Correcting cis-trans-transgressions in macromolecular structure models

Franz Waibl¹, Klaus R. Liedl¹ and Bernhard Rupp²,³

¹ Department of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innsbruck, Austria
² k.-k.Hofkristallamt, San Diego, CA, USA
³ Institute of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria

Keywords
cis-trans isomers; cryo-EM; macromolecular models; stereochemical restraints; unsaturated fatty acids; X-ray crystallography

Correspondence
B. Rupp, k.-k. Hofkristallamt, San Diego, CA 92084, USA; Institute of Genetic Epidemiology, Medical University of Innsbruck, A-6020, Austria
Tel: +1(925) 209-7429, +43(676) 571-0536
E-mails: br@hofkristallamt.org, bernhard.rupp@i-med.ac.at

(Received 26 February 2021, revised 1 April 2021, accepted 16 April 2021)
doi:10.1111/febs.15884

Introduction

Alkene cis-trans isomers in crystallographic and cryo-EM models

Models of biomolecular structures deposited in the PDB [1] are the primary source for 3-D conformation and stereochemistry of proteins and nucleic acids. The structures are predominantly determined by X-ray crystallography, nuclear magnetic resonance spectroscopy or cryo-EM, and in all cases, during reciprocal or real space model refinement, the models require application of stereochemical restraints [2] to assure compliance with known stereochemistry. Owing to the limited number of different residues or nucleotides as building blocks of the proteins and nucleic acids, the macromolecular components of the models are subject to reasonably reliable validation, which reduces severe geometry errors and assures plausible stereochemistry of the macromolecular component of the models [3–5]. The macromolecular geometry restraint target values and variance have been generated from accurate stereochemical data in the CSD [6] and assembled into dictionaries for proteins [7,8] and nucleic acids [9–11].

Small molecule ligands

Most macromolecular structure models contain additional small molecule components (over 25 000

Abbreviations
cryo-EM, cryo-electron microscopy; CSD, Cambridge Structural Database; e.s.d., estimated standard deviation; LCP, lipidic cubic phase; MAG, monoacylglycerate; PDB, Protein Data Bank.
currently exist in the PDB monomer (HET) library, and for each of them, a correct stereochemical restraint file is required for model building, refinement and validation [12,13]. The validation process of the small molecule model stereochemistry against the expectation values in the restraint files depends on the accurate description and interpretation of each individual molecular entity, and errors in ligand models frequently go undetected, cf. reviews by [14,15]. A surprisingly frequent problem is the violation of cis-trans stereochemistry in C-C= C-C alkene moieties.

Relevance
Carbon–carbon alkene double bonds are present in numerous naturally occurring small molecule compounds such as unsaturated fatty acids, lipids or related compounds which have either (predominantly) cis- or trans geometry [16]. Typical examples are unsaturated fatty acids such as palmitoleic acid (PAM) important as a physiologically relevant substrate in cellular signalling [17–19], or mono-olein (OLB, OLC) used in LCP crystallization of membrane proteins [20]. Membrane proteins are also frequently crystallized with specifically designed MAGs and other lipids to protect their hydrophobic transmembrane domains. Lipid components that are present either natively or added during purification of membrane molecule complexes are frequently modelled in both X-ray- and cryo-EM-structure models.

Consequences
The energy barrier for the cis-trans isomerisation in alkenes is substantial. Rotation around the C=C double bond of an alkene requires breaking of the double π-bond of about 240 kJ mol⁻¹ bond energy [21], and in vivo, cis-trans isomerisation generally is catalysed by a respective cis-trans isomerase. The refinement programs do not automatically remedy incorrect cis-trans assignment (cf. Discussion). Without proper restraints and given ambiguous experimental electron density or electrostatic potential maps, the ligand models can become severely distorted during refinement.

Given that cis-trans isomers have different stereochemistry and thus distinctly different physical properties [22], the following points should be observed: for each cis- or trans isomer, a unique three-letter hetero-monomer identifier (HET) needs to be associated with the monomer restraint file (HET.cif) completely and correctly defining the proper isomer; the same three-letter HET identifier cannot be used to describe both isomers; a given HET isomer can only be modelled in its assigned cis or trans configuration, not in both; validation programs should clearly flag deviations from the expected cis- or trans geometry to allow the depositor to intercept the error already prior to or during deposition, or at least inform the user in the final validation report that a seriously improbable configuration is present.

Examination of models containing commonly occurring HET moieties with alkene double bonds has revealed problems in the above categories (a-d) ranging from severe errors potentially affecting biological conclusions to inconsistent presentation in search results and nuisance errors in modelling of these HET ligands.

Results
Cis-trans transgressions
Out of 24 743 small molecule HET monomers in the PDB, 187 were cis-only, 286 trans-only and 46 monomers contained both cis- and trans bonds. We examined the most frequent molecules, such as unsaturated fatty acids, unsaturated MAGs and related entities. Nomenclature and correct cis- and trans configuration were cross-checked for consistency with PubChem [23].

