Expert Advice

Fitting Tip #23 – Alternate conformations always want to spread

Jane Richardson, Duke University, Nigel W. Moriarty, Lawrence Berkeley National Laboratory, and Daniel Keedy, CUNY Advanced Science Research Center

Introduction

In macromolecules at very high resolution, one sees increasingly many regions that have two, or even more, alternate conformations, with at least some of their atoms clearly distinct. They come in all sizes, from a single atom to many consecutive residues, and in a wide range of separation distance or conformational difference. They can be quite simple, as for the two proline puckers in Figure 1A, or quite confusing, as for the 4 hierarchical Lys alternates in Figure 1B with two distinct sidechains on each of two shifted backbones. If we want them to represent a physically possible ensemble, each individual local alternate must be internally consistent (free of geometry outliers) and consistent with the neighboring structure (free of all-atom clashes with its surroundings). Ensuring this is a far from trivial task. A more straightforward, but seldom done, task is to extend the ends of an alternate far enough not to produce geometry outliers.

Internal consistency

Either for sidechain or for backbone, alternates very often crisscross back and forth, as seen in Figure 2 below. That makes it tricky to be sure each alternate goes through the right subset of the atom peaks to be correctly bonded and with an acceptable rotamer conformation. If the two alternates have significantly different occupancies, their peak densities will be different for well separated atoms and add up for atoms much closer than the resolution. That is quite noticeable along the backbone in Figure 2, where the front carbonyl O is much stronger than the back one. Alt A (white) should be assigned to the higher-occupancy

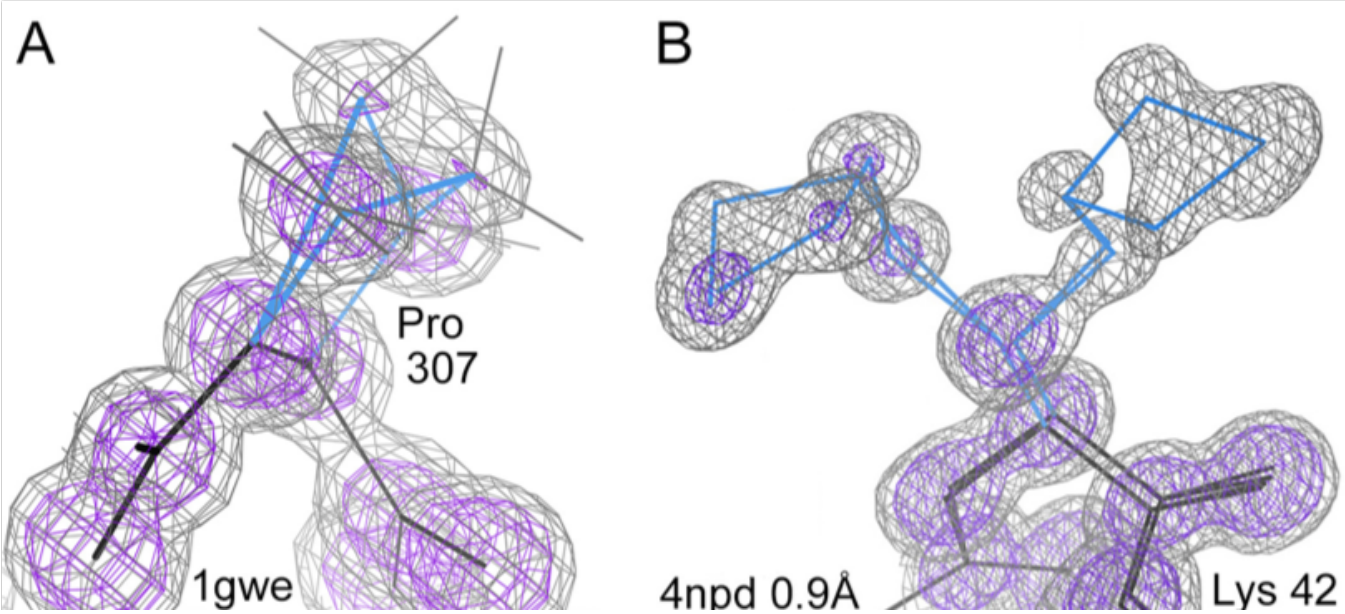

Figure 1: Some very high-resolution alternate conformations. A) A simple alternate for proline ring pucker, involving just two moving atoms, in 1gwe at 0.88Å (Murshudov 2002). B) Four hierarchical alternates of a lysine, two on each of two backbone alternates, in 4npd (Deis 2014). Contours are at 1.2σ (gray) and 3σ (purple).

