnDowser to diagnose nonwater HOHs

At high to mid resolutions better than about 2.5 Å, density peaks are routinely seen for individual water molecules bound dynamically but reproducibly to favorable sites where they H-bond to protein, nucleic acid, or ligands. On an initial model, these will be positive difference peaks, into which waters are then fit either manually or automatically. However, it has long been known that not all such peaks actually represent waters.[32] The clashes of our all-atom contact analysis provide powerful criteria especially useful for distinguishing among the wide variety of cases that are something other than water, and applicable even to low-coordination sites at the molecular surface. A new feature in MolProbity called UnDowser now produces a table to show such diagnosis, to guide the user in making these decisions.
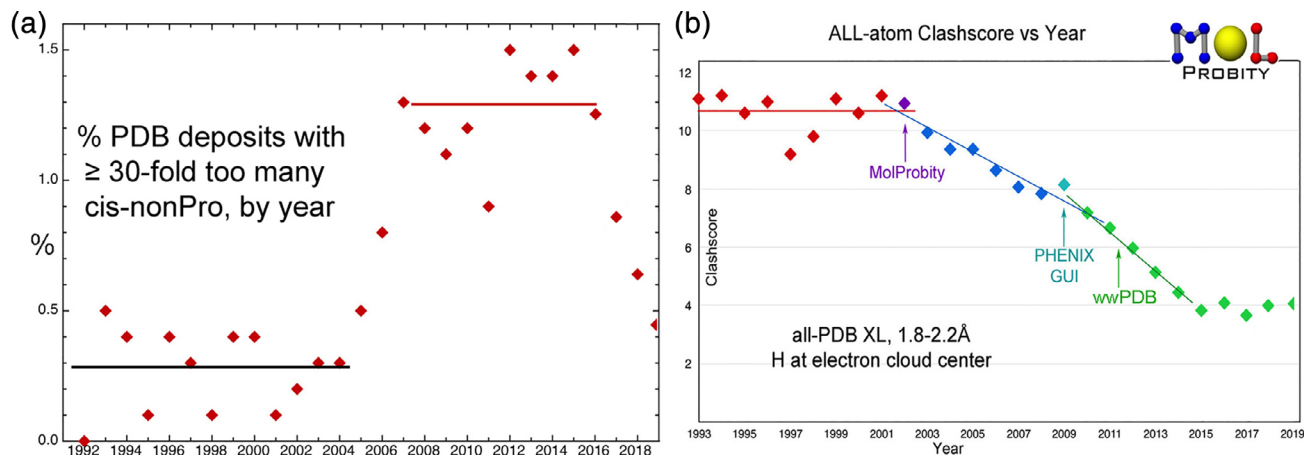
FIGURE 1 Timeline of MolProbity validation metrics. (a) Overuse of very rare *cis*-nonPro peptides by year in deposits to the worldwide PDB,[27] showing the abrupt rise around 2006[28,29] and the now successful return to pre-epidemic levels since *cis* and twisted peptides have been prominently flagged in MolProbity, Phenix, and Coot. (b) Midresolution (1.8–2.2 Å) clashscores (number of non-H-bond steric overlaps ≥0.4 Å per 1,000 atoms) by year, steadily decreasing since 2002 and now leveled off at about four, only somewhat above the 2.7 average for our good reference data

The table includes all clashing HOHs, sorted in approximate order of severity [by Sum(overlap − 0.2 Å)]. Type and charge of the atom(s) with which they clash are noted, and probable diagnoses of the underlying problem are given. Examples are shown and discussed here below for about a dozen distinct scenarios, to aid user comparison with their own cases.

An HOH that clashes with two or more atoms of the same polarity, and with no nonpolars or opposite polars, is almost certainly an ion. If all clashes or H-bonds are with negative atoms, then the HOH is a positive ion; if all interacting atoms are positive, then the HOH is a negative ion. Such interactions show up graphically as hotpink clash spikes inside the green dots of a putative H-bond. A doubly charged ion (e.g., Mg^{++}) almost always interacts with at least one fully charged atom (e.g., a phosphate or carboxyl O), while a singly charged ion (e.g., Na^+) often interacts just with partial charges (e.g., OH or backbone CO).

Occasionally, full coordination is seen, as shown in Figure 2a for the positive ion modeled as HOH 606 in the 6hhm sulfatase,[33] with two clashes to oxygen atoms. That HOH has six ligands—three backbone CO, two Thr Oγ, and a water—in closely octahedral geometry, although some at longer than ionic-bond distances. Once an ion has been diagnosed or explicitly modeled, coordination-geometry tools[35,36] can identify the most probable ion species, in this case Na^+.

If there is only one HOH-to-polar clash the diagnosis can be complex and depends on many factors of the context, such as clashes or close interactions with nonpolar or wrong-charge atoms, and relative density strength, B-factors, and shape of the groups involved. Clearly

diagnosable cases that include a single HOH-polar clash show up in three later figures, and an ambiguous case is discussed here. A partially-occupied +ion at a carboxyl O is relatively common, where the interacting O usually shifts somewhat when unbound. At high resolution a clashing water might be modeled, but at mid resolution this may just produce a strangely shaped sidechain density with an extra lobe. Figure 2b shows a strange triangular density shape with an extra lobe for Asp 388 of 6hhm at 1.23 Å resolution. The HOH modeled in that lobe cannot be either a water or an alternate conformation of the Asp (the lobe is too far from the Cα and too close to the Oδ). An ion is however rendered unlikely by the good H-bond to Arg 341 Hh1. In this confusing case the HOH should perhaps be deleted.

In contrast, genuine waters often H-bond to atoms of opposite charge, since their tetrahedral coordination includes two donors and two acceptors. Figure 2c shows a real water in the all-helix viral ORF of 6a4v, with good 2.2 Å density and weak but reasonable H-bonds to backbone NH and CO. A nonclashing HOH with a well-separated, round density peak at good H-bonding distance and angle from one or more polar atoms is nearly always a genuine water.

An extreme, unambiguous case is when after refinement there is extremely weak or absent density at the clashing HOH position. Such a "water" should be deleted, as for the case shown in Figure 2d of an HOH at 2.22 Å in the OmpU trimer of 5onu.[34] It superficially looks rather like a coordinated ion, but four of the clashes are to nonpolar atoms and the HOH peak has no density even at zero contour level. At present, the user must diagnose this, since inside the MolProbity website
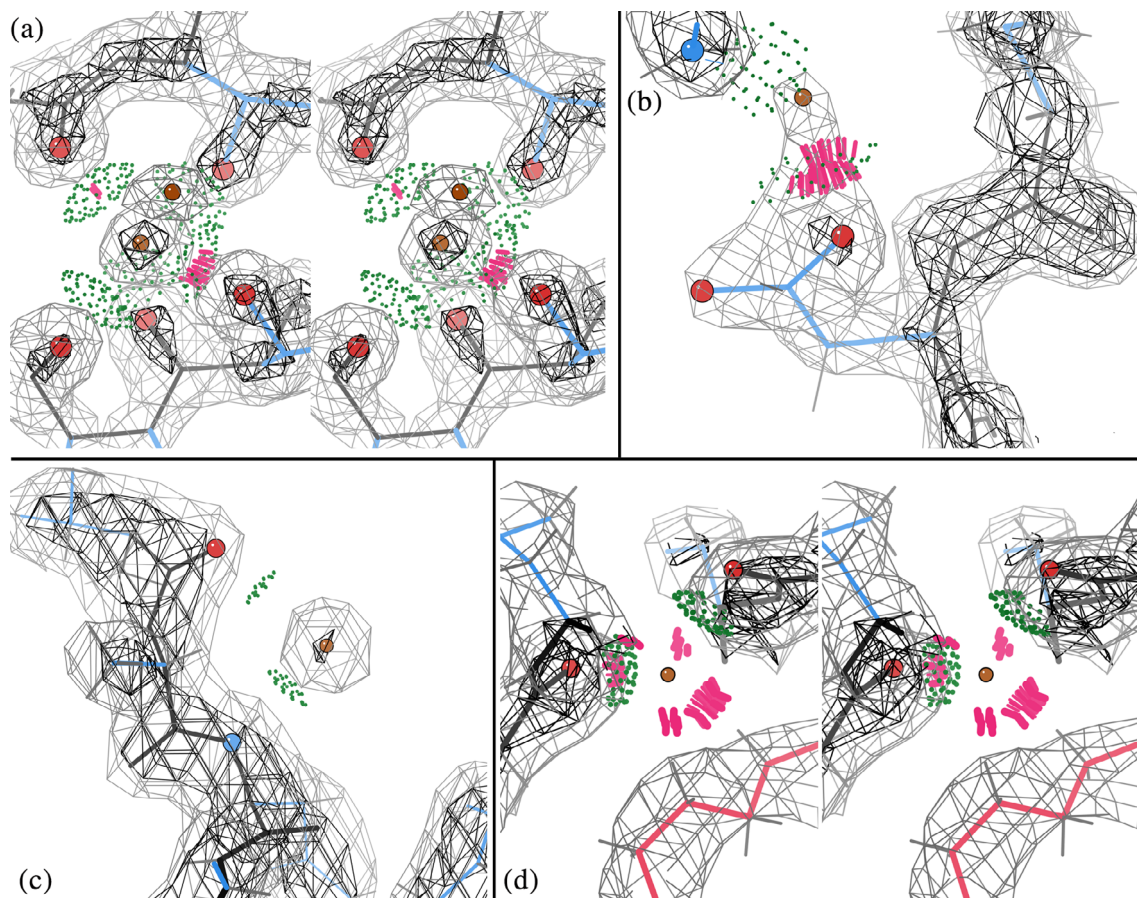
FIGURE 2 UnDowser diagnosis of HOH "waters": (a) Stereo image of a positive ion modeled as HOH, with six near-octahedral oxygen ligands. Lenses of green all-atom contact dots show putative donor–acceptor H-bond overlaps, but two distances are so short they also have serious clashes (clusters of hotpink spikes). HOH 606 in 6hhm.[33] (b) A confusing HOH with one clash to an Asp carboxyl O, but also a good H-bond to an Arg guanidinium, so not interpretable as an ion of either charge. HOH 603 in 6hhm. (c) A genuine water, with a strong, round density peak and 2 H-bonds with backbone atoms of opposite partial charge. HOH 414 in 6a4v (Hwang, unpublished). (d) Stereo image of an HOH with one polar and four nonpolar clashes. There is no electron density even at zero, suggesting that it should be deleted. HOH 504 in 5onu[34]

