

Standards for Crystallographic Publishing

Manfred S. Weiss

Team Leader in Macromolecular Crystallography

EMBL Hamburg Outstation, Hamburg, Germany

Co-Editor of Acta Crystallographica Sections D & F



Why This Workshop?



Aug 2005: IUCr Journal
Commission Meeting
in Montecatini, Italy.

Jun 2006: First NMR structure
published in an IUCr
journal ^{a)}.



^{a)} Betsy L. Lytle *et al.* (2006). Solution structure of *Arabidopsis thaliana* protein At5g39720.1, a member of the AIG2-like protein family. *Acta Cryst.* **F62**, 490-493.

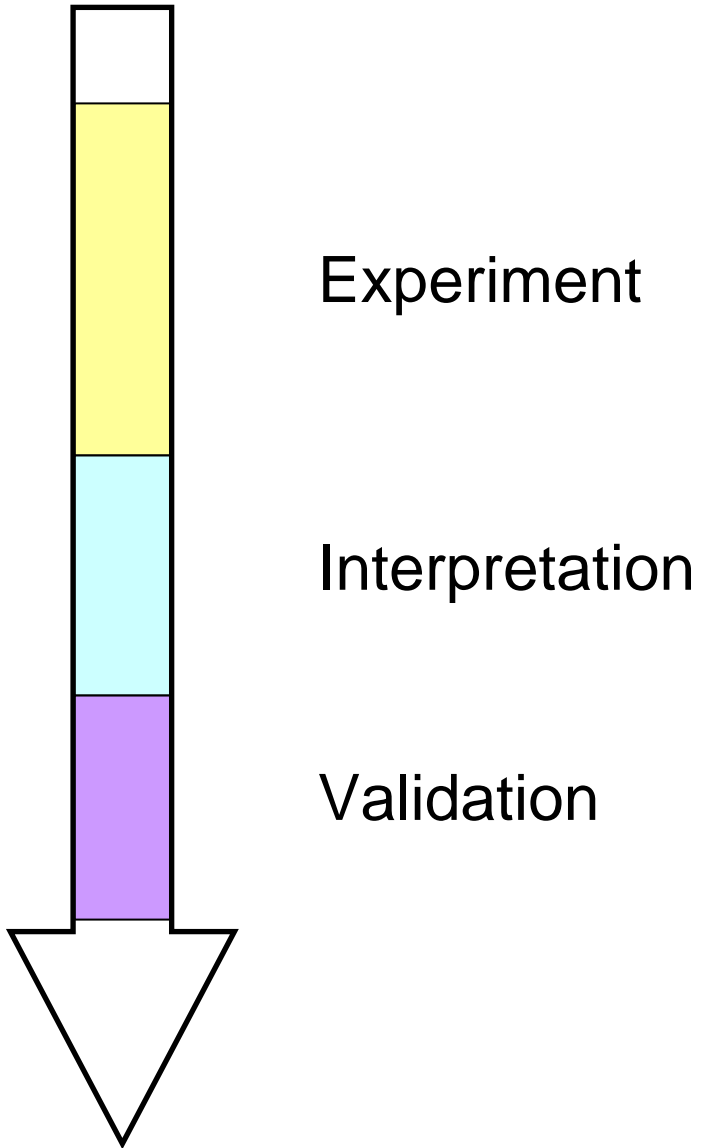
Why Standards?

- To facilitate judgement of the veracity of results
- To ensure archival access to results

Questions in Structure Publishing

- What is the identity and quality of the sample ?
- How well is the experiment carried out ?
- How is the experiment interpreted ?
- What is the quality of the resulting model ?
- Which data are made available to the community ?

In some more detail ...



In some more detail ...