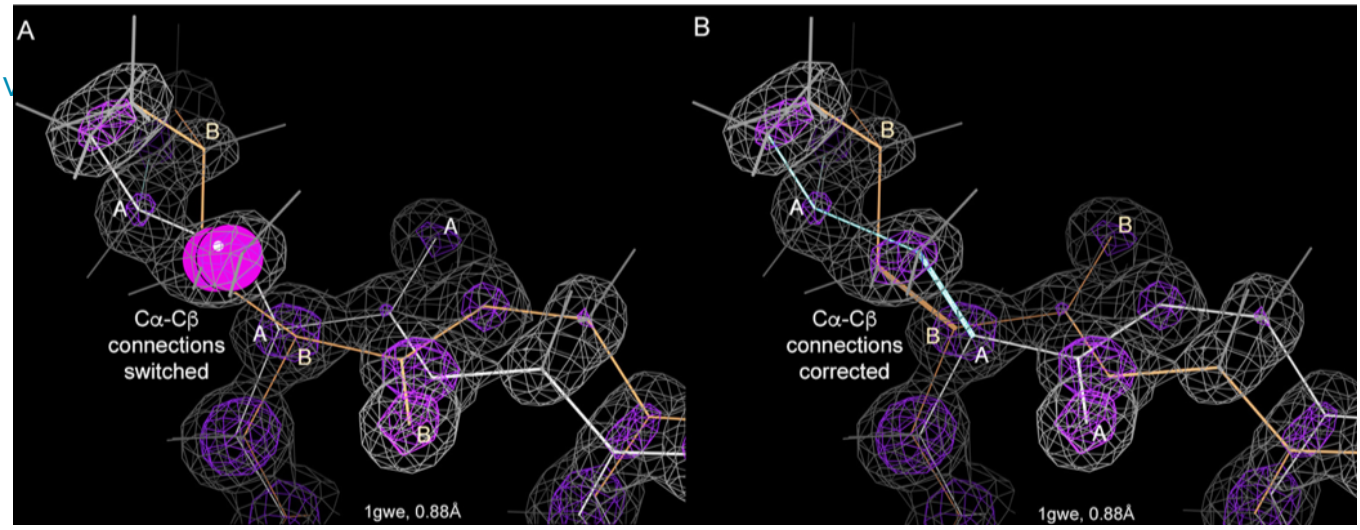

Figure 2: An example of internally inconsistent modeling of alternate conformations. A) Original PDB entry for Leu 105 of 1gwe, where Cβ deviation geometry outliers (magenta balls) flag the switched connectivities between the two alternates for the Cα-Cβ bonds. B) Alternate identities for backbone and Cβ atoms have been corrected, producing internally correct alternates and good geometry.

peaks, which is done wrong in the PDB entry (Fig. 2a). For the Cγ atoms on the sidechain, however, alt A is assigned correctly to the denser peak. This is one clue to the fact that the Cα and Cβ atoms in this residue have been connected across alternates rather than within them.

The other, even better clue is the severe Cβ deviation outliers (large magenta balls on the Cβ atoms); they combine the effects of multiple bad bond angles at the Cα (Lovell 2003). Wrong connections are also often flagged by bond length outliers unless refinement restraints are too tight. Wrong connections change dihedral angles and thus sometimes produce rotamer or Ramachandran outliers, but not in this case. In Fig 2b the alternate identities of the atoms have been corrected for all of the backbone and for Cβ, making each alternate internally consistent and removing the geometry outliers.

Consistency with surroundings

Consistency with the surroundings means that the individual alternates have no severe all-atom clashes with neighboring parts of the model (other parts of the macromolecules, het groups, or waters), and also that their polar atoms successfully make H-bonds with the nearby partners. As seen in Figure 3a, the backbone of 1gwe Leu 105 fails at that, since its alt A carbonyl O clashes with the nearby water that is much too close to H-bond. The close pair of waters are only 1.2Å apart, but have been modeled as HOH 181 and 182 instead of as alternates. If the waters are assigned as alternates, with the stronger one as alt A, and the backbone alternates are switched as in Figure 2B, then it all works consistently. Not only is there no clash in Figure 3b, but the alt A water can H-bond with the alt A backbone N (white) and the alt B water with the alt B backbone carbonyl O (peach) at excellent distances.

Most often, consistent interactions with the surroundings necessitates defining alternates for neighboring structure where the differences are too small to have been fitted originally, but are necessary to make a consistent ensemble. The small differences needed can often be confirmed and fitted using difference density peaks. In Figure 4A below, alt B of 1gwe Asp 410 is somewhat too close to its own backbone carbonyl O, and difference peaks suggest alternate orientations of the peptide, provided in the refit of Figure 4B.

