Is it a "good" model?

X-ray structure of a voltage-dependent K⁺ channel

Youxing Jiang*, Alice Lee, Jiayun Chen, Vanessa Ruta, Martine Cadene, Brian T. Chait & Roderick MacKinnon
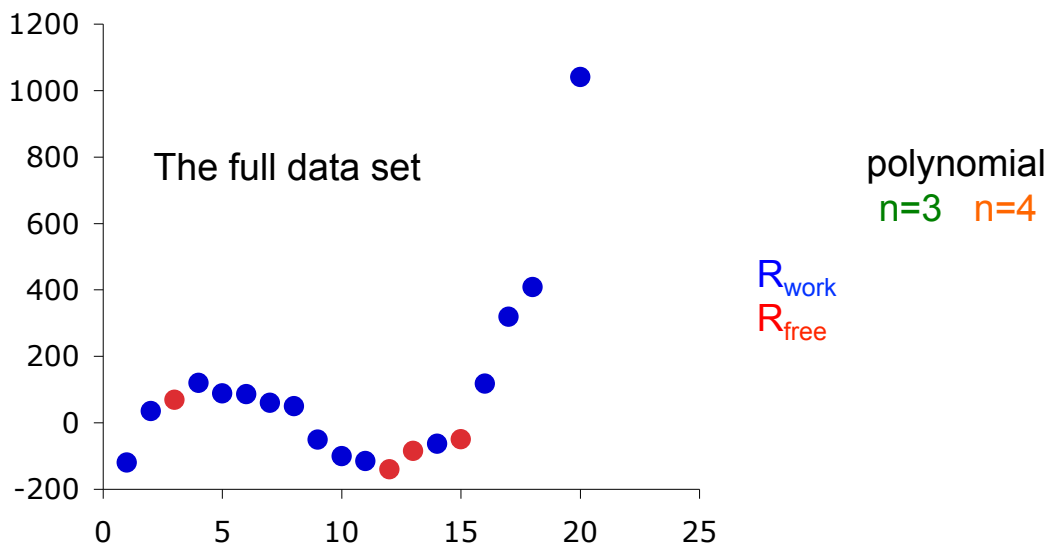
Table 1 **Data and refinement statistics**

Protein	Fab 6E1–KvAP	Fab 33H1–isolated voltage sensor
Space group	I422	C2
Data source*	CHESS A1	CHESS A1
Resolution (Å)	30–3.2	30–1.9
Completeness (%)	98.8 (95.2)	97.1 (92.8)
Redundancy†	5.5	2.9
R_{sym} (%)‡	8.0 (42.4)	5.8 (17.7)
I/σ	18 (2.2)	15.5 (5.8)
Unique reflections	22,476	56,778
Atoms refined	5,046 protein, 6 H_2O, 6 K, 7 Cd	4,365 protein, 403 H_2O
R_{work}/R_{free} (%)§	25.6 / 29.9	23.1 / 25.1
r.m.s.d. of bond‖	1.44°/0.009 Å	1.48°/0.006 Å

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$$

To Add Another Variable?