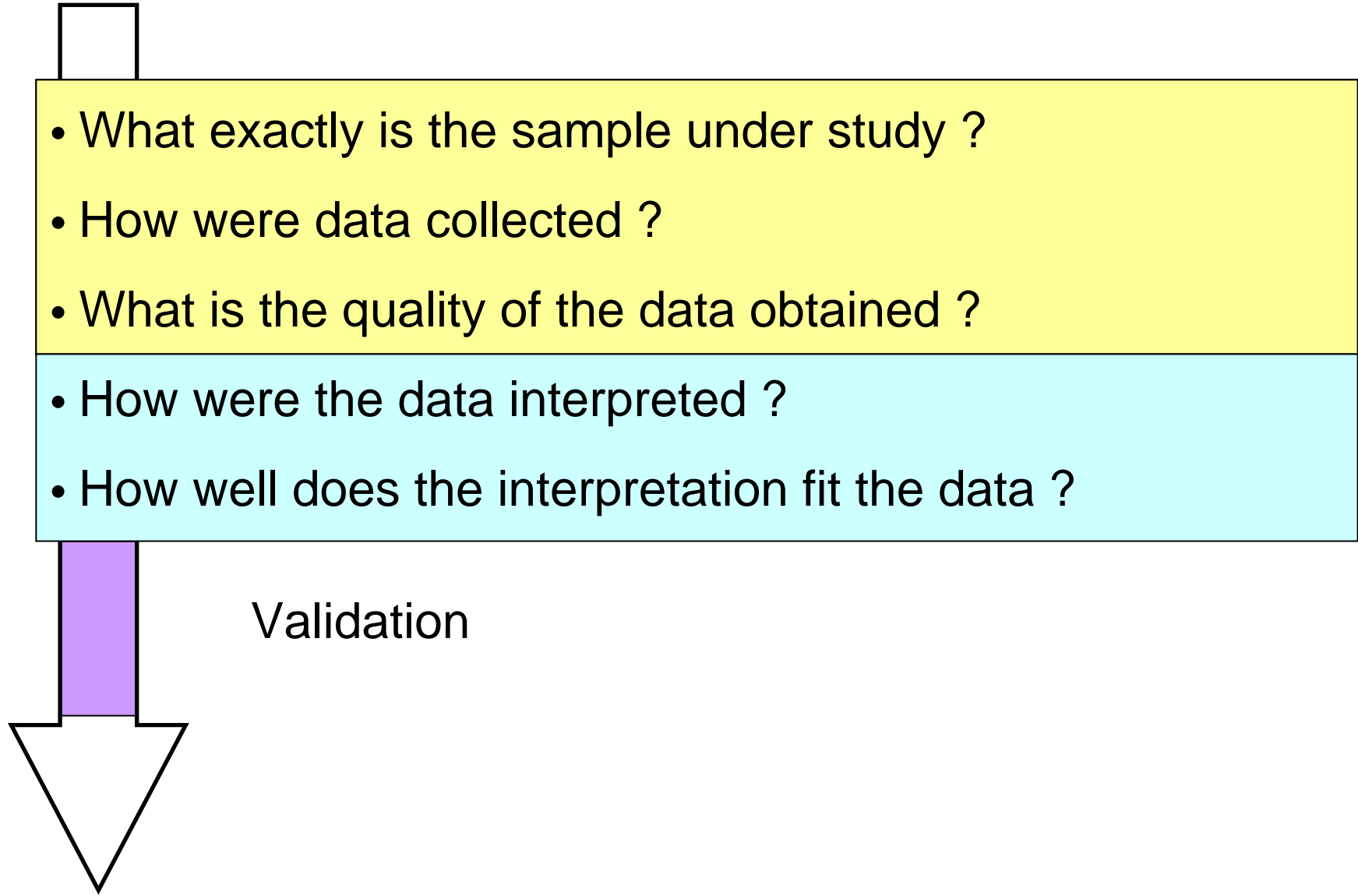
- What exactly is the sample under study ?
- How were data collected ?
- What is the quality of the data obtained ?