UnDowser does not have access to density maps (a future implementation inside Phenix will use information from electron density).

A similar-appearing case of many large clashes, including nonpolars with gorgeous density, is seen for HOH 51 and others in 3azd at 0.98 Å.[37] Apparently this rarely-seen problem is that the data themselves were detwinned rather than using a twin target in refinement. Detwinning lowered the information content enough that the 2mFo-DFc map is very nearly a calculated map which thus slavishly follows even incorrectly modeled features.

Comparison of B-factors between the HOH and surrounding atoms is often inconclusive because of B-factor restraints and partial-occupancy waters, but when extreme mismatches occur they offer a reason to look. For the 5onu case above, the ghost water has a B of 137 while the surrounding protein atoms average 45. On the other hand,

an HOH B-factor much lower than surrounding atoms or than well-ordered backbone O suggests a heavier atom than O even if it does not clash with anything.

If a water looks good but clashing atoms have higher B-factors, weaker density, or poorly fit density, then try fitting a different conformation for the clashing group that preferably H-bonds rather than clashes with the water. This often happens at lysine Nz, as shown in Figure 3a for the strong HOH B283 at 2.0 Å in 6aht,[38] which clashes with the weak end of Lys B106. Figure 3b shows a rebuilt, H-bonding lysine rotamer. The clashing atom can be at fault even with good density, if it is in a backward-fit sidechain amide or histidine, as shown in Figure 3c for HOH 1033 clashing with Cδ2 of the backward-fit His 73 of 1bkr at 1.1 Å.[39]

At high resolution, an HOH that clashes with a nonpolar atom (especially if along a sidechain) is often the next atom in an unmodeled alternate conformation.

FIGURE 3 HOH "waters," context interpreted: (a) Original fitting shows clash with Lysine methylene: HOH B283 at 2.0 Å in 6aht,[38] and (b) shows a rebuilt, H-bonding lysine rotamer. (c) Clash of HOH 1033 with Cδ2 of the backward-fit His 73 of 1bkr at 1.1 Å.[39] (d) Shows a "water" where an alternate conformation of Cys186 Sγ can be modeled for the 91–186 disulfide bond in 3ajd at 1.27 Å.[40] (e) Shows an HOH pair modeled at full occupancy that clash with each other and with mislabeled backbone alternates of Leu 105 at 0.88 Å in the 1gwe bacterial catalase,[41] and (f) shows a rebuild with the higher-density Leu and water alternates as altA at 60% occupancy and the lower-density ones at 40%
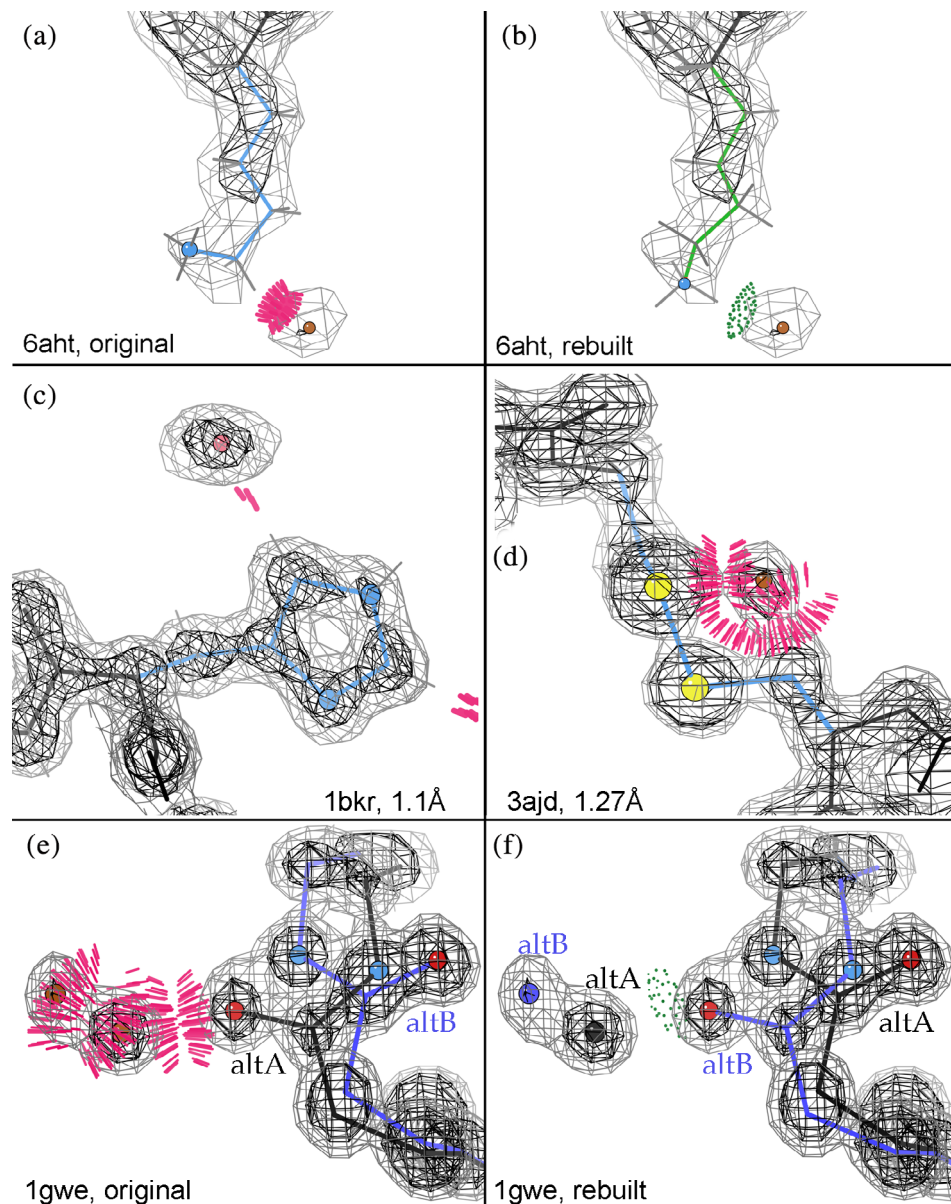

Fig. 8 of Richardson et al.[42] shows two HOHs clashing with the Cβ methylene of Asp 9 in the Zn protease of 1eb6 at 1.0 Å,[43] cleanly corrected with a small "backrub" adjustment[44] and the most common Asp rotamer. A similar situation is shown in Figure 3d for an HOH where the 186S should be, for an unmodeled alternate conformation of the 91–186 disulfide bond in 3ajd at 1.27 Å.[40] A partially reduced disulfide, common from radiation damage during data collection, might also have an SH incorrectly modeled as HOH.