Even more complicated interactions of alternate conformations arise, such as interacting cooperative networks of alternates, as seen in the Spa-N protein of 4npd (Deis 2014), or clashing alternates across a symmetry axis that cannot be

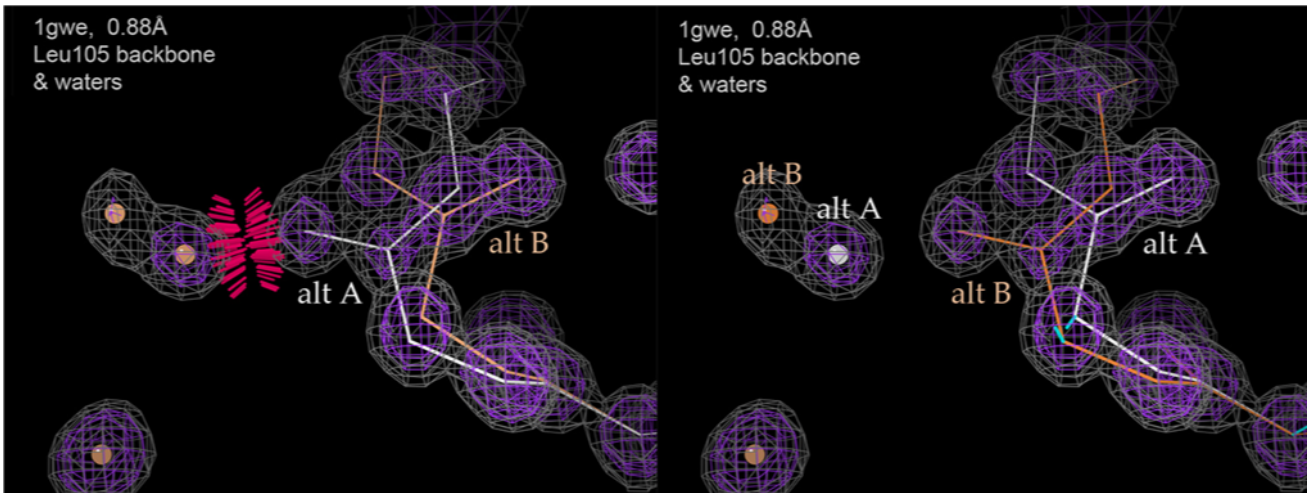

Figure 3: A case of backbone alternates inconsistent with neighboring waters. A) 1gwe Leu 105 alt A backbone clashing with a non-alternate water. B) With alternates assigned for the two close water peaks and the backbone alternates switched as in Fig, 2B above, now the protein alternates are consistent with the neighboring waters, forming two H-bonds rather than a clash.

expressed correctly without changing the spacegroup, such as shown in Figure 5. 1gwe is a homo-tetramer, but not quite perfectly symmetric, as demonstrated by the thorough overlap of Tyr 378 in what has to be labeled as the same alternate from neighboring molecules, but must actually be instantiated in each individual contact as one of each alternate. Note also that the two clashes at lower right mean alternates should be defined for that peptide, as also suggested by the very elongated density for the carbonyl O.

Don't stop too soon at alternate ends

Default treatment in nearly all software allows definition of alternates either for a sidechain ending at Cβ, which leaves the Cα as a single atom,