Common unsaturated fatty acids
PAM
The PDB contains 79 PAM ((9Z)-hexadec-9-enoic acid, PubChem CID 445638) molecules in 34 entries, 70% of which have rather improbable double-bond configuration (Fig. 1, cf. Data Availability). PAM has a C8-C9=C10-C11 cis double bond (torsion angle \( \tau = 0° \)) and occurs as a free ligand in fatty acid transporters or covalently bound, for example, to acylated serine residues in Wnt proteins that play a significant role in cellular signalling [17–19]. Incorrect C8-C9=C10-C11 torsion angles is concerning beyond a nuisance in cases where PAM carries biological relevance, as in 6c0b [24] (180°) or 40a [18] (92°). The respective publications do not address the unphysical bond distortions in the structure models.

Free PAM is frequently presented in cryo-EM models with almost random torsion angles to predominantly trans-like orientations (6ueb) [25] (cf. Data Availability). Even when modelled only as a component of a
fatty acid membrane collar, completely implausible—that is, plainly wrong—configurations of double bonds could be avoided by use of proper restraints in model refinement (cf. Discussion).

OL, EL

Oleic acid (OLA), \((9Z)\)-octadec-9-enoic acid, is found in 230 PDB entries, and of 470 molecules where the torsion angle could be calculated (in some instances, only truncated fragments are present), about half have implausible torsion angles (Fig. 2A). There are significant differences between individual entries, with some having all cis torsions correctly assigned, while others show varying cis/trans mixtures up to almost exclusively trans configurations (cf. Data Availability). Its trans-stereoisomer, elaidic acid (ELA), \((9E)\)-octadec-9-enoic acid, is modelled in only 18 instances, of which two thirds have implausible torsion angles (Fig. 2B).

VCA

Vaccenic acid is inconsistently annotated in the PDBe as \(E\) and \(Z\) isoforms (cf. section Chemical library presentation). Out of 21 instances in multiple entries, 20 are modelled as the rare cis-vaccenic acid (asclepic acid, \((11Z)\)-octadec-11-enoic acid) while one is modelled with an implausible torsion angle of \(-58^\circ\).

Monoacylglycerates

OL, OLC

A frequently modelled group of compounds with possibility for cis-trans transgression are MAGs. Mono-olein \((1-(cis-9-octadecenoyl)-rac-glycerol)\) is used in LCP crystallization [20,26,27]. Due to the chiral centre at \(C2\), two enantiomer restraint descriptions exist, OLB, \((2S)\)-2,3-dihydroxypropyl \((9Z)\)-octadec-9-enoate; and OLC, \((2R)\)-2,3-dihydroxypropyl \((9Z)\)-octadec-9-enoate. One redundant and mostly incorrectly modelled recent addition, A6L [26,27], exists as well.

In contrast to the almost 50% incorrect assignments for oleic acid (OLA), MAGs in general seem to be modelled less frequently with incorrect configuration. The distributions for both mono-olein \((9Z)\) enantiomers are almost bimodal, with about 35% OLB and 23% OLC in implausible, largely trans, configuration (Fig. 2C,D).

Designer MAGs, 78M, 78N

Various designed MAGs [28] are used in LCP crystallization [29]. 78M \((2S)\)-2,3-dihydroxypropyl \((7Z)\)-pentadec-7-enoate) and its \(C2\) enantiomer 78N \((2R)\)-2,3-dihydroxypropyl \((7Z)\)-pentadec-7-enoate) may serve as an example of a designer MAG. The MAG 78M (Fig. 3B) introduced later than OLB and OLC still has about 12% of trans-outliers (Fig. 3B), but already considerably less than the \((9Z)\) mono-oleins OLB and OLC (Fig. 2C,D). Finally, all 78N (enantiomers of 78M) are modelled in the correct cis isoform, which may be the result of increasing awareness of proper lipid stereochemistry in the lipidic phase user community, and indicative that modellers concerned about selecting the proper enantiomer likely also pay better attention to the correct cis isomer assignment. The decreasing number of torsion angles between \(\sim 0^\circ\) and \(\sim 180^\circ\) indicates that the planarity restraints are more consistently applied.

SIP

Another biologically relevant trans-monoacylglycerate derivative is sphingosine 1-phosphate, SIP, modelled in 5ksi and 5ksj [30]. In multiple instances, the same molecule is present in split conformations, where each partial model except one violates the expected trans configuration and, notably, with scarce electron...
density support even for the few atoms of the additional S1P fragments modelled. No battery of sophisticated nonparsimonious modelling and poor electron density fit can justify these—easily correctable—configuration errors. In two other PDB entries, S1P is modelled correctly (cf. Data Availability).

**Fig. 2.** The distribution of double-bond torsion angles in C18:1 unsaturated fatty acids and monoacylglycerates in PDB entries. Oleic acid (OLA, cis), elaidic acid (ELA, trans) and the two OLA-derived C2 enantiomers of monoacylglycerate monolein, OLB (2S) and OLC (2R). About 35% of OLA, 67% of ELA, 35% of OLB and 23% of OLC are incorrectly modelled.