On the other hand, an HOH that clashes with an already-modeled alternate conformation protein atom most likely is a real water that needs an occupancy <1.0, or sometimes the altA, altB naming is incorrect. Both those situations occur in Figure 3e for the full-occupancy HOH pair and Leu 105 mislabeled backbone alternates at

0.88 Å in the 1gwe bacterial catalase.[41] Figure 3f shows a rebuild with the higher-density Leu and water alternates as altA at 60% occupancy and the lower-density ones at 40%. Each alternate-conformation model is now clash-free and forms a water-backbone H-bond.

Two or more HOHs that clash with one another usually just need compatible partial occupancies <1.0. In fact, this is so common that the standard clashscore in MolProbity does not calculate or include clashes between waters; that is implemented only within UnDowser. However, two or more clashing HOHs may well represent an unmodeled ligand such as sulfate, GOL, or PEG. Figure 4a shows a set of three very badly clashing HOHs at 1.77 Å in the 1LpL CAP-Gly domain,[45] which H-bond with Lys 175 and Arg 212. Figure 4b shows this density rebuilt as a very convincing SO4 in 1tov.[17]
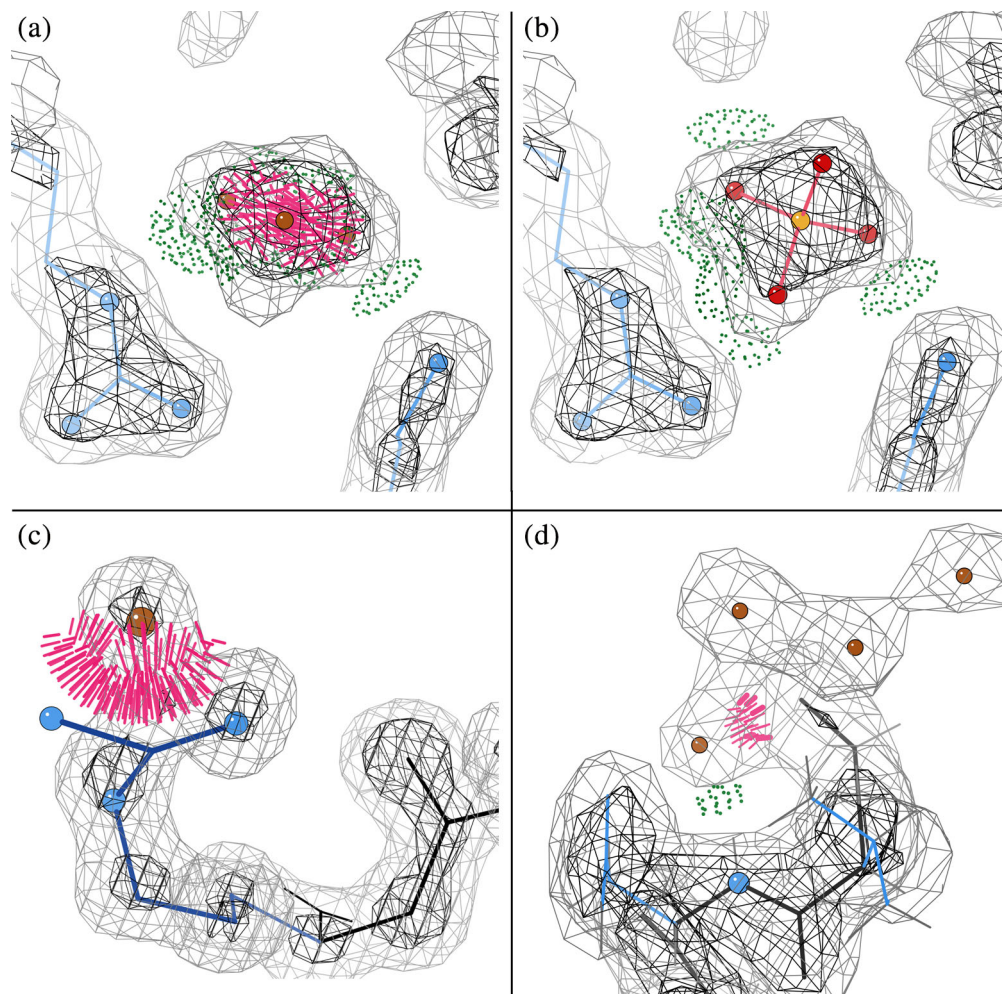
FIGURE 4 Sets of clashing HOH "waters" suggest reinterpretation of density. (a) Three very badly clashing HOHs at 1.77 Å in the 1LpL CAP-Gly domain,[45] which H-bond with Lys 175 and Arg 212, in (b) reinterpreted as a very convincing SO4 in 1tov.[17] (c) Suggests that premature placement of HOH A2090 forced an Arg A59 guanidinium N atom out of density in 1qLw at 1.09 Å.[46] (d) Shows four HOH loosely fit into connected density, one clashing with the backbone N of Gly 32, the first-modeled residue in the 2.22 Å OmpU of 5onu,[34] almost certainly representing an additional, earlier helical turn of backbone

A quite disruptive local problem is an HOH fit into density that actually belongs to some other atom which is thus kept out of its correct position, most often in a sidechain. This happens when the initial model (either molecular-replacement or ab initio) has the wrong rotamer. That produces a positive difference peak where the displaced atom should have been, and then automated water picking places an HOH there. Refinement cannot recover from this type of mistake, except for a procedure that deletes all waters before a rotamer-rebuilding step. However, an informed look at severely clashing, apparently strong "waters" can diagnose corrections on the basis of short distance and connecting density to sidechain atom(s) and poor fit for the sidechain. A stunningly obvious high-resolution example is shown in Figure 4c where HOH A2090 displaces an Arg A59 guanidinium N atom in the novel esterase of 1qLw at 1.09 Å,[46] to serve as a guide for recognizing such patterns even at lower resolutions.

For backbone atoms, incorrect HOH mimicry is most commonly found at a nonterminal chain end that should be extended. As an example, Figure 4d shows four HOH loosely fit into connected density, one clashing with the backbone N of Gly 32, the first-modeled residue in the 2.22 Å OmpU of 5onu.[34] That density almost certainly represents an additional, earlier turn of backbone, but the clashing HOH has displaced the carbonyl O atom of peptide 31–32.

For clashing waters, as well as for other validation outliers, the aim is to fix cases that are wrong in ways apt to be both significant and stable to further refinement: conformations in the wrong local minimum or groups that change existence or identity. The ability for clashes to flag meaningful errors of HOH identity depends on a good balance in the refinement target between steric overlap and density-fit terms. If sterics are overweighted, there will be clashes only for the very worst cases. If density fit is overweighted, then sometimes, at mid resolution and especially for cryoEM, there are many small water clashes spread throughout the structure, as seen for instance in the excellent 2.2 Å 5a1a β-galactosidase structure.[47] Most of these are not worth trying to fix individually. It might, however, be worth looking only at clashes >0.5 Å (dark pink in the MolProbity multichart), to find the more important and correctable problems.

This new water diagnosis, especially when it finds ions or ligands, can be of biological importance and is always an issue for computations based on the structure. We hope the UnDowser tool will make such choices much easier, even in this initial form, and its diagnostic criteria should continue to improve with experience in large-scale use.

2.3 | Application of CaBLAM to CryoEM

An especially influential change in direction and emphasis since the previous Tools issue[4] has come from our experience in using CaBLAM as assessors in two community exercises sponsored by the EMDB (Electron Microscopy Data Bank; https://www.ebi.ac.uk/pdbe/emdb/): the CryEM Model Challenge in 2016[48] and the CryoEM Model Metrics Challenge in 2019 (http://challenges.emdataresource.org/?q=model-metrics-challenge-2019). We also studied its use for crystal structures at 3–4 Å where a reliable answer was available at higher resolution.

That experience has demonstrated, in actual use, that CaBLAM supplies newly discriminating and robust overall and local validation information in the regime where traditional model validations fail: starting at 2.5–3 Å resolution and especially severe by 3.5–4 Å, for either cryoEM or X-ray structures.[48] Significantly different models can fit equally well to those broad density maps. If traditional outliers do occur, they still indicate real problems, but at these resolutions they can be "gamed" simply by applying restraints that shove outliers across the nearest boundary, which seldom fixes and often worsens the underlying problems, so that validation scores make the models look much better than they actually are.