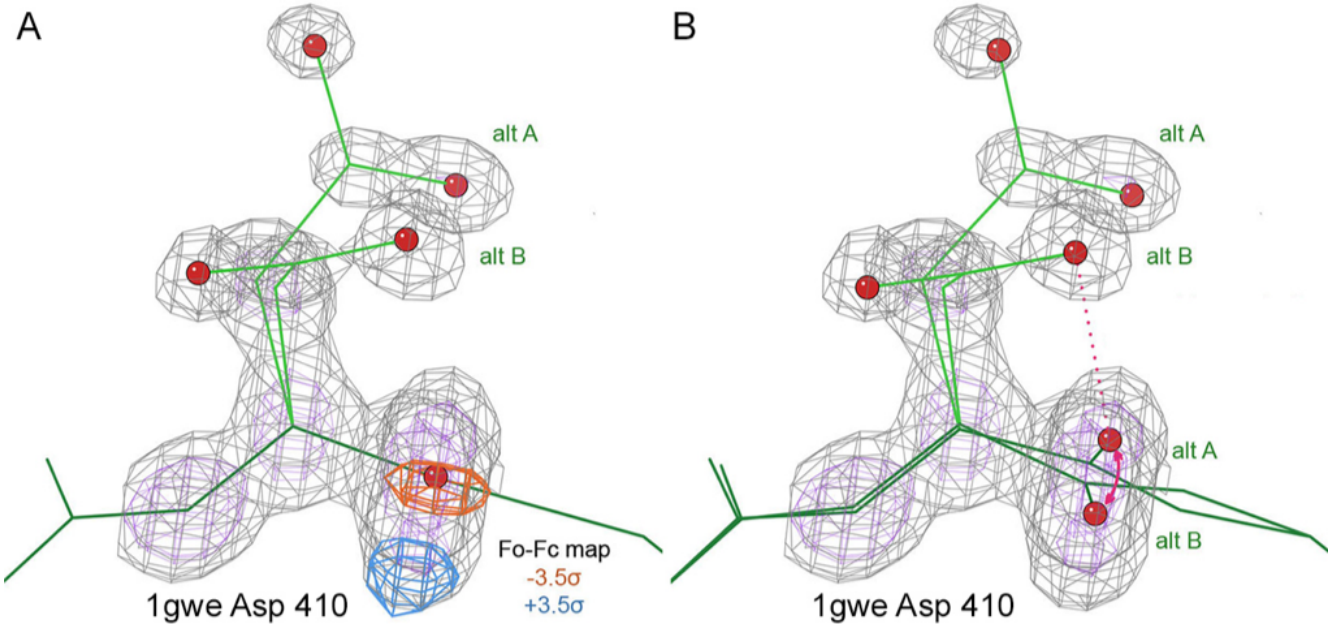

Figure 4: Sidechain alternates that need a backbone shift as well. A) Alternates for the sidechain of Asp 410 in 1gwe, with + (blue) and – (orange) difference density peaks that suggest the need for a small orientation shift of the peptide backbone. B) Providing those backbone alternates fits the 2Fo-Fc density better and removes the difference peaks.

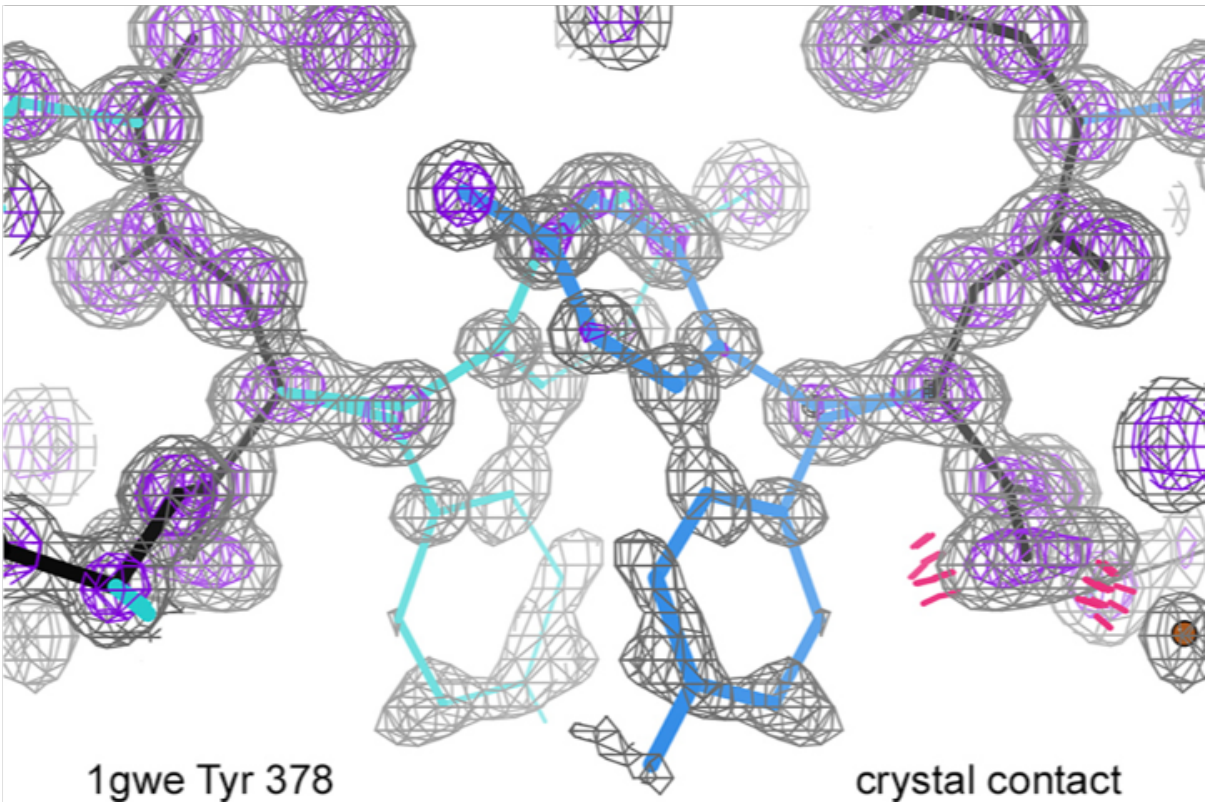

Figure 5: An alternate-conformation relationship that breaks the space-group definition across a 2-fold symmetry axis. The overlap of the alt A Tyr sidechains is so interpenetrated that the viewer thinks they are covalently bonded and does not show their clashes.