**Fig. 3.** Torsion angle distributions of unsaturated double bonds. The restraint file for SPH is inconsistent and specifies a cis bond contradicting the true trans geometry. Twelve per cent of 78M are modelled in trans configuration, but its enantiomer 78N is correctly modelled in all 150 instances. Of about 12 000 C-C=C-C torsion angles in β-carotene molecules, most are modelled with correct trans configuration, with similar outlier tails for cryo-EM and crystallographic models.
The general difficulty to distinguish fatty acid acyl chains from polyethylene glycol chains based on electron density alone (which is likely a contributing factor to incorrect modelling) has been pointed out earlier [31,32] and needs to be addressed separately.

**Multiple isomers—carotenoids**

The class of carotenoids with their multiple conjugated isoprene double-bond chains present a separate challenge for modelling and validation. The core of the conjugated diene—CH=CH—CH=CH—bonds is almost planar due to electron delocalization, with the central bond length of \( \sim 1.44 \) Å intermediate between a single and double C-C bond.

Although \( \beta \)-carotene all-(\( E \)) isomers are the most abundant, some (\( Z \)) isomers such as (9\( Z \)), (3\( Z \)) or (15\( Z \)) \( \beta \)-carotene are found in nature [16]. Note that in the \( x \)-cis nomenclature for all-(\( E \)) \( \beta \)-carotene (BCR), the \( cis \) prefix here refers to the C6-C7 single bond between the ring and the isoprene chain and not to any double bonds of the \( \beta \)-carotene chain [33]. BCR is modelled in 144 X-ray and cryo-EM (membrane) protein structures as an integral component. While identification of each (legitimate) isomer with an individual HET might be considered impractical or overly cumbersome during modelling, each isomer is a different chemical entity with different physical properties, and as pointed out, individual HETs are required for consistent elimination of the ambiguities in cis-trans isomer modelling, refinement and validation.

About 2% of the \( \sim 12000 \) double bonds identified in BCR molecules in the PDB are modelled in incorrect cis configurations and the torsion angle distribution shows an extended tail of trans-torsions with improbable torsion angles (Fig. 3C,D). In almost all cases, the molecules disproportionally contributing to aberrant torsion angles are found in regions of poor to no electron density (X-ray) or electrostatic potential (cryo-EM) maps. Particularly, in these situations with weak experimental evidence, correct interpretation and application of the restraints by the modelling and refinement programs are required to keep the stereochemistry plausible. Similar observations hold for related carotenoids such as the plant xanthophylls lutein (LUT) and violaxanthin (XAT).

**Inconsistent dictionary presentation**

**Chemical library presentation**

A source of confusion for model builders and users that could be easily corrected during the ongoing remedia-
tion of the PDB HET libraries is the inconsistent annotation of HET compounds and their 2D presentation in the PDB chemical compound libraries (PDBChem, Fig. 4). As an example, for 1EX ((11\( Z \),13\( Z \))-hexadeca-
11,13-dien-1-ol) the SMILES and IUPAC name match, but contradict the PDBe images, while for ELD (elaidoylamide, (9E)-octadec-9-enamide), the correct IUPAC molecule name contradicts the cis-SMILES but matches the PDBe image; for ELA, the corresponding elaidic acid, all records match. Similarly, in the case of VCA (Fig. 4), the synonym name for the predominant trans-vaccenic acid, (11E)-octadec-11-enoic acid, contradicts the cis-SMILES and image for (11Z)-octadec-11-enoic acid, which could be contributing to all VCA molecules in PDB deposits being modelled in the cis configuration, although cis-vaccenic acid (asclepic acid) is the rare isomer.

In addition, discrepancies exist between the PDBeChem and RCSB ligand summary. 1EX is correct in RCSB PDB, the VCA 2D representation is correct in RCSB, but the 3D presentation is wrong, and the SMILES are unspecified. ELD SMILES are unspecified in the RCSB ligand summary.

The inconsistent chemical nomenclature in restraint files and description was highlighted [34], and remediation announced [35], almost a decade ago.

### Cis-trans annotation

Part of the confusion in presentation, modelling and restraint application originates from the inconsistent and hard to parse content in cif (crystallographic information file) format of unremediated legacy restraint files, which are then used by the refinement programs to set up the stereochromic restraints [36,37]. Cis-trans stereochromy can be annotated in multiple places in the HET.cif files. In addition to the SMILES strings [38], double bonds are identified as cis (Z, ‘zusammen’) or trans (E, ‘entgegen’) in the _chem_comp_bond section. Also the _pdbe_chem_comp_bond_depiiction section carries—although not necessarily consistently—cis and trans information in the ENDUPRIGHT(ENDDOWNRIGHT annotation. An example of inconsistent annotation is SPH ((2R,3S)-sphingosine), where the SMILES and ‘Z’ double-bond  annotation contradict the actual trans stereochemistry, which could be a contributing factor to ~ 75% incorrect modelling (Fig. 3A). However, the (2S,3R) enantiomer SQS is described correctly.