The recent cryoEM "resolution revolution" was enabled by direct electron detector hardware and by image collection as movies with software correction of specimen motion. It has resulted in dramatic numbers of new cryoEM structures at unprecedented resolutions better than 4 Å, where *de novo* chain tracing is possible. Relative to X-ray crystal structures at these resolutions, refinement of models built into the cryoEM 3D reconstructions is done in real space rather than reciprocal space, phases are measured directly and start out quite good but are not improved by refinement, and backbone connectivity is typically a bit better for cryoEM maps. The differences across a structure in effective resolution/disorder/uncertainty are even larger than for crystal structures. Since electrons are sensitive to local electrostatic potential, negatively charged carboxyls show weak density and Arg and Lys sidechains are strong. For nucleic acids, bases are strong and phosphates relatively weak although still fairly spherical and recognizable.

Somewhere between 3.5 and 4 Å resolution, nucleic acid double helices switch confusingly between connectivity across clear basepairs and connectivity along the direction of base stacking. For protein helices, in the range between 2.5 and 5 Å resolution, the density gradually shifts confusingly between a spiral around a vacant axis and a tube with maximal density along the axis. In these transition regions some details in the density maps will give the wrong answer if followed too slavishly.

Relative to a density map at 2 Å or better, 3–4 Å density is broad, ambiguous, and sometimes even misleading. Therefore, model building and refinement must make use of more outside information to achieve physically reasonable models and well behaved refinement; even so, it is an extremely challenging task with extant tools. Covalent geometry (bond lengths and angles, planarity, etc.) needs to be quite tightly restrained. This causes no large problems because these are single-valued targets that cannot jump to some other allowed but incorrect local minimum. However, such tight restraints do destroy MolProbity's Cβ deviation criterion,[49] which flags incompatibility of sidechain and backbone conformation, such as a backward-fit branched sidechain. If geometry is perfect, then a Cβ cannot deviate from ideality even if perfect geometry puts it in the wrong position. In contrast, tight restraints on Ramachandran values or other multiple-minimum criteria make the scores better but the structure worse, since many of the changes pull the conformation into the wrong local minimum, as shown by examples below.

A major underlying problem is that at 2.5–3 Å or worse, separate protrusions for the backbone carbonyl oxygens disappear into the tube of backbone density. Since those are the strongest cue for fitting local backbone conformation, this makes misoriented peptides the commonest type of misfitting in 2.5–4 Å cryoEM (or X-ray) structures. Whenever a peptide orientation is off significantly, the preceding ψ and following ϕ angles are grossly wrong. That almost always means that optimizing the Ramachandran values moves them into the wrong local minimum on the plot.[48]

In order to attack this serious problem of misoriented peptides, CaBLAM was designed to directly measure incompatibility of modeled CO directions with the local Cα trace. For each five-residue stretch, it works from the relatively well-determined Cα virtual dihedrals to analysis of the much more dubious CO directions in the built model. (For more procedural detail, see Williams et al. in the 2018 Tools issue.[4]) Very rare pairings of the Cα virtual dihedrals themselves (plus the Cα virtual angle) are called Cα-geometry outliers and marked in red. They provide an effective diagnosis of probable errors in Cα-only models, which are deposited fairly often near 4 Å but have lacked suitable validation.

Since CaBLAM's parameters span several residues and are not part of any current refinement program's target set, its Cα-Cα-CO outliers can provide useful and independent validation at typical cryoEM resolutions. The 1% CaBLAM outliers are flagged graphically by magenta lines that follow the bad CO–CO dihedral, and on mouse-click or in numerical tables they provide conservative, quantitative diagnosis of which segments should be tried as regular α-helix or β-strand (Figure 5a, c). As true for other empirical-frequency validations, some outliers are genuine, and those are often of functional importance such as the ion-channel peptides of 6cju in Figure 5b.[49]

CaBLAM markup for a β-sheet region is shown in Figure 5a, from the recent 6cmx cryoEM structure at 3.1 Å resolution.[51] The model is correctly trying to fit antiparallel β strands, but not successfully, because they make almost no β-type backbone H-bonds, and ϕ, ψ for 7/10 residues in the lower two strands are favorable but in the α or Lα Ramachandran regions, not in β. Here, CaBLAM 1% outliers flag a common lower-resolution pathology where three or more carbonyl O atoms point in the same direction rather than alternating, while the probabilities given for β conformation are in a highly suggestive range of 0.1–0.25. The central CO of such a triplet is almost always misoriented. For each of the COs marked with an asterisk, a near-opposite peptide
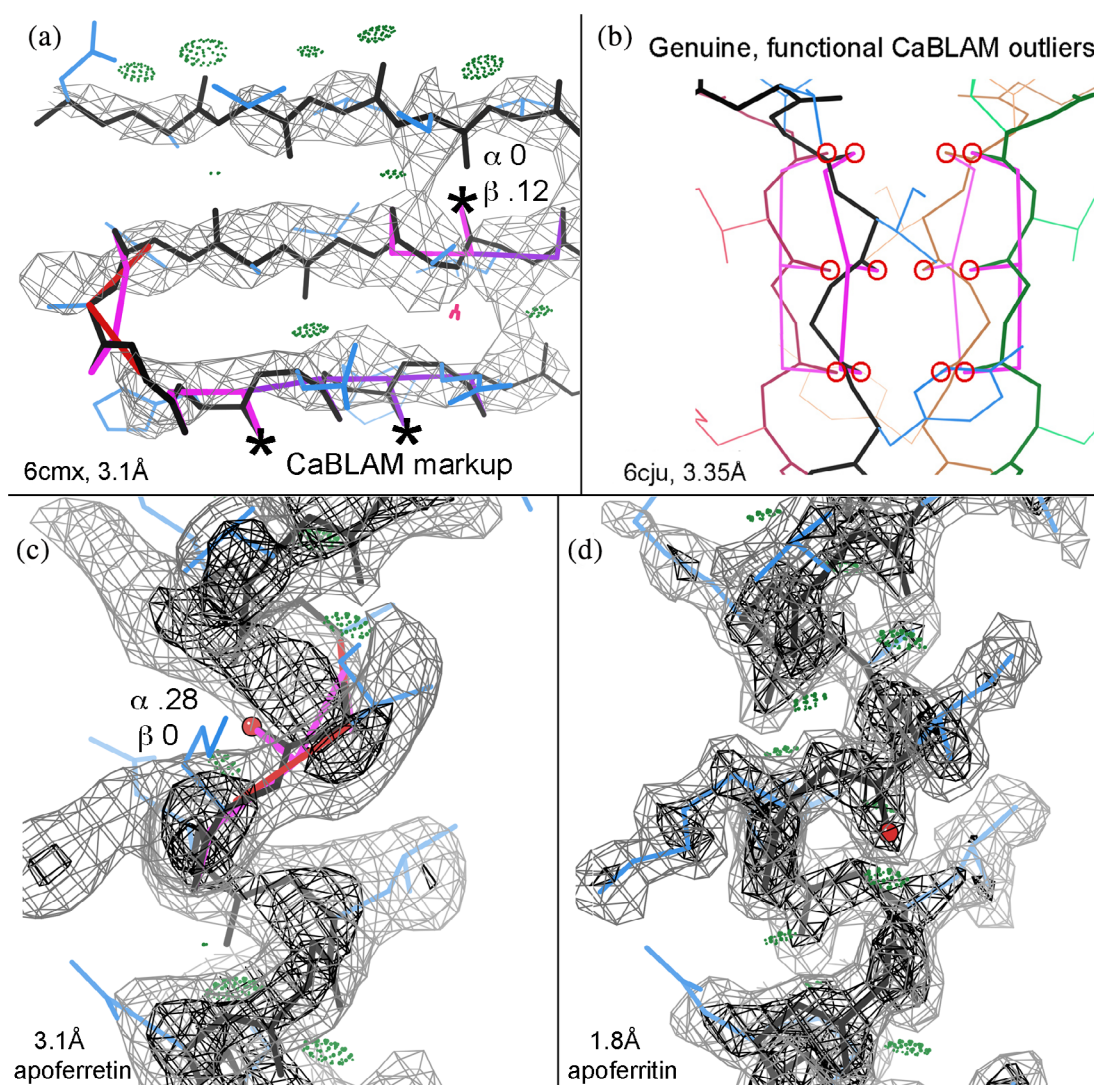

FIGURE 5 CaBLAM markup (Williams et al.[4] in the 2018 Tools issue) is (at present) orthogonal to fitting/overfitting and suggests diagnosis of which misfit segments should be tried as regular α-helix or β-strand. (a) CaBLAM suggests regular β-strands in 6cmx at 3.1 Å,[50] in spite of multiple successive outliers and almost no H-bonds. Each starred CO just needs a near-180° peptide rotation to achieve good β conformation and H-bonding. (b) Shows genuine, functional CaBLAM outliers of extended strands forming the ion-selectivity pore in 3.35 Å 6cju.[51] (c) CaBLAM suggests regular α-helix at a 0.28 score across a backward-pointing CO outlier in the broad 3.1 Å apoferritin density, and (d) shows the CO correctly placed in its clear density at 1.8 Å

rotation allows the CO to H-bond with an NH on a neighboring strand and removes the CaBLAM outliers. Once the two strands are corrected (not shown), the problematic turn can be refit as classic.