or for full residues ending at the near side of the peptide bonds, which leaves several atoms in the planar peptide as singletons. Refinement cannot possibly fix the resulting bond angle outliers, which would require separate positions of at least that next atom. These reasonable-sounding and easy-to-program definitions guarantee bad geometry unless the last alternate atom pairs are very close together (less than 0.2Å or preferably 0.15Å apart). Examples of this effect are shown in Figure 6A for a protein case and in Figure 6B for a nucleic acid case. The resulting distortions are often quite extreme, mostly in bond angles, but occasionally in bond lengths, Cβ deviations, or omega values.

Fortunately, this problem is fairly straightforward to fix. Make copies of the next atom or atoms to extend the definition of alternates, and let refinement make the geometry acceptable. Phenix has a utility that makes that process very easy. Some version of this should always be done when validation shows geometry outliers at the ends of alternates.

Backrubs for good backbone in single-residue alternates

When the problem is a sidechain alternate ended at Cβ, the bond-angle outliers often show as Cβ deviations, as in the example of Figure 7A, since those see combined effects of all the bond angles around the Cα atom. Such a problem clearly requires definition of alternates for the Cα atom, but unless the Cβ alternates are very close together, it actually requires spreading the alternates along the backbone up to single Cα atoms at n-1 and n+1. That can be done, and the smooth motions defined, by a backrub motion (Davis 2006) between the two conformations. The backrub is a hinge motion of the entire residue around the axis between Cα atoms at n-1 and n+1,

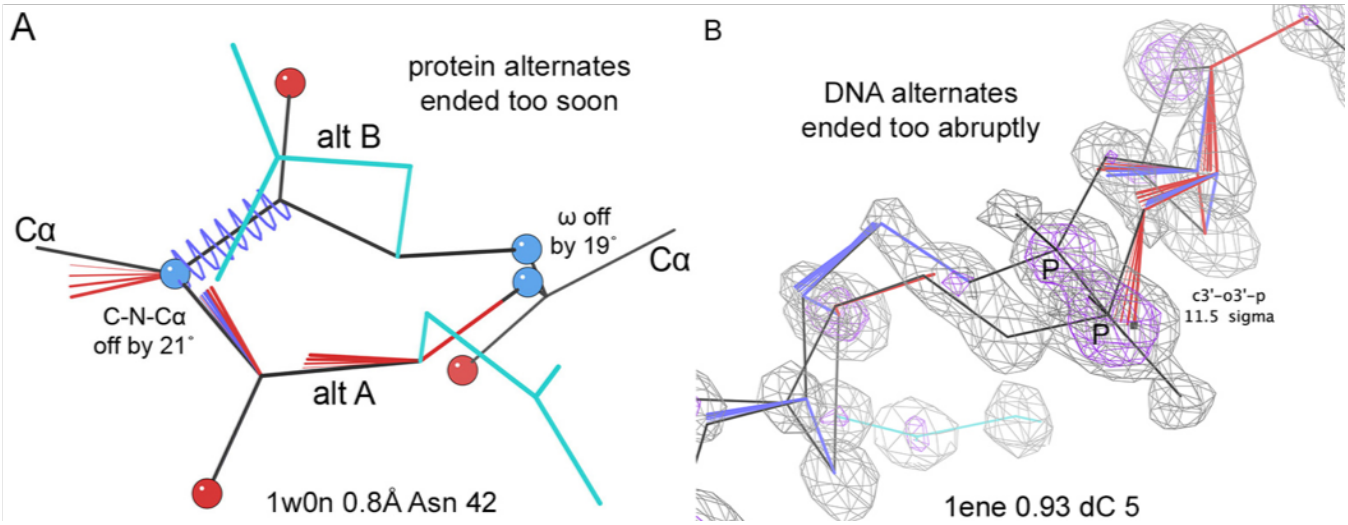

Figure 6: Two examples of the very common problem of serious geometry outliers where alternate conformations are ended too soon. A) Widely separated sidechain alternates of Asn 42 in 1w0n (Jamal 2004), with bond angle outliers up to 21°. B) Backbone DNA alternates at dC 5 of 1ene (Chiu 2000), with bond angle outliers up to 11.5σ (red and blue fans).

with minor rotations of the individual peptides to maintain their orientations and H-bonding, as seen for the correction in Figure 7B. The new backbone conformations instantiate the subtler changes that are strongly implied by the two clearly-resolved sets of sidechain-atom positions. Backrubs can be done in Coot (Emsley 2010) or in KiNG (Chen 2009) graphics and modeling programs, or can be approximated by duplicating the intervening atoms, shifting one slightly in the right direction, and refining.