### Discussion

### Restraint implementation

#### Restraining the cis-trans configurations

The observed violations of cis-trans stereochemistry fall into two categories: (a) selection of the wrong stereoisomer while still maintaining planarity, or (b) completely incorrect stereochemistry where not even the planarity of the double bond resulting from the C sp² hybridization is preserved. Refinement programs, except REFMAC5 which minimizes the eigenvalue of the product moment matrix of the atoms in the plane [39], generally restrain the double-bond torsion angles by minimizing the normal distance residual of the four carbon atoms from their least-squares plane (Fig. 5), which automatically implies a perfect 0° or 180° target value. The planarity restraint can thus be fulfilled in either, cis- or trans configuration. If the model represents the incorrect isomer, the refinement programs will not flip the incorrect configuration that is trapped in a local minimum. Note that the deviation from the arbitrary LSQ plane is only a precision measure because the true location of that plane is not known.

Currently, the outcome of a rerefinement of a PDB model depends both on the choice of restraint

![Fig. 5. Basic planar and angle restraint implementation. (A) Example of the C8-C9=C10-C11 cis bond of PAM or OLA with an improbable torsion angle $\phi$ of 42° and its restraining least square plane. (B) The supporting additional 1-2-4 and 1-3-4 ‘bond’ angles are different (≈ 98° vs 155°) for the cis and trans dihedral angles, respectively, and can serve as ‘flipping’ restraints.](image-url)
implementation and on the restraint file version. As an example, we re-refined the PDB entry 4f0a (91.8° τ for PAM) with default settings in REFMAC5 [39] using the still distributed legacy PDB restraint file (resulting in τ of −105.0°). Using the Grade [40] restraint file that will be provided as a part of ongoing PDB remediation, resulted in correct cis configuration with τ of 0.25°. Refinement with PHENIX [41] in default mode (three macrocycles) using ReadySet!/eLBOW restraints results in a correct τ angle of 1.01°.

**Flipping the double bond**

Because the planarity restraints trap a 1-2=3-4 bond in a local minimum around either 0° or 180°, an additional restraint that exacts a penalty for incorrect isomer assignment is required. Additional tight 1-2-4 and 1-3-4 pseudo-angle restraints across the double bond, which differ for cis and trans, with a tight target e.s.d. of 1.0° are effective and suffice to flip incorrect models but are not enough to assure adequate planarity (Table 1). The addition of these pseudo-bond-angle restraints to the standard Grade restraint file (or successive application) is surprisingly robust both in refinement and in regularization. Table 1 summarizes the results of REFMAC5 restrained maximum-likelihood refinement of 5wiu [42] which contained a mixture of variably incorrect modelled OLA molecules.

The flipping restraints can also be applied before final refinement in idealization mode as shown in Table 3 in the Materials and methods section for 6peq. Both the sequential regularization and regularization with the modified Grade restraints flip all isomers into the correct cis configuration.

**Recommendations**

Examination of PDB entries and PDB validation reports revealed numerous problems largely resulting from neglecting the rules listed above, in section categories (a-d), ranging from severe errors potentially affecting biological conclusions to inconsistent presentation in search results and nuisance mistakes in modelling of these HETS. While most of these instances may be quite easily remediated or corrected by the depositors or the PDB, at present the user of such entries must validate the corresponding HETS in the context of the supported claim or intended use of the structure models.

**Users**

Trust but verify [14]. When a structure model including a ligand or compound is of interest or importance for a user, it is not sufficient to simply download the model from the PDB and to consult the validation report. Global metrics generally cannot inform about local deviations in a small part of the model such as the key ligand. The validation results for the ligand are located at the end of the validation report, and not all errors, particularly C-C=C-C cis-trans violations, are flagged. Inspection of the primary evidence in form of electron density (for example, with the 'Fetch PDB and map' options in COOT [43]) is highly recommended, because then with a
single click the fit of the model to the electron density (or its absence) can be displayed, the double bond inspected, and its torsion angle measured with the distance tool. Comparison of the model with the electron density (or its absence) often provides direct clues why a torsion angle is incorrectly modelled [15,44]. In case of discrepancy, the HET.cif file (links are provided in the PDB ChemExplorer download pages) should be examined for consistency with the expected stereochemistry. For cryo-EM structures, the electrostatic potential maps need to be downloaded separately, but the inspection process outlined above remains the same. Other options for graphical display programs exist, such as Chimera or PyMol [45,46].

Model builders

Be aware that the model refinement programs—both in reciprocal space for crystallography and in real space for cryo-EM—do not automatically fix incorrectly assigned cis-trans stereoisomers. If specified (which is not always the case), only planarity restraints are applied (cf. Fig. 3). The initial model then must be placed with proper stereochemistry. In case of doubt, consult the HET.cif file via the PDB ChemExplorer pages but be aware that inconsistencies such as shown in Fig. 4 do exist.

Not every spurious map density needs to be filled with a contorted cis-trans ligand, and in reciprocal space refinement, a poorly placed ligand actually makes the entire model worse [44]. Please keep in mind that: the same three-letter HET identifier cannot be used to describe both cis-trans isomers—one of them is simply wrong.