A very simple case of bad peptide orientation in α-helix is shown in Figure 5c. It comes from a model in the 2019 CryoEM Challenge, fit into a 3.1 Å apoferritin map. The CO of Ile 145 is flanked by two CaBLAM outliers, it points nearly opposite the other carbonyls, the α probability is 0.28, and preceding sidechains are pushed up from their density. The 1.8 Å apoferritin map in Figure 5d shows the O density side peak clearly, with a well-fit, helical model. If initial model building starts with ideal secondary structures (which we would recommend at resolutions poorer than 2.5 Å), then such outliers will not occur inside a helix or strand. However, outliers are still likely at the ends, and extending the regular α or β conformation should be tried as a likely correction. For instance, such a correctable C-terminal case (not shown) occurs at Lys 242 of 4heL at 3.2 Å (Meena and Saxena, unpublished), and an N-terminal X-ray case at 3.5 Å is shown and discussed by Moriarty et al.[52]

Backbone conformations, CaBLAM outliers, and their corrections are much more diverse in turns and loops. A few are genuine, some are minor, and some flag major errors that may or may not show any other validation outliers. The probabilities given by CaBLAM in loops for α or β conformation are seldom as high as 0.01. Figure 6a shows a CaBLAM outlier in the 99–104 loop of the 1de9 DNA-complex crystal structure at 3.1 Å.[53] Inspection shows residues out of density, which can be rebuilt (Figure 6b) once the problem is noticed.

If a model has a stretch of sequence misalignment, a disturbingly likely and serious problem at 3 Å or worse, CaBLAM outliers very often flag one or both ends of the misaligned region. Figure 6c shows an example from the alcohol dehydrogenase target at 2.9 Å in the 2019 CryoEM Challenge. A human alerted by multiple outliers can see that this model curves too tightly around the loop to fill its density, suggesting a missing residue. Indeed,
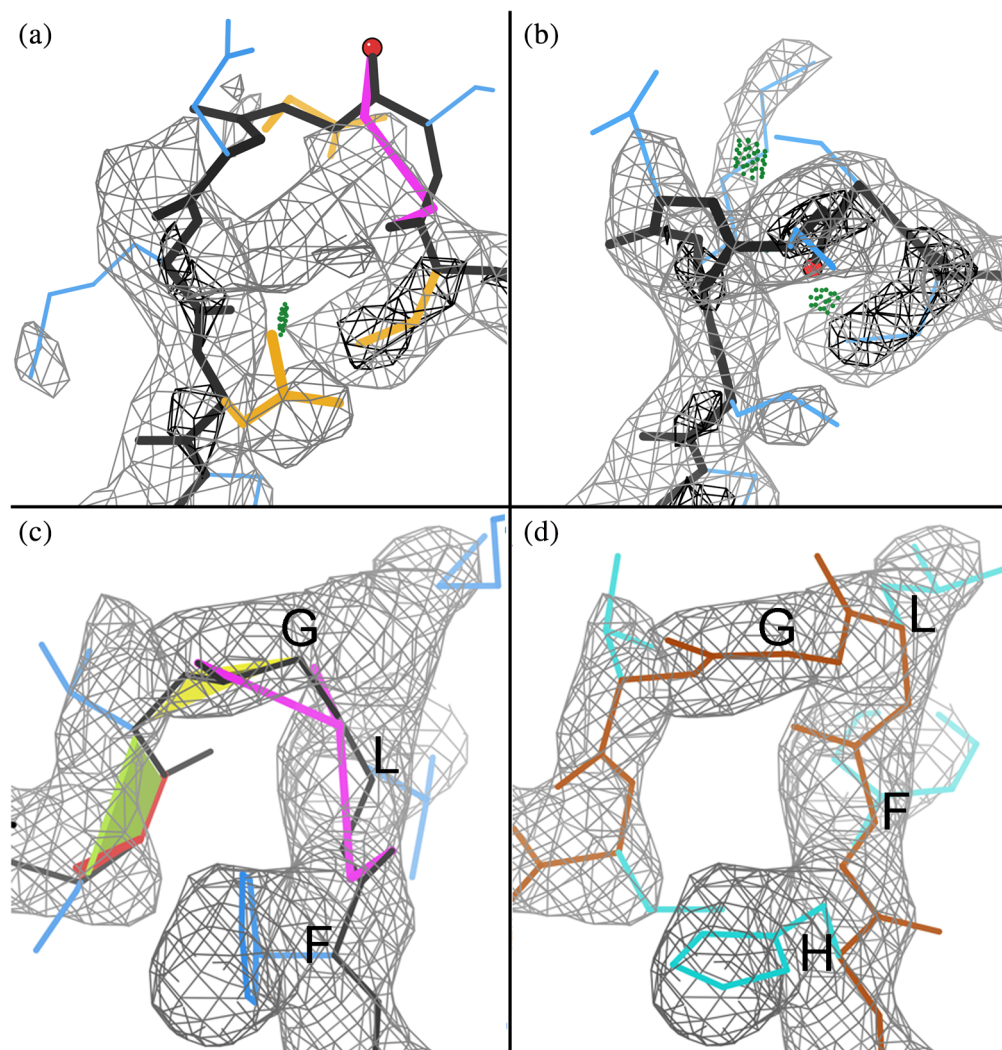

FIGURE 6 CABLAM markup as a guide to rebuilding regions poorly fitting into density: (a) shows a CaBLAM outlier in the 99–104 loop of the 1de9 DNA-complex crystal structure at 3.1 Å,[53] and (b) shows the rebuilt region fitting into density. (c) Indicates a missing residue by misalignment and poor fit of a loop in the alcohol dehydrogenase target at 2.9 Å in the 2019 CryoEM Challenge, and (d) shows a different, correct Challenge model which has the sequence aligned correctly and no outliers

Figure 6d shows a different, correct Challenge model which has the sequence aligned correctly and no outliers.

We and others are working toward automated correction for certain classes of CaBLAM outliers, or perhaps even initial avoidance of them. With guidance from the clear examples above at about 3 Å, manual correction should be feasible in Coot[54] or Isolde[55] for similar cases across the 2.5–4 Å range. Start with what looks easiest: in or at the end of secondary structure, or two successive outliers in an otherwise good context.

2.4 | Expanded treatment of alternate-conformation validation

Validation assessments that draw atoms from multiple residues present a particular challenge in bookkeeping. For a given residue, Ramachandran analysis requires the C atom of the preceding residue in sequence for the calculation of phi, and the N atom of the succeeding residue for psi. Similarly, the calculations in CaBLAM require O and CA atoms across a span of five residues. Thus multiple alternate positions for an atom in an adjacent residue can result in multiple alternate validation results for a residue that contains no alternate positions itself.

Previous versions of MolProbity calculated and presented alternate validations only for the simple case of residues that themselves contained alternate positions. The new version calculates and presents all validations resulting from alternate atom positions. In the results table, some residues that do not contain alternate atom positions are thus now labeled as having alternates because of their proximity to residues that do. This more complete reporting of alternate conformations should guide users to a more complete understanding of alternates, especially at the sometimes problematic points at which alternate conformations rejoin nonalternate structure.

For clarity in defining and presenting percentages and residue counts, the overall statistics presented in MolProbity's structure-level summary table are still calculated from non-alternate residues plus alternate A only.