Alternates must often be added when "waters" clash with nonpolar atoms

Density peaks fit as HOH but that clash with polar atoms are often actually ions or larger het groups. Those that clash with nonpolar atoms have sometimes displaced an atom of the macromolecule, but at high resolution they are often part of an unmodeled alternate conformation (Prisant 2020). In Figure 3d of that paper, an HOH that clashes with both S atoms and a Cβ of disulfide 91-186 in 3ajd at 1.27Å (Kuratani 2010) is actually one sulfur of an unmodeled alternate conformation of the disulfide. In Figure 1 of Headd 2013, reproduced in Figure 8 below, two waters in pear-shaped density clash with the nonpolar Cβ H atoms of Asp 9 in 1eb6 at 1.0Å. Panel 8B shows that they are actually the Oδ atoms of an alternate sidechain conformation in the most favored **m-20** rotamer of Asp, with the pear shapes pointing to the Cγ between them. A small backrub motion makes the fit perfect.

The bottom line

Each individual alternate conformation needs to be validated for rotamer and Ramachandran outliers, and especially for bond length and angle outliers, Cβ deviation outliers, and all-atom clashes. Then those problems need to be fixed, which sometimes involves correcting connectivity among the alternates, but which usually requires adding more alternates or more alternate atoms. This Fit Tip explains methods that help in making those corrections, such as defining atoms as duplicated even when their peaks are not separated, or using backrub motions for smooth transitions along the backbone. Almost all these things involve "spread" of alternates: new alternates, more extensive alternates, or more alternate atoms.

Although currently most alternate conformations in crystal structures are modeled manually, the software qFit (Riley 2021) seeks to automate this

process, yielding multi-conformer models with a parsimonious set of 1-4 conformations for each residue. qFit remains in active development, and will benefit from ongoing codification of the lessons reported here, based on detailed inspection and remodeling of specific local examples.

[Note that although many of the examples shown here are from fairly early structures, these