In addition, if chiral centres are present, assure the proper assignment of diastereomers (example OLB vs OLC).

Before inventing new HETs, check the HET libraries for existing equivalent entries.

Use the latest, updated PDB restraint file from the ongoing restraint file remediation effort, or a ligand restraint generation program such as phenix.elbow [47] or the Grade Server (http://grade.globalphasing.org) to produce a complete restraint file that includes adequate planarity information and check for proper isomer description. If desired, add the respective pseudo-angle restraints for flipping the configuration of incorrectly modelled isomers.

Even if a given small molecule ligand is not central to the research of the model builder, it should be modelled correctly. Later studies or large-scale data mining with a different focus may rely on that ligand model.

Proper restraint files are easy to obtain and massively improve the plausibility of the model.

Validation

The current ligand validation reports do not flag incorrect cis-trans isomer assignment as an error. Validation should clearly flag such transgressions and intercept them preferably already during the deposition process, where they can be easily corrected by the depositor. To correct a deposited model is much harder and requires a major remediation effort. Voluntary participation of the original depositors should be considered, by notifying them and encouraging redeposition of corrected models refined with the latest, remediated restraint files. Redeposition should become straightforward with the forthcoming improvements of the current versioning system [48] of PDB entries. In cases where major scientific claims would be affected, voluntary compliance is less likely, and correction of the public record by appropriate action [15] might be necessary. The restraint file modification we have described here are easy to automate and could be implemented in future remediation efforts until a generally accepted modification to the restraint files is implemented.

Inconsistencies in ChemExplorer as shown in Fig. 4, resulting probably from ambiguous, incomplete or contradictory information in the HET.cif restraint files should be addressed in future ligand validation efforts [12].

Materials and methods

Data scraping and analysis

The HET.cif restraint files available in the ccp4 7.1 monomer library (Winn et al., 2011) were scraped using SMILES strings [49] C/CC/C or C/CC/C, and C/CC/C or C/CC/C for any presence of cis and trans geometry, respectively. Out of 24743 HET monomers in the January 2021 ccp4 7.1 monomer library, 187 were cis-only, 286 trans-only and 46 monomers contained both cis- and trans C-C=C=C bonds. For each HET serving as an example, the HET cif files were scanned for double bonds, their location and status determined, and the torsion angles calculated (cf. Data Availability).

Estimate of target value variance

The CONQUEST program from the CSD suite of crystallographic software [50] was used to extract geometry data for unsubstituted alkenes. We searched for X-CH2-CH=CH2-X motifs, where X is any atom and none of...
| Ligand Number | Original deposit | Legacy cif | Grade cif | Flipping cif | Grade after flipping cif | Grade 0.02 + flipping combined | Grade 0.04 + flipping combined |
|---------------|-----------------|-----------|-----------|-------------|-------------------------|-------------------------------|-------------------------------|
| PAM A 903     | 110.2           | 169.6     | 0.6       | 29.3        | 1.7                     | 0.8                           | 4.2                           |
| PAM A 904     | 116.9           | 174.4     | 0.7       | 19.3        | 0.5                     | 0.3                           | 1.1                           |
| PAM A 905     | 154.7           | 178.5     | 180.0     | 11.8        | 0.1                     | 0.1                           | 0.2                           |
| PAM B 904     | −168.5          | −179.3    | −180.0    | 14.2        | 0.1                     | 0.3                           | 0.3                           |
| PAM B 905     | −162.4          | −176.8    | −179.7    | 3.6         | 0.4                     | 0.6                           |                               |
| PAM B 906     | −138.3          | −167.9    | −179.8    | −28.6       | −0.3                    | −0.3                          | −0.8                          |
| PAM F 202     | 78.8            | 3.5       | 0.4       | 20.7        | 0.4                     | 0.2                           | 0.8                           |
| PAM H 202     | 79.9            | −0.5      | −0.1      | 14.9        | −0.0                    | −0.0                          | −0.3                          |
| PAM C 904     | 106.7           | 177.5     | −1.6      | 25.8        | 0.3                     | −0.4                          | −1.7                          |
| PAM C 905     | 28.1            | 0.3       | 0.6       | 9.5         | 0.1                     | 0.1                           | 0.1                           |
| PAM C 906     | −178.3          | 179.1     | 179.9     | −6.7        | 0.3                     | 0.2                           | 0.7                           |
| PAM D 905     | 31.3            | 4.3       | 0.9       | 17.6        | 0.8                     | 0.7                           | 2.9                           |
| PAM D 906     | 170.7           | −179.8    | −180.0    | 25.9        | 0.2                     | −0.6                          | −2.5                          |
| PAM D 907     | 152.0           | 173.4     | 179.7     | 19.0        | 0.3                     | 0.2                           | 0.8                           |
| OLC A 902     | 102.2           | 178.0     | 0.6       | 28.4        | 0.2                     | 0.0                           | 0.5                           |
| OLC B 902     | 157.4           | −178.7    | −179.9    | 21.4        | −0.1                    | −0.2                          | −0.6                          |
| OLC B 903     | 138.1           | 177.6     | 179.9     | 22.2        | 0.1                     | 0.2                           | 1.1                           |
| OLC C 902     | 70.6            | 0.8       | 0.2       | 24.8        | 0.1                     | 0.1                           | 0.5                           |
| OLC C 903     | −25.2           | 1.0       | 0.3       | 13.1        | 0.3                     | 0.2                           | 1.3                           |
| OLC D 902     | −140.7          | −174.7    | −179.9    | −23.2       | 0.0                     | 0.1                           | 0.5                           |
| OLC D 903     | 93.3            | 178.0     | −0.5      | 14.3        | −0.7                    | −0.6                          | −2.4                          |
| OLC D 904     | −78.3           | 4.7       | 0.6       | −5.0        | 0.7                     | 0.5                           | 1.7                           |

