SFCHECK

Documentation

SFCHECK is a program for assessing the agreement between an atomic model and X-ray data.

The program requires one or two input files, one with the coordinates of the model (in PDB, CIF or BLANC format) and one with the structure factors (in PDB, CIF or BLANC format). It runs completely automatically.

Non-crystallographic symmetry information is extracted from the coordinates file and is used when necessary to restore the content of the asymmetric part of the unit cell.

I. Output information produced by SFCHECK

SFCHECK produces a PostScript output file with the following set of information:

1. Crystal

cell parameters and space group

2. Model

number of atoms

number of water molecules

solvent content

<B> for model

reported resolution

reported R-factor

3. Refinement

refinement program

resolution range for refinement

reported sigma cut-off for refinement

reported R-factor

reported Rfree

4. Structure factors

number of reflections

number of reflections with I > σ

number of reflections with I > 3σ

resolution range

completeness

R-standard (Σσ(F)/ΣF)

Wilson plot (amplitude *vs* resolution)

overall B-factor by Patterson origin peak and by Wilson plot

optical resolution

expected minimal error in coordinates

5. Model *vs* structure factors

R-factor

Correlation coefficient

R-factor for reported resolution range and sigma cut-off

Rfree

Luzzati plot (R-factor *vs* resolution)

coordinate error from Luzzati plot

expected maximal error in coordinates

DPI

Optical resolution

Optical resolution is defined as an expected minimum distance between two resolved peaks in the electron density map. It is computed as $\Delta^2_{patt}\big/\sqrt{2}$, where Δ_{patt} is the width of the Patterson origin peak. The "Expected optical resolution for complete data set" is calculated as above but using all reflections; for missing reflections, the average value in the corresponding resolution shell is used.

The plot of "Optical resolution for an atom with B=0" demonstrates the behaviour of the part of the "Optical resolution" corresponding to the series termination.

Low resolution cut-off

Disordered solvent contributes to the diffraction at low resolution. However, removing low resolution data from calculations results in a series termination effect which is noticeable in the electron density at the surface of the molecule. To reduce the influence of the low resolutions terms, **SFCHECK** applies the *soft* low resolution cut-off to structure factors according to the formula:

$$F_{new} = F_{old} \cdot \left(1 - e^{-B_{off} \cdot s^2}\right)$$

where

$$B_{off} = 4 \cdot d_{\max}^2$$

The program uses $B_{off} = 256$.

Patterson scaling

Scaling in **SFCHECK** is based on the Patterson origin peak which is approximated as a gaussian. Compared to the conventional scaling by the Wilson plot, this method is particularly advantageous when only low resolution data are available. The program gives overall B-factors estimated with both methods.

The program scales F_{calc} to F_{obs} by the Patterson origin peak using all the data and applying B_{off}.

It first computes $B_{overall}$'s for observed and calculated amplitudes. Next, it makes the width of the calculated peak equal to the observed one, i.e. computes an additional thermal factor B_{add}. Finally, it computes the scale factor for F_{calc}.

$$F_{calc\_scaled} = F_{calc} \cdot scale \cdot e^{-B_{add} \cdot s^2}$$

where

$$B_{add} = B_{overall\_obs} - B_{overall\_calc}$$

$$scale = \sqrt{\frac{\sum\left(F_{obs}^2 \cdot \left(1 - e^{-B_{off} \cdot s^2}\right)\right)}{\sum\left(F_{calc}^2 \cdot e^{-B_{add} \cdot s^2} \cdot \left(1 - e^{-B_{off} \cdot s^2}\right)\right)}}$$

The program computes the *R-factor* and the *Correlation coefficient* for the reported resolution range and sigma cut-off without applying B_{off}. If the F_{obs} file contains reflections marked with the *Rfree* flag, the program computes *Rfree*.

Completeness

For the calculation of the map, missing data are restored by using the average value of the intensities for the corresponding resolution shell. The program produces a plot of completeness *vs* resolution and a plot of the average radial completeness in the polar coordinates θ and ϕ.

Expected minimal error

The minimal coordinate error is estimated using the experimental $\sigma(F)$. The standard deviation of the atomic coordinates is

$$\sigma(x) = \frac{\sigma(slope)}{curvature}$$

where

σ(slope) is the slope of electron density in the *x* direction

curvature is the average curvature of the electron density in the atomic peak centrer

It is computed as:

$$\sigma(slope) = \frac{2\pi \cdot \sqrt{\sum\left(h^2 \cdot \sigma(F)^2\right)}}{Volume\_cell\ .\ A}$$

$$curvature = (2\pi^2 \Sigma(h^2 F)) \ / \ (Volume\_cell\ .\ A^2)$$

where A is the cell parameter, h - the Miller index, the sum over all reflections

(Cruickshank, D.W.J. (1949) Acta crys. **2**, 65)
(Murshudov et al., (1997) Acta crys. **D53**,240-255)

If there is no experimental sigma for the observed data, the program uses $\sigma = 0.04 \cdot F_{obs}$ for all the reflections.

Expected maximal error

The expected maximal error in the coordinates is estimated by the difference between F_{obs} and F_{calc} :

$$\sigma(slope) = \frac{2\pi^2 \cdot \sqrt{\sum\left(h^2 \cdot (F_{obs} - F_{calc})^2\right)}}{Volume\_cell\ .A^2}$$

For the missing reflections the program uses the average value of *σ(F)* for the corresponding resolution shell instead of $(F_{obs}-F_{calc})$.