Interpretation

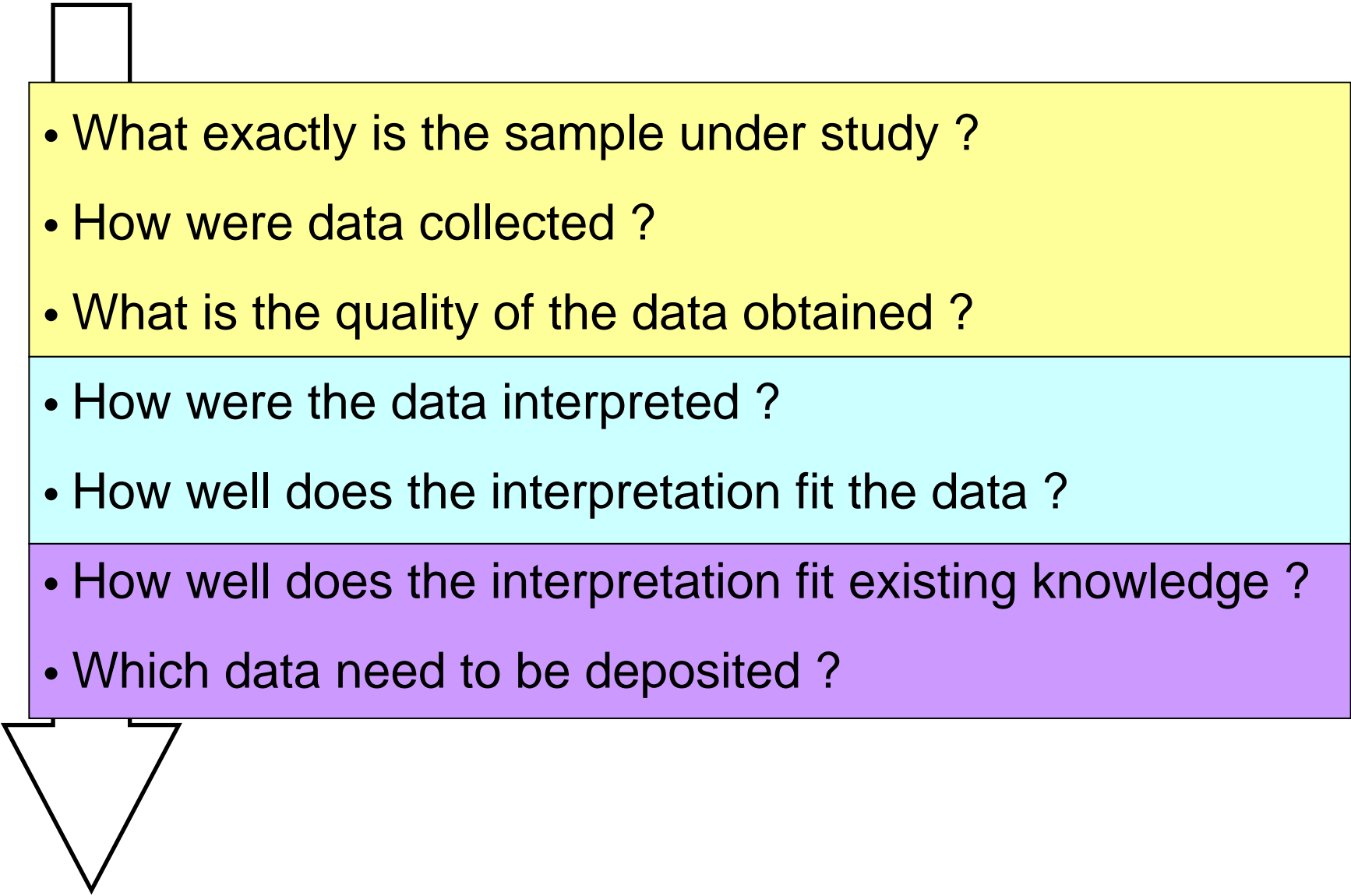
Validation

In some more detail ...

- 
- What exactly is the sample under study ?
 - How were data collected ?
 - What is the quality of the data obtained ?
 - How were the data interpreted ?
 - How well does the interpretation fit the data ?

Validation

In some more detail ...

- 
- What exactly is the sample under study ?
 - How were data collected ?
 - What is the quality of the data obtained ?
 - How were the data interpreted ?
 - How well does the interpretation fit the data ?
 - How well does the interpretation fit existing knowledge ?
 - Which data need to be deposited ?

The Reality in Acta Cryst. D & F

Notes for Authors §11, Evaluation criteria

Data Collection

Effective resolution

R_{merge}

Multiplicity / Redundancy

$I/\sigma(I)$

Completeness of data

Example 1: Acta F

Table 1

Data-collection and processing statistics.

Values in parentheses correspond to the highest resolution shell.

No. of crystals	1
Beamline	X13
Wavelength (Å)	0.8031
Temperature (K)	100
Crystal-to-detector distance (mm)	180
Rotation range per image (°)	1.0
Total rotation range (°)	156
Space group	$P2_12_12$
Unit-cell parameters (Å)	$a = 87.42, b = 180.65, c = 35.11$
Mosaicity (°)	0.60
Resolution limits (Å)	50.0–1.93 (2.00–1.93)
Total No. of reflections	268981
Unique reflections	43027
Redundancy	6.3
$I/\sigma(I)$	23.3 (3.7)
Completeness (%)	99.9 (99.8)
R_{merge} (%)	7.6 (49.8)
$R_{\text{r.i.m.}}$ (%)	8.3 (54.8)
$R_{\text{p.i.m.}}$ (%)	3.3 (22.5)
Overall B factor from Wilson plot (Å ²)	24.2
Optical resolution (Å)	1.55

Some Problems

- Lack of clear definitions, e.g. for effective resolution.
- Inconsistent criteria, e.g. R_{merge} (better R-values are around but not routinely calculated by the popular data reduction programs).
- Standard deviations may be calculated differently by different programs.