**Example for creating a flipping restraint file**

Edit the Grade restraint file for PAM, insert pseudo-angle flipping restraint and optionally, relax planarity as necessary to avoid over-restraining (0.03 Å as a planarity target SD for ~2.0 Å resolution data, in this example).

- PAM C8 C9 C11 97.5 1.0
- PAM C8 C10 C11 97.5 1.0
- PAM csdf-01 C8 0.030
- PAM csdf-01 C9 0.030
- PAM csdf-01 C10 0.030

**Table 2.** Empirical outlier-corrected CSD restraint targets and their variances. Target values are ideal values. The target values for pseudo-angles and -distances are computed from the bond lengths and angles in the Grade restraint file for ELA (C=C: 1.32 Å, C=C-C single bond: 1.50 Å, C=C-C angle: 126.2°).

| restraint | Target (°) | Variance (°²) | SD (°) |
|-----------|------------|---------------|--------|
| cis-1234 torsion | 0.0        | 5.6          | 2.4    |
| trans-1234 torsion | 180.0      | 10.4         | 3.2    |
| cis-124/134 pseudo-angle | 97.5     | 3.3          | 1.8    |
| trans-124/134 pseudo-angle | 154.9     | 2.7           | 1.6    |
| cis-14 pseudo-distance | 3.09 Å    | 0.003 Å²     | 0.06 Å |
| trans-14 pseudo-distance | 3.92 Å    | 0.0005 Å²    | 0.02 Å |

the carbons is part of a ring system. Powder diffraction structures, as well as structures with an R-value higher than 0.075, and structures containing atoms heavier than Ca were excluded. We obtained 152 hits and queried the torsion angle, the 1-2-4 and the 1-3-4 pseudo-bond (flipping) angle, and the 1-4 distance. Data were separated into cis (torsion < 90°) and trans (torsion > 90°) configurations.

For the analysis of torsion angles, all hits with a torsion angle of exactly 0° or exactly 180° were excluded since those exact values occur when the torsion is crystallographically constrained. The resulting torsion distributions are depicted in the background of Figure 1 (blue and orange histograms).

Outlier detection using a modified Z-score test [11,51] was applied and variances from restraint target values for all nonoutlier data points were calculated and summarized in Table 2.

**Table 3.** Regularization of a cryo-EM structure with flipping restraints. REFMAC5 regularization with the legacy cif improves only few torsions, while regularization with the Grade restraints groups into cis and trans isomers around the respective planarity target values. If flipping restraints are included in the Grade cif (or applied before the Grade planarity restraints), all monomers are corrected. Relaxing the pane restraints from leads to a looser geometry. Colour code: red, more than 40° off target, yellow 40-10° off target, green within 10° off target.
Example for application of flipping restraints for regularization

Table 3 provides the result of the regularization via flipping restraints for a cryo-EM model.

Alternate flipping procedure with 1-4 distance restraint

As an alternate approach, the use of a single 1-4 distance restraint (3.09 Å for cis, 3.92 Å for trans) was explored. Starting from incorrect trans configurations, the 1-4 distance restraint works primarily against the other bond restraints while achieving little towards flipping the torsion angle. While successive application of this single restraint and then standard Grade restraints can achieve correct results, single-step refinement with a 1-4 distance restraint combined with planarity restraints did not result in correct cis-trans configurations. The two pseudo-angle restraints accomplish this readily, but for the single 1-4 distance restraint we were unable to find a reliable single-step weighting scheme.

Another possibly preferable configuration-specific restraint would be a hard torsion angle restraint with a period of 1 with an empirical variance. Such torsion restraints are currently provided with target of 180°, a degenerate periodicity of 2 and practically infinite variance and thus not implemented. With proper implementation, unique torsion angles could be an alternative to the cis-trans-degenerate planarity restraints.

Acknowledgements

BR acknowledges partial funding through FWF (Austrian Science Foundation) grants P32821 and P28395 and KRL through P30565. We thank Clemens Vonrhein, Global Phasing Limited, for suggestions how to scrape the PDB monomer libraries, and Mariusz Jaskolski, University of Poznan, for advice on chemical restraint implementation.