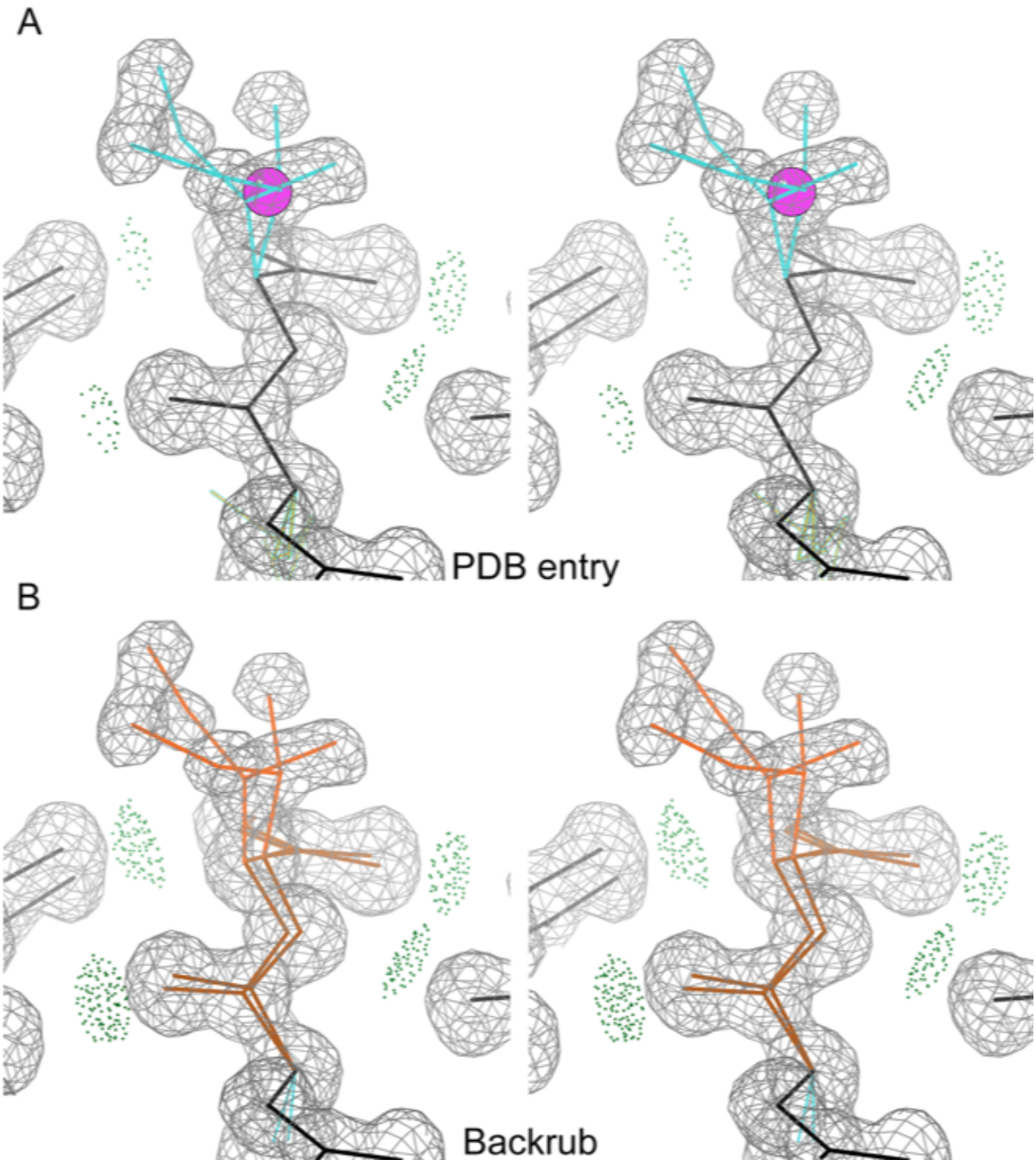

Figure 7: Stereo images of using a backrub motion to correct serious geometry outliers from not propagating sidechain alternates into the backbone. A) Clear alternate rotamers for the sidechain of Ile 47 in 1n9b at 0.9Å (Nukaga 2003), with a bad Cβ deviation outlier (magenta ball). B) Producing valid geometry by spreading the definition of alternates into the backbone and separating them with a small backrub motion between Cαs n-1 to n+1.