DPI - Diffraction-data Precision Indicator

DPI is the Cruickshank's estimation of the coordinate error (*The Refinement of Macromolecular Structure*, Proccedings of CCP4 Study Week-End, pp11-22, 1996).

$$\sigma(x) = \sqrt{\frac{N_{atoms}}{N_{obs} - 4 \cdot N_{atoms}}} \cdot c^{-1/3} \cdot d_{\min} \cdot R_{factor}$$

where

c	: fraction completeness
R_{factor}	: conventional crystallographic R-factor
N_{obs}	: number of reflections
d_{min}	: maximal resolution

If *Rfree* flags are specified, the program uses the Murshudov's approach to calculate the DPI (Newsletter on protein crystallography, Daresbury Laboratory, **33**, 25-30, 1997):

$$\sigma(x) \quad = \quad \sqrt{\frac{N_{atoms}}{N_{obs}}} \cdot c^{-1/3} \cdot d_{\min} \cdot R_{free}$$

Sovent content

The solvent content is the fraction of the unit cell volume not occupied by the model. The model consists of ALL the atoms present in the coordinate file.

Residual factor R_{merge}

$$R_{merge}(I) = \frac{\sum_i \sum_j \left(I_j - \langle I \rangle \right)}{\sum_i \sum_j \langle I \rangle}$$

where

I_j : the intensity of the jth observation of reflection *i*

$\langle I \rangle$: the mean of the intensities of all observations of reflection *i*

$\sum_i$ is taken over all reflections

$\sum_j$ is taken over all observations of each reflection

II. Local error estimation

The local error estimation (plotted for each residue, for the backbone and for the side chain) is given by the following information

1. the amplitude of the atom displacement from the electron density
2. the density correlation coefficient
3. the density index
4. the B-factor
5. the index of connectivity

Displacement

The displacement of the atoms of the elctron density is estimated from the*(F_{obs}-F_{calc})* difference map. The displacement vector is the ratio of the gradient of the density difference to the curvature at the atomic peak centre. The plot shows for each residue the *shift* as

$$shift = \frac{Displacement}{\sigma_{Displacement}}$$

where

$\sigma_{Displacement}$ is the standard deviation of the *Displacement*; the actual value of σ is given in the plot

Correlation coefficient

The density correlation coefficient is calculated for each residue from the *($2F_{obs}$-F_{calc})* map and the (F_{calc}) map:

$$D_{corr} = \frac{\langle \rho_{obs} \cdot \rho_{calc} \rangle}{\sqrt{\langle \rho_{obs}^2 \rangle \cdot \langle \rho_{calc}^2 \rangle}}$$

where

ρ_{obs} is the $(2F_{obs}\text{-}F_{calc})$ map

ρ_{calc} is the (F_{calc}) map

The value of the density for some atom of the map $\rho(x)$ is:

$$Dens = \frac{\sum_i \left(\rho(x_i) \cdot \rho_{atom}(x_i - x_a)\right)}{\sum_i \rho_{atom}(x_i - x_a)}$$

where

ρ_{obs} is the actual electron density

ρ_{calc} is the atomic electron density

x_a : the vector of the atom centre

x_i : the vector of the ith point of the grid

The sum is taken over all the grid points at a distance from the atom centre less than Radius_limit;

for all the atoms, Radius_limit = 2.5 Å.

Index of density and index of connectivity

The index of density is the geometric mean of the $(2F_{obs}\text{-}F_{calc})$ electron density values for the backbone (or side chain) atoms divided by the mean density. The actual values of the mean density, $\langle Dens \rangle$, and the standard deviation of the density, σ, are given in the plot.

The index of connectivity is a similar indicator which is calculated for each residue, in the case of proteins, for the backbone atoms (N, C_α and C), and, in the case of DNA/RNA, for the P, O5', C5', C3', O3' atoms. Low values of this index indicate breaks in the backbone electron density which may be due to flexibility of the chain or incorrect tracing.

III. Omit procedure

An omit map is a way to reduce the model bias in the electron density calculated with model phases. **SFCHECK** produces the so called total omit map by an automatic procedure. First, the initial (F_{obs} , PH_{model}) map is divided into *N* boxes. For each box, the electron density in it is set to zero and new phases are calculated from this modified map. A new map is calculated using these phases and F_{obs}. This map contains the omit map for the given box which is kept until the procedure is repeated for all boxes. At the end, all the boxes with omit maps compose the total omit map. Phases calculated from the total omit map are combined with the initial phases. The whole procedure may be repeated *n* times (keyword NOMIT). Note: it is time consuming !

IV. Incomplete information

The program can use only one input file of coordinates or structure factors. In this case, the program gives the information derived from the input file without local estimation.

V. Authors

A.A. Vagin, J. Richelle, S.J. Wodak

VI. Reference

Wodak, S.J., Pontius, J., Vagin, A.A, and Richelle, J.

Procedures for assessing the quality of X-ray structures of macromolecules

IUCr Computing school (in press) (1996)

VII. Version

This document describe version 4.0 of **SFCHECK**.

April 1997