Conflict of interest

The authors declare no conflicts of interest.

Data Accessibility

For each example and additional HETs, relevant C-C=C-C angle numbering:
https://www.dropbox.com/sh/0ofhwj1isuj86qy/AAAB9hHcjvtxhgNGm5NAF7ffa?dl=0

For each example and additional HETs, a complete listing of torsion angles: https://www.dropbox.com/sh/hoapghdqnrcpceh/AAChEeq9j0KHBJgGv37RH3za?dl=0

Author contributions

FW mined and analysed data; KRL supervised work by FW; BR conceived and supervised the study, scraped and analysed data, and prepared the manuscript.

References

1. wwwPDB_consortium (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucl Acids Res 47, D520–D528.
2. Konnert J (1976) A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. Acta Crystallogr A 32, 614–617.
3. Yang H, Peisach E, Westbrook JD, Young J, Berman HM & Burley SK (2016) DCC: a Swiss army knife for structure factor analysis and validation. J Appl Crystallogr 49, 1081–1084.
4. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N et al. (2012) Outcome of the first electron microscopy validation task force meeting. Structure 20, 205–214.
5. Read RJ, Adams PD, Arendahl WB 3rd, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lutteke T, Otwinowski Z et al. (2011) A new generation of crystallographic validation tools for the protein data bank. Structure 19, 1395–1412.
6. Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr B 58, 380–388.
7. Engh RA & Huber R (2001) Structure quality and target parameters. In International Tables for Crystallography F (Rossmann MG & Arnold E eds), pp. 382–392.Kluwer Academic Publishers, Dordrecht/Boston/London.
8. Moriarty NW, Tronrud DE, Adams PD & Karplus PA (2016) A new default restraint library for the protein backbone in Phenix: a conformation-dependent geometry goes mainstream. Acta Crystallogr D 72, 176–179.
9. Kowiel M, Brzezinski D & Jaskolski M (2016) Conformation-dependent restraints for polynucleotides: I. Clustering of the geometry of the phosphodiester group. Nucl Acids Res 44, 8479–8489.
10. Gilski M, Zhao J, Kowiel M, Brzezinski D, Turner DH & Jaskolski M (2019) Accurate geometrical restraints for Watson-Crick base pairs. Acta Crystallogr B 75, 235–245.
11 Kowiel M, Brzezinski D, Gilski M & Jaskolski M (2020) Conformation-dependent restraints for polynucleotides: the sugar moiety. Nucl Acids Res 48, 962–973.
12 Smart OS, Horsky V, Gore S, Svobodova Varekova R, Bendova V, Kleywegt GJ & Velankar S (2018) Validation of ligands in macromolecular structures determined by X-ray crystallography. Acta Crystallogr D74, 228–236.
13 Adams PD, Aertgeerts K, Bauer C, Bell JA, Berman HM, Bhat TN, Blaney JM, Bolten E, Bricegne G, Brown D et al. (2016) Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop. Structure 24, 502–508.
14 Deller MC & Rupp B (2015) Models of protein–ligand crystal structures: trust, but verify. J Comp Aided Mol Des 29, 817–836.
15 Wlodawer A, Dauter Z, Porebski PJ, Minor W, Stanfield R, Jaskolski M, Pozharski E, Weichenberger CX & Rupp B (2018) Detect, correct, retract: How to manage incorrect structural models. FEBS J 285, 444–466.
16 Meléndez-Martínez AJ, Mapelli-Brahm P, Hornero-Méndez D & Vicario IM (2019) Structures, Nomenclature and General Chemistry of Carotenoids and their Esters in Carotenoid Esters in Foods: Physical, Chemical and Biological Properties. pp. 1–50, The Royal Society of Chemistry.
17 Naschberger A, Orry A, Lechner S, Bowler MW, Nurizzo D, Novokmet M, Keller MA, Oemer G, Seppi D, Hasbeck M et al. (2017) Structural evidence for a role of the multiple-functional human glycoprotein Afamin in Wnt transport. Structure 25, 1907–1915.
18 Janda CY, Waghary D, Levin AM, Thomas C & Garcia KC (2012) Structural basis of Wnt recognition by frizzled. Science 337, 59–64.
19 Nile AH, Mukund S, Stanger K, Wang W & Hannoush RN (2017) Unsaturated fatty acyl recognition by Frizzled receptors mediates dimerization upon Wnt ligand binding. Proc Nat Acad Sci USA 114, 4147–4152.
20 Caffrey M (2015) A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Crystallogr F71, 3–18.
21 Ouellette RJ & Rawn JD (2018) Alkenes: Structures and Properties. In Organic Chemistry (Ouellette RJ & Rawn JD eds), 2nd edn. pp. 135–165, Academic Press, Cambridge, MA.
22 Caffrey M (1987) Kinetics and mechanism of transitions involving the lamellar, cubic, inverted hexagonal, and fluid isotropic phases of hydrated monoacylglycerides monitored by time-resolved X-ray diffraction. Biochemistry 26, 6349–6363.
23 Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B et al. (2021) PubChem in 2021: new data content and improved web interfaces. Nucl Acids Res 49, D1388–D1395.
24 Chen P, Tao L, Wang T, Zhang J, He A, Lam K-H, Liu Z, He X, Perry K, Dong M et al. (2018) Structural basis for recognition of frizzled proteins by Clostridium difficile toxin B. Science 360, 664–669.
25 Nakagawa T (2019) Structures of the AMPA receptor in complex with its auxiliary subunit corinichon. Science 366, 1259–1263.
26 Wang X, Lyu Y, Ji Y, Sun Z & Zhou X (2021) Substrate binding in the bile acid transporter ASBTYf from Yersinia frederiksenii. Acta Crystallogr D77, 117–125.
27 Wang X, Lyu Y, Ji Y, Sun Z & Zhou X (2021) An engineered disulfide bridge traps and validates an outward-facing conformation in a bile acid transporter. Acta Crystallogr D77, 108–116.
28 Misquitta LV, Misquitta Y, Cherezov V, Slattery O, Mohan JM, Hart D, Zhalnina M, Cramer WA & Caffrey M (2004) Membrane protein crystallization in lipidic mesophases with Tailored Bilayers. Structure 12, 2113–2124.
29 Li D, Lee J & Caffrey M (2011) Crystallizing membrane proteins in lipidic mesophases. A Host Lipid Screen. Crystal Growth Des 11, 530–537.
30 Sun K, D’Alessandro A, Ahmed MH, Zhang Y, Song A, Ko T-P, Nemkov T, Reisz JA, Wu H, Adebiyi M et al. (2017) Structural and functional insight of sphingosine 1-phosphate-mediated pathogenic metabolic reprogramming in sickle cell disease. Sci Rep 7, 15281.
31 Dym O, Song W, Felder C, Roth E, Shnyrov V, Ashani Y, Xu Y, Joosten RP, Weiner L, Sussman JL et al. (2016) The impact of crystallization conditions on structure-based drug design: a case study on the methylene blue/acyethylcholinesterase complex. Protein Sci 25, 1096–1114.
32 Naschberger A, Juyoux P, von Velsen J, Rupp B & Bowler MW (2019) Controlled dehydration, structural flexibility, and Gadolinium MRI contrast compound binding in human plasma glycoprotein afamin. Acta Crystallogr D75, 1071–1083.
33 Schlücker S, Szeghalmi A, Schmitt M, Popp J & Kiefer W (2003) Density functional and vibrational spectroscopic analysis of β-carotene. J Raman Spectroscopy 34, 413–419.
34 Jaskolski M (2013) On the propagation of errors. Acta Crystallogr D69, 1865–1866.
35 Berman H, Kleywegt GJ, Nakamura H & Markley JL (2013) Comment on On the propagation of errors by Jaskolski (2013). Acta Crystallogr D 69, 2297.
36 Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F & Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr D60, 2184–2195.
Cis-trans transgressions in structure models