2.5 | Chirality checking

MolProbity has not included an explicit check for chirality, because (a) model-building libraries or fragments include only the correct forms, (b) there are no chiral groups that span protein or nucleic acid monomer units, (c) standard refinement could not change chirality, and (d) inadvertent D-amino-acids are flagged by very large Cβ deviations. However, we have recently encountered a few cases,

including one where Amber refinement changed chirality at Cβ for a backward-fit Thr in order to pull all sidechain atoms into density peaks.[52] We have now implemented a check of chiral volume outliers, which reports only if there is a problem >4σ. It covers all the atoms defined as chiral centers in the Phenix Geo_std or monomer_library dictionaries. Three major cases are: (a) at Cα for D-amino-acids, either produced unintentionally or for genuine cases named as normal (e.g., Ala) rather than as D-amino-acids (e.g., Dal); (b) at Cβ for Thr or Ile; (c) at substituents on RNA or DNA sugar rings. Some apparent chiral-volume outliers are due to incorrect numbering in atom names, such as for OP1 versus OP2 on phosphates, or for numbering in Fe-S clusters, which are only pseudochiral. When distortions are large, but not large enough to flip chirality, we report them as geometry outliers.

2.6 | New reference dataset in progress

High-resolution PDB deposits have more than doubled since our current Top8000 quality-filtered dataset of protein chains, and we are now set up to use map-density criteria for even better residue-level filtering,[56] so that we can obtain even cleaner data from about twice as many residues. Also, graphical databases are now available which are more suitable for our purposes than relational databases; we are in the process of switching to Neo4j (https://neo4j.com/docs/2.1.5/introduction.html).

At our present stage of database completion, we have already generated complete Cβ deviation plots as a function of amino-acid type and of χ1, which were not as exciting as we had hoped, and have identified the 200 chain-terminal *cis*-nonPro peptides in the unfiltered data, all of which are incorrect. We have also started to explore multidimensional ϕ, ψ, χ plots, which can differ even more dramatically by χ than by amino acid; Asn and Gln are shown and contrasted in Reference 57. This new Neo4j database will soon be organized and populated enough to enable improved updates of the previous data distributions for all our conformational model validation criteria, especially valuable to improve the sensitivity level of CaBLAM outliers and occurrence statistics for UnDowser.

2.7 | Security issues

For the last several years, our main MolProbity web service has come under specifically tailored, sophisticated web hacking attacks, serious enough to threaten continuing existence of the site and forcing intensive defense efforts. Here we give a brief history of the nature of these

security threats, describe our defense strategies to keep this widely used community resource running, and note specific changes that may be visible to our users.

The foundational design of MolProbity occurred in a very different early internet era when one did not expect a strictly research service to be a hacking target. Since 2017, attacks on Molprobity have escalated to include thoroughly researched, specifically targeted hacks, presumably because it is located inside the network of a major medical center whose walls the attackers hope to breach.

Since its initial deployment, MolProbity has grown enormously in an organic fashion, with contributions from many authors. It is not a monolithic program but rather an ecosystem of diverse tools in many programming languages: C, C++, Java, Bash script, Perl, and more recently Python and Javascript. These executable tools are presented to the user through a dynamic, hand-coded web framework written in PHP. Users upload files for active analysis on the site and download the results. All of the code is open source, on GitHub. The PHP Query String Parameters that drive specific program execution are visible in the web URL request. Session management is file-directory based, with session state dynamically determined by session directory content. Those directories are also visible to an outside user, given modest reverse-engineering. Finally, our email bug-report system is susceptible to constant spamming. In short, MolProbity is too large and complex for us to rewrite completely, but is intrinsically vulnerable in our current internet environment.

Given these realities, our approach has been a patchwork of partial changes to the system, combined with constant monitoring.

Changes to the system include: (a) The actual Mol-Probity server machine is now housed outside the lab on a private, secured subnet. So we do not have direct physical access to the machine, and thus restoration of service in case of outage is no longer fully under our control. If the main site is down, try the mirror site at molprobity.manchester.ac.uk, but note that it may serve an older version. (b) We have limited the variety of uploads or fetches, and filter all uploads for Trojan scripts. Users must now begin with a coordinate file and cannot upload maps or kinemages for online display. We also check for well-formed PDB files, both at input and during processing, with the useful by-product of more helpful error messages for the commonest problems. (c) We have made modest changes to obstruct wide-open access to MolProbity's data directories. (d) We have migrated to a different bug-report mechanism less susceptible to spamming.

Constant monitoring is in part automated and in part administrator managed. System updates are performed daily. Monitoring includes: (a) Our servers are under constant packet-based and root-kit monitoring by the Duke University Health System security team, using a variety of enterprise tools. We also do similar monitoring using an independent set of tools. (b) We regularly monitor system memory, disk, runtime execution, and internet access profiles on the server through hand-written tools. This allows us to spot unusual activity as it occurs, and also to kill hung jobs, freeing capacity for other users. (c) The data directories are hand-scanned for Trojans on a regular basis, and are wiped several time a day. (d) We monitor the httpd and MolProbity logs for unusual activity and use firewall banning against suspicious URLs or those submitting batch runs too large for the server to handle. This too is accomplished with a set of hand-written scripts which are frequently modified as new threats emerge. We are committed to open, worldwide access for all types of legitimate users, so if your address seems to be incorrectly blocked, please send us email.

Unfortunately, it seems likely that serious, targeted attacks will be a growing problem for any scientific web site that performs active, open services for a worldwide user community. This puts a very significant burden on system management, in order to avoid exploit, denial-of-service or other mischief that could force closure of the site.

2.8 | Restoration of online kinemage graphics with NGL Viewer

Our Java-based KiNG display and modeling program[1,58] has historically served two separate functions. First, it provided interactive viewing, as a Java applet, of the validation results in 3D on the query model, seamlessly online with no download or installation needed. Second, when installed on the user's computer, KiNG offers a very wide set of user-friendly modeling and analysis capabilities. The deprecation of the Java applet functionality still allows all the locally-installed usage of KiNG but has destroyed its online use as part of a MolProbity run.

In response to this problem, we have replaced the online viewing functionality of KiNG with a modified version of the Javascript NGL Viewer developed for the RCSB PDB.[59] NGL Viewer gives excellent user perception of the 3D relationships and utilizes WebGL to provide blazingly fast interaction. In order to display kinemages in NGL Viewer, with the assistance of Alex Rose, we created a parser to translate the various kinemage objects (vectors, ribbons, balls, dots, etc.) into JavaScript objects that can be used in NGL Viewer. For ease of creating the initial interface, we used a demo GUI provided in the NGL Viewer code to display and control these translated kinemage objects. That GUI is currently capable of displaying and controlling all types of MolProbity validation
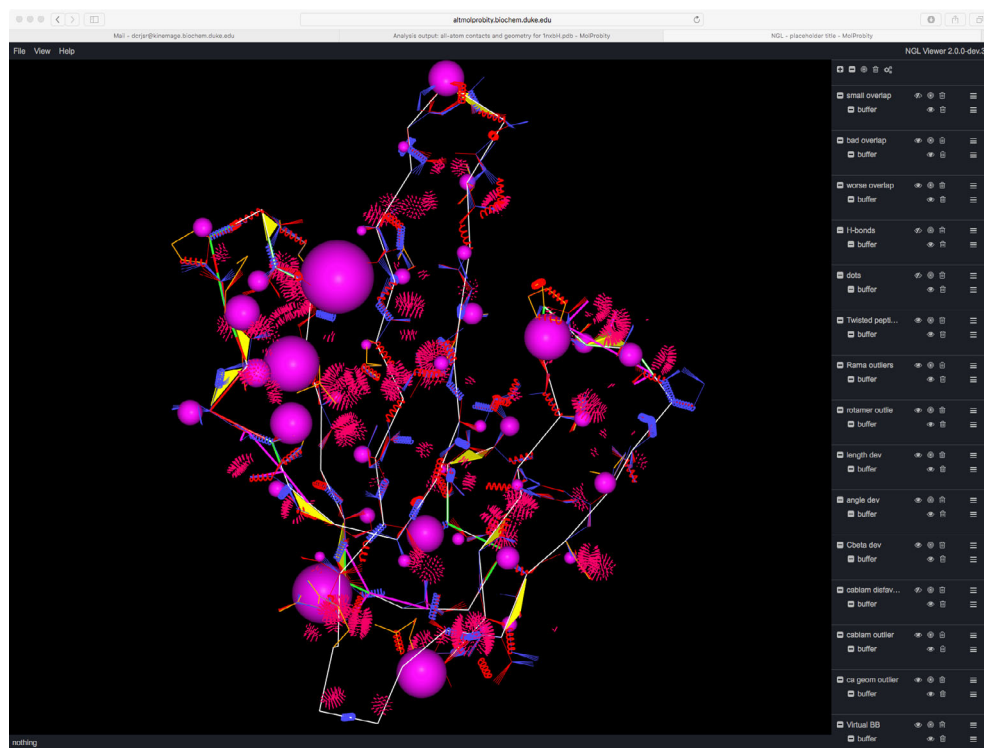
FIGURE 7 The demo GUI provided in the NGL Viewer code[59] has been modified to display and control kinemages of MolProbity validation markup. This screen capture shows a startup multicriterion kinemage view with all validation flags on the Cα trace. This small structure from the 1970's was chosen because it includes instances of all the types of problems within a simple view

markup (see Figure 7) and is being tested on the MolProbity beta site. This restores a user's ability, online within the web browser, during the run, to explore the multicriterion kinemage that shows local clustering and severity of all validation outliers on the 3D model.