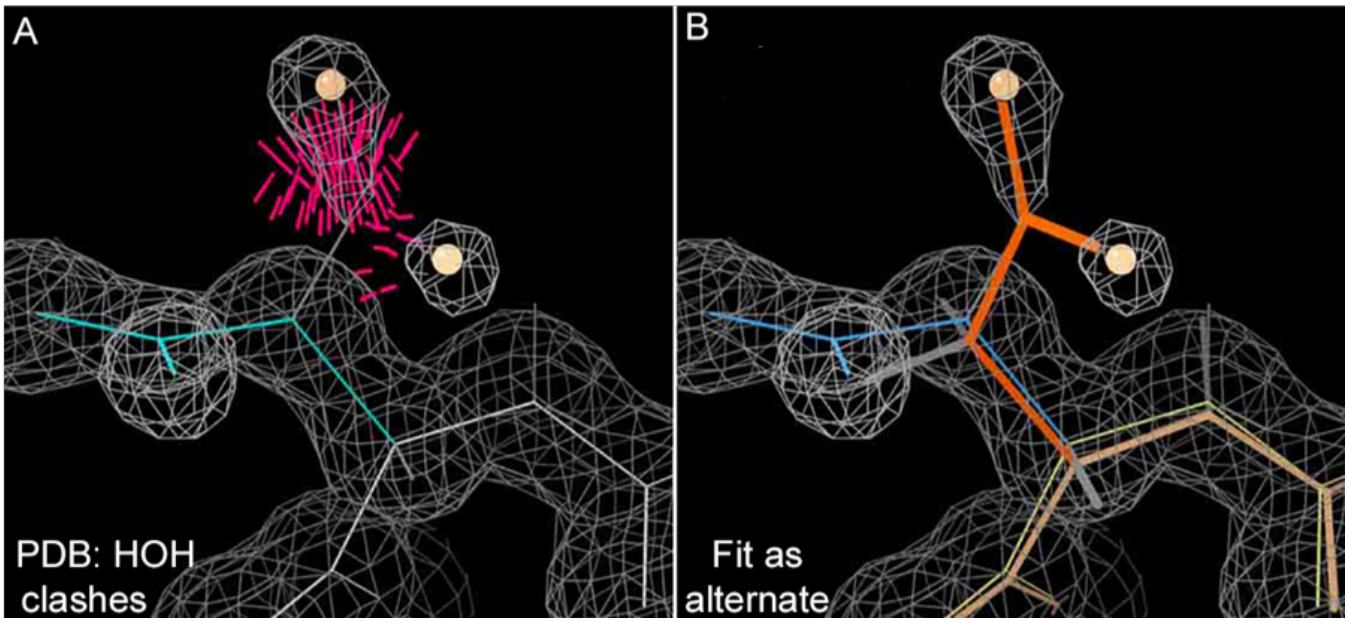

Figure 8: Stereo images of density peaks fit as HOH that clash with nonpolar atoms but are actually an unmodeled alternate conformation. A) Asp 9 and clashing waters as fit in 1eb6 (Mcauley 2001). B) The clashes are absent when those peaks are fit as an alternate conformation of the Asp.

alternate-conformation problems are still very common in current structures.]

Although currently most alternate conformations in crystal structures are modeled manually, the software qFit (Riley 2021) seeks to automate this process, yielding multi-conformer models with a parsimonious set of 1-4 conformations for each residue. qFit remains in active development, and will benefit from ongoing codification of the lessons reported here, based on detailed inspection and remodeling of specific local examples.

References:

Chen VB, Davis IW, Richardson DC (2009) KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program, *Protein Sci* **18**: 2403-2409

Chiu TK, Dickerson RE (2000) Crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium, *J Mol Biol* **301**: 915-945

Davis IW, Arendall WB III, Richardson DC, Richardson JS (2006) The backrub motion: How protein backbone shrugs when a sidechain dances, *Structure* **14**: 265-274

Deis LN, Pemble CW, Qi Y, Hagarman A, Richardson DC, Richardson JS, Oas TG (2014) Multiscale conformational heterogeneity in staphylococcal protein a: a possible determinant of functional plasticity, *Structure* **22**: 1467-1477

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot, *Acta Crystallogr* **D68**: 486-501

Fuhrmann CN, Kelch BA, Ota N, Agard DA (2004) The 0.83Å resolution crystal structure of alpha-lytic protease reveals the detailed structure of the active site and identifies a source of conformational strain, *J Mol Biol* **338**: 999-1013

Headd JJ, Richardson J (2013) "Fitting Tips #5: What's with water?", *Comput Cryst Newsletter* **4**: 2-5

Jamal S, Boraston AB, Turkenburg JP, Tarbouriech N, Ducros VM-A, Davies GJ (2004) *Ab initio* structure determination and functional characterization of Cbm36: A new family of calcium-dependent carbohydrate binding modules, *Structure* **12**: 1177-1187

Kuratani M, Hirano M, Goto-Ito S, (2010) Crystal structure of *Methanococcus jannashii* Trn4 complexed with sinefungin, *J Mol Biol* **401**: 323-333

Lovell SC Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure Validation by Cα Geometry: ϕ,ψ and Cβ Deviation, *Proteins: Struct Funct Genet* **50**: 437-450

Mcauley KE, Jia-Xing Y, Dodson EJ, Lehmbeck J, Ostergaard PR, Wilson KS (2001 A quick solution: *Ab initio* structure determination of a 19 kDa metalloproteinase using Acorn, *Acta Crystallogr* **D57**: 1571-1578

Murshudov GN, Grebenko AAI, Brannigan JA, Antson AA, Barynin VV, Dodson GG, Dauter Z, Wilson KS, Melik-Adamyan WR (2002) The structures of Micrococcus lysodeikticus catalase, its ferryl intermediate (compound II) and Nadph complex, *Acta Crystallogr* **D58**: 1972-1982

Nukaga M, Mayama K, Hujer AM, Bonomo RA, Knox JR (2003) Ultrahigh resolution structure of a class A beta-lactamase: On the mechanism and specificity of the extended-spectrum SHV-2 enzyme, *J Mol Biol* **328**: 289-301

Prisant MG, Williams CJ, Chen VB, Richardson JS, Richardson DC (2020) New tools in MolProbity validation: CaBLAM for cryoEM backbone, UnDowser to rethink "waters", and NGL Viewer to recapture online 3D graphics, *Protein Sci* **29**: 315-329

Riley BT, Wankowicz SA, de Oliveira SHP, van Zundert GCP, Hogan DW, Fraser JS, Keedy DA, van den Bedem H (2021) qFit 3: Protein and ligand multi-conformer modeling for X-ray crystallographic and single-particle cryo-EM density maps, *Protein Sci* **30**: 270-285

FAQ

crystallographic software users can be submitted to the editor at any time for consideration. Submission of text by email or word-processing files using the CCN templates is requested. The CCN is not a formal publication and the authors retain full copyright on their contributions. The articles reproduced here may