F. Waibl et al.

37 Evans PR (2007) An introduction to stereochemical restraints. Acta Crystallogr D63, 58–61.
38 Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28, 31–36.
39 Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F & Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D67, 355–367.
40 Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C & Bricogne G (2011) Better ligand representation in BUSTER protein–complex structure determination. Acta Crystallogr A 67, C134.
41 Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung L-W, Jain S, McCoy AJ et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D75, 861–877.
42 Wang S, Wacker D, Levit A, Che T, Betz RM, McCory JD, Venkatakrishnan AJ, Huang X-P, Dror RO, Shoichet BK et al. (2017) D4 dopamine receptor high-resolution structures enable the discovery of selective agonists. Science 358, 381–386.
43 Debreczeni JE & Emsley P (2012) Handling ligands with Coot. Acta Crystallogr D68, 425–430.
44 Pozharski E, Weichenberger CX & Rupp B (2013) Techniques, tools and best practices for ligand electron-density analysis and results from their application to deposited crystal structures. Acta Crystallogr D69, 150–167.
45 Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH & Ferrin TE (2018) UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci 27, 14–25.
46 DeLano WL (2008) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA.
47 Moriarty NW, Janowski PA, Swails JM, Nguyen H, Richardson JS, Case DA & Adams PD (2020) Improved chemistry restraints for crystallographic refinement by integrating the Amber force field into Phenix. Acta Crystallogr D76, 51–62.
48 Zardecki C, Young J, Westbrook J, Feng Z, Peisach E, Persikova I, Liang Y, Gutmanas A, Berrisford J, Ikegawa Y et al. (2019) Enabling depositor-initiated PDB coordinate replacement through file versioning. Acta Crystallogr A 75, 25.
49 Weininger D, Weininger A & Weininger JL (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf Comput Sci 29, 97–101.
50 Bruno IJ, Cole JC, Edgington PR, Kessler M, Macrae CF, McCabe P, Pearson J & Taylor R (2002) New software for searching the Cambridge Structural Database and visualizing crystal structures. Acta Crystallogr B 58, 389–397.
51 Crosby T (1994) How to detect and handle outliers. Technometrics 36, 315–316.