The internal organization of the button-panel and mouse control for this GUI in NGL Viewer is quite different from the interface provided by KiNG and Mage,[60] so future plans for the software include a rewrite of the GUI. The rewrite will enable animation and better control over the kinemage groups and category "masters," as well as a more streamlined interface.

3 | DISCUSSION

We are very pleased to now have NGL Viewer as a replacement for online kinemage viewing, and UnDowser as a good initial utility for broad diagnosis of problematic peaks modeled as HOH. This new UnDowser tool, along with examples and descriptions for each scenario, should make it quick and easy for a structural biologist, or even an end-user, to come up with better reassignments for most clashing "waters" in a model. When implemented inside Phenix, it can become synergistic with existing tools for ion identification and other ligand identification, which have complementary strengths and shortcomings to those in UnDowser.

We have so far survived, at considerable cost, the highly targeted hacks on the MolProbity server. There is likely to become an increasingly serious need to develop and share protocols for providing open, worldwide access to important scientific web servers without a crippling level of vulnerability.

The recent cryoEM "revolution" is rapidly generating an unprecedented treasure of large, complex, dynamic, and biologically important structures. But, as explained here, there is a potentially dangerous disconnect in the process that could allow incorrect models and incorrect conclusions to go undetected until scientific contradictions pile up, as happened for similar reasons in crystallography around 1990 (see Section 1). The old measures developed since then (Rfree, Ramachandran, all-atom clashes, etc.) are still necessary but no longer sufficient, because the broad density at resolutions poorer than about 2.5 Å is compatible with multiple quite distinct local models, both correct and incorrect. Therefore a new set of validation criteria are badly needed that are independent, not easy to refine against directly, and diagnostic of local conformation but spanning more than a single residue to couple better with the larger-scale shapes in lower-resolution density. Our CaBLAM is almost the only current criterion with those properties, and it has indeed proven both sensitive and useful for making corrections, in practical use with cryoEM models. CaBLAM serves as a "chiropraxis" guide for realigning the "vertebrae" of peptide backbone into healthier and better-relaxed relationships, by much larger changes than our gentler "backrub"[44] motion. We and others will be working to develop additional independent criteria that can collectively fill this lower-resolution validation gap.

4 | METHODS

The overall workflow, and the underlying methods, for the prior state of MolProbity were described in the previous Tools issue[4] and have not changed.

UnDowser runs automatically along with the clashscore assessment, as it uses the same Reduce'd (hydrogens added) file and all-atom contact calculations but adds water–water interactions. Its input is the set of HOH entities in the coordinate file that have at least one all-atom clash (non-H-bond overlap ≥ 0.4 Å). Its output table is modeled on the MolProbity multicriterion chart, with a row for each clash interaction, sorted first by total severity for each water estimated as Sum(overlap − 0.2 Å), and within each water by individual clash severity. The row groups for each water are distinguished by alternate pale gray or white coloring. After complete identifiers and B-factors for each atom in the clash, succeeding columns classify each clash by the type of atom(s) the HOH clashes with (polar, nonpolar, other water, or alternate-conformation atom). Clash severity is color-coded progressively as pink, hotpink, or red by the same divisions of clash overlap (0.4, 0.5, 0.9 Å) as used in the main MolProbity multicriterion chart. The assignment of possible interpretation(s) for consideration and inspection are based on the characteristics of manually-evaluated examples such as those shown in the text. For instance, for an HOH that has ≥ 2 clashes with full or partial negative charges, no clashes with nonpolars, no clashes or H-bonds with positive polars, a B-factor close to or less than the average of surrounding atoms, and no alternate conformations involved, a positive ion will be the major suggestion. For an HOH that clashes with the first or last modeled backbone atom at a nonterminal chain end, the major suggestion is to try extending the chain further, starting with that clashing HOH. The rules are preliminary and will be reformulated after large-scale analysis with the new reference dataset.

The chirality check by chiral volume is run automatically along with covalent geometry validation, since it works from the same file produced by mmtbx.mp_geo.

The modified Javascript code to show multicriterion kinemages in NGL Viewer is available on GitHub in a fork from the main NGL Viewer repository, at https://github.com/vbchen/ngl.

The methods behind CaBLAM are described in the previous Tools issue.[4] Both the Python code and the underlying reference-data distributions are available from the rlabduke repository on GitHub, and are also distributed with Phenix.[30] Most CaBLAM examples and conclusions here were drawn from our assessments at the 2016–17 EMDB CryoEM Model Challenge[48] and at the 2019 EMDB Model Metrics Challenge (not yet published). In 2019, we used both our own MolProbity runs, including CaBLAM, and the scores, superpositions, and comparisons made available by Andriy Kryshtafovych on the Challenge web site (http://challenges.emdataresource.org/?q=model-metrics-challenge-2019), on all submitted models for all four targets. Reference models (assumed as essentially correct) were 3ajo[61] for 1.8, 2.3, and 3.1 Å effective resolutions of apoferritin, and 6nbb[62] for the alcohol dehydrogenase dimer target at 2.9 Å. We visualized CaBLAM outliers and density maps on multicriterion kinemages in KiNG[58] and analyzed their deleterious effect on ϕ, ψ values in kinemage Ramachandran plots with clickable and searchable residue datapoints. Outlier correction was defined on the Challenge targets by the reference-structure conformation, and for other structures the outliers were manually corrected in KiNG or Coot, with success declared if the result had no outliers and same or better map fit.

All figures except 1 and 7 were made in KiNG. PDB codes are given as lowercase except for L (4heL,1qLw), a convention that gives no ambiguities in any font, such as 1/I/l or O/0.[63]

ORCID

Michael G. Prisant ⓘ https://orcid.org/0000-0003-4095-387X
Christopher J. Williams ⓘ https://orcid.org/0000-0002-5808-8768
Vincent B. Chen ⓘ https://orcid.org/0000-0003-2492-979X
Jane S. Richardson ⓘ https://orcid.org/0000-0002-3311-2944
David C. Richardson ⓘ https://orcid.org/0000-0001-5069-343X

REFERENCES

1. Davis IW, Murray LW, Richardson JS, Richardson DC. MolProbity: Structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. 2004;32:W615–W619.
2. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007;35:W375–W383.
3. Chen VB, Arendall WB III, Headd JJ, et al. MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr. 2010;D66:12–21.
4. Williams CJ, Headd JJ, Moriarty NW, et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 2018;27:293–315.

5. Brunger AT. Crystallographic refinement by simulated annealing: Application to a 2.8 Å resolution structure of aspartate aminotransferase. J Mol Biol. 1988;203:803–816.

6. Knight S, Andersson I, Branden C-I. Reexamination of the three-dimensional structure of the small subunit of RuBisCo from higher plants. Science. 1989;244:702–705.

7. Pai EF, Krengel U, Petsko GA, Goody RS, Kabsch W, Wittinghofer A. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: Implications for the mechanism of GTP hydrolysis. EMBO J. 1990;9: 2351–2359.

8. Wlodawer A, Miller M, Jaskolski M, et al. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. Science. 1989;245:616–621.

9. Brunger AT. The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. Nature. 1992;355: 472–474.

10. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. 1991;A47:110–119.

11. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Cryst. 1993;26:283–291.

12. Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature. 1996;381:272.

13. Richardson JS, Richardson DC, Tweedy NB, et al. Looking at proteins: Representations, folding, packing, and design. Biophys J. 1992;63:1186–1209.