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_{i,hkl} - \bar{I}_{hkl}|}{\sum_{hkl} \sum_i I_{i,hkl}}$$

Some Problems

- Lack of clear definitions, e.g. for effective resolution.
- Inconsistent criteria, e.g. R_{merge} (better R-values are around but not routinely calculated by the popular data reduction programs).
- Standard deviations may be calculated differently by different programs.

$$R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i | I_{i,\text{hkl}} - \bar{I}_{\text{hkl}} |}{\sum_{\text{hkl}} \sum_i | I_{i,\text{hkl}} |}$$

$$R_{\text{r.i.m.}} = \frac{\sum_{\text{hkl}} (N/(N-1))^{1/2} \sum_i | I_{i,\text{hkl}} - \bar{I}_{\text{hkl}} |}{\sum_{\text{hkl}} \sum_i | I_{i,\text{hkl}} |} = R_{\text{meas}}$$

$$R_{\text{p.i.m.}} = \frac{\sum_{\text{hkl}} (1/(N-1))^{1/2} \sum_i | I_{i,\text{hkl}} - \bar{I}_{\text{hkl}} |}{\sum_{\text{hkl}} \sum_i | I_{i,\text{hkl}} |}$$

Some More Problems

- No real standards for sample description, macromolecule production and crystallization.

The Reality in Acta Cryst. D & F

Notes for Authors §11, Evaluation criteria

Data Collection

Effective resolution

R_{merge}

Multiplicity / Redundancy

$I/\sigma(I)$

Completeness of data

Refinement

R , R_{free}

Resolution limits

Number of atoms

Restraints

Example 2: PNAS

Table 1. Refinement and model statistics for the two hARH3 crystal forms

	P2 ₁ 2 ₁ 2 ₁	P2 ₁
Refinement		
Resolution limits, Å	40.0–1.60	30.0–2.05
No. of reflections		
Working set	42,579	42,868
Test set	834	852
<i>R</i> _{cryst} , %	17.5	18.8
<i>R</i> _{free} , %	20.9	23.1
No. of atoms		
Protein	2,613	5,186
Ions	2	4
Water	240	147
Stereochemistry		
R.m.s. deviations		
Bonds, Å	0.012	0.023
Angles, °	1.34	1.93
Average <i>B</i> factors		
Protein, Å ²	24.3	48.6
Ions, Å ²	11.4	28.2
Water, Å ²	29.7	45.3
Ramachandran plot		
Most favored, %	92.9	93.1
Allowed, %	6.4	6.1

Some More Problems

- No real standards for sample description, macromolecule production and crystallization.
- No real standards for structure solution statistics
- Without refinement history, it is impossible to judge, whether the structure has been refined to convergence or whether it has been over-refined.

The Reality in Acta Cryst. D & F

Notes for Authors §11, Evaluation criteria

Data Collection

Effective resolution

R_{merge}

Multiplicity / Redundancy

$I/\sigma(I)$

Completeness of data

Refinement

R , R_{free}

Resolution limits

Number of atoms

Restraints

Validation

Ramachandran plot

PDB validation report

Some More Problems

- No real standards for sample description, macromolecule production and crystallization.
- No real standards for structure solution statistics
- Without refinement history, it is impossible to judge, whether the structure has been refined to convergence or whether it has been over-refined.
- Only a subset of all available validation approaches is applied.

ActaBioStandards

<http://www.iucr.ac.uk/iucr-top/lists/actabiostandards/>

Actabiostandards

- Unmoderated, private list for members of the *Acta Cryst. D* and *F* Editorial Boards.
- Subscribers to the list are strictly limited to the members of the relevant Boards, those IUCr officials (*viz.* President, General Secretary and Treasurer, and Executive Secretary) who request membership of the list, and observers selected by the Chair.
- Chair: Howard Einspahr, hmeinspahr@yahoo.com

List of Recommended Items

<http://journals.iucr.org/d/services/evaluationcriteria/>

<http://journals.iucr.org/f/services/structuralcommunications/requirements.html>

1. Sample information
 - 1.1. Macromolecule and source information
 - 1.2. Macromolecule production
 - 1.3. Crystallization
 - 1.4. Crystal data
2. Data collection and structure solution statistics
 - 2.1. Data collection, refinement data set
 - 2.2. Phasing
 - 2.2.1. MAD/SAD data and structure solution statistics
 - 2.2.2. MIR/MIRAS/SIR/SIRAS data and structure solution statistics
 - 2.2.3. Molecular replacement data and structure solution statistics
3. Model generation and refinement
4. Model validation

List of Recommended Items

<http://journals.iucr.org/f/services/structuralcommunications/requirements.html>

Even More Problems

- Author compliance.
- Some journals don't provide the necessary space for publishing all relevant information.
- More and more journals publish the relevant information as supplementary material only.

Summary

- Standards for crystallographic publishing exist, but need to be updated and developed further.
- There is significant overlap between standards for crystallographic publishing and standards for NMR.
- Acta editors and co-editors take the standards issue very seriously.
- The situation is somewhat different for other journals.