14. Richardson JS, Richardson DC. Doing molecular biophysics: Finding, naming, and picturing signal within complexity. Annu Rev Biophys. 2013;42:1–28.

15. Word JM, Lovell SC, LaBean TH, et al. Visualizing and quantitating molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. J Mol Biol. 1999;285:1711–1733.

16. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol. 1999;285: 1735–1747.

17. Arendall BW III, Tempel W, Richardson JS, et al. A test of enhancing model accuracy in high-throughput crystallography. J Struct Funct Genomics. 2005;6:1–11.

18. Janssen BJC, Read RJ, Brunger AT, Gros P. Crystallography: Crystallographic evidence for deviating C3b structure. Nature. 2007;448:E1–E3.

19. Borell B. Fraud rocks protein community. Nature. 2009; 462:970.

20. Read RJ, Adams PD, Arendall WB III, et al. A new generation of crystallographic validation tools for the Protein Data Bank. Structure. 2011;19:1395–1412.

21. Montelione GT, Nilges M, Bax A, et al. Recommendations of the NMR structure validation task force. Structure. 2013;21:1563–1570.

22. Gore S, Velankar S, Kleywegt GJ. Implementing an X-ray validation pipeline for the Protein Data Bank. Acta Crystallogr. 2012;D68:478–483.

23. Williams CJ. Using Calpha geometry to describe protein secondary structure and motifs [PhD thesis]. Department of Biochemistry, North Carolina, USA: Duke University, 2015;p. 248.

24. Barad BA, Echols N, Wang RY-R, et al. EMRinger: Side-chain-directed model and map validation for 3D electron microscopy. Nat Methods. 2015;12:943–946.

25. Pintile G, Chiu W. Assessment of structural features in cryo-EM density maps using SSE and sidechain Z-scores. J Struct Biol. 2018;204:564–571.

26. Williams CJ, Videau LL, Hintze BJ, Richardson JS, Richardson DC. Cis-nonPro peptides: Genuine occurrences and their functional roles. bioRxiv. 2018;324517.

27. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat Struct Mol Biol. 2003;10:980.

28. Croll TI. The rate of cis–trans conformational errors is increasing in low-resolution crystal structures. Acta Crystallogr. 2015; D71:706–709.

29. Williams CJ, Richardson JS. Fitting tips #9: Avoid excess cis peptides at low resolution or high B. Comput Crystallogr Newsl. 2015;6:2–6.

30. Liebschner D, Afonine PV, Baker ML, et al. Macromolecular structure determination using X-rays, neutrons, and electrons: Recent developments in Phenix. Acta Crystallogr. 2019;D75:861–877.

31. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. 2011; D67:235–242.

32. Van den Akker F, Hol WGJ. Difference density quality (DDQ): A method to assess the global and local correctness of macromolecular crystal structures. Acta Crystallogr. 1999;D55: 206–218.

33. Reisky L, Prechoux A, Zuhlke MK, et al. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. Nat Chem Biol. 2019;15:803–812.

34. Li H, Zhang W, Dong C. Crystal structure of the outer membrane protein OmpU from Vibrio cholerae at 2.2 Å resolution. Acta Crystallogr. 2018;D74:21–29.

35. Echols N, Morshed N, Afonine PV, et al. Automated identification of elemental ions in macromolecular crystal structures. Acta Crystallogr. 2014;D70:1104–1114.

36. Zheng H, Cooper DR, Porebski PJ, Habalin IG, Handing KB, Minor W. CheckMyMetal: A macromolecular metal-binding validation tool. Acta Crystallogr. 2017;D73:223–233.

37. Meshcheryakov VA, Krieger I, Kostyukova AS, Samatey FA. Structure of a tropomyosin N-terminal fragment at 0.98 Å resolution. Acta Crystallogr. 2011;D67:822–825.

38. Hayashi I, Oda T, Sato M, Fuchigami S. Cooperative DNA binding of the plasmid partitioning protein TubR from the Bacillus cereus pXO1 plasmid. J Mol Biol. 2018;430:5015–5028.

39. Banuelos S, Saraste M, Carugo KD. Structural comparisons of calponin homology domains: Implications for actin binding. Structure. 1998;6:1419–1431.

40. Kuratani M, Hirano M, Goto-Ito S, et al. Crystal structure of Methanocaldococcus jannaschii Trm4 complexed with sinefungin. J Mol Biol. 2010;401:323–333.

41. Murshudov GN, Grebenko AI, Brannigan JA, et al. The structure of Micrococcus lysodeikticus catalase, its ferryl intermediate (compound II) and Nadph complex. Acta Crystallogr. 2002; D58:1972–1982.

42. Richardson JS, Williams CJ, Hintze BJ, et al. Model validation—Local diagnosis, correction, and when to quit. Acta Crystallogr. 2018;D74:132–142.

43. Macauley KE, Jia-Xing Y, Dodson EJ, Lehmbeck J, Ostegaard PR, Wilson KS. A quick solution: Ab initio structure determination of a 19 kDa metalloproteinase using Acorn. Acta Crystallogr. 2001;D57:1571–1578.

44. Davis IW, Arendall WB III, Richardson JS, Richardson DC. The backrub motion: How protein backbone shrugs when a sidechain dances. Structure. 2006;14:265–274.

45. Li S, Finley J, Liu Z-J, et al. Crystal structure of the cytoskeleton-associated protein glycine-rich (CAP-Gly) domain. J Biol Chem. 2002;277:48596–48601.

46. Bourne PC, Isupov MN, Littlechild JA. The atomic resolution structure of a novel bacterial esterase. Structure. 2000;8: 143–151.

47. Bartesaghi A, Merk A, Banerjee S, et al. 2.2 Å resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science. 2015;348:1147–1151.

48. Richardson JS, Williams CJ, Videau LL, Chen VB, Richardson DC. Assessment of detailed conformations suggests strategies for improving cryoEM models: Helix at lower resolution, ensembles, pre-refinement fixups, and validation at a multi-residue length scale. J Struct Biol. 2018;204:301–312.

49. Lovell SC, Davis IW, Arendall WB III, et al. Structure validation by Cα geometry: φ, ψ and Cβ deviation. Proteins. 2003;50: 437–450.

50. Li J, Shalev-Benami M, Sando R, et al. Structural basis for teneurin function in circuit-wiring: A toxin motif at the synapse. Cell. 2018;183:735–748.

51. Rheinberger J, Gao X, Schmidpeter PA, Nimijean CM. Ligand discrimination and gating in cyclic-nucleotide-gated ion channels from apo and partial agonist-bound cryo-EM structures. Elife. 2018;7:e39775.

52. Moriarty NW, Janowski PA, Swails JM, et al. Improved chemistry restraints for crystallographic refinement by integrating Amber molecular mechanics into Phenix. bioRxiv. 2019;724567.

53. Mol CD, Izumi T, Mitra S, Tainer JA. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. Nature. 2000;403:451–456.

54. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr. 2010;D66:486–501.

55. Croll TI. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. 2018;D74:519–530.

56. Hintze BJ, Lewis SM, Richardson JS, Richardson DC. MolProbity's ultimate rotamer-library distributions for model validation. Proteins. 2016;84:1177–1189.

57. Richardson J, Richardson D, Williams C. Fitting tip #17—Asn and Gln are remarkably different. Comput Crystallogr Newsl. 2019;10:1–6.

58. Chen VB, Davis IW, Richardson DC. KiNG (Kinemage, next generation): A versatile interactive molecular and scientific visualization program. Protein Sci. 2009;18:2403–2409.

59. Rose AS, Hildebrand PW. NGL Viewer: A web application for molecular visualization. Nucleic Acids Res. 2015;43:W576–W579.

60. Richardson DC, Richardson JS. Mage, probe, and Kinemages, chapter 25.2.8. In: Rossmann MG, Arnold E, editors. IUCr's international tables of crystallography, volume F: Crystallography of biological macromolecules. Dortrecht: Kluwer Academic Press, 2001.

61. Masuda T, Goto F, Yoshihara T, Mikami B. The universal mechanism for iron translocation to the ferroxidase site in ferritin, which is mediated by the well conserved transit site. Biochem Biophys Res Commun. 2010;400:94–99.

62. Herzik MA Jr, Wu M, Lander GC. High-resolution structure determination of sub-100 kDa complexes using conventional cryoEM. Nat Commun. 2019;10:1032–1032.

63. Moriarty NW. Editor's note. Comput Crystallogr Newsl. 2015;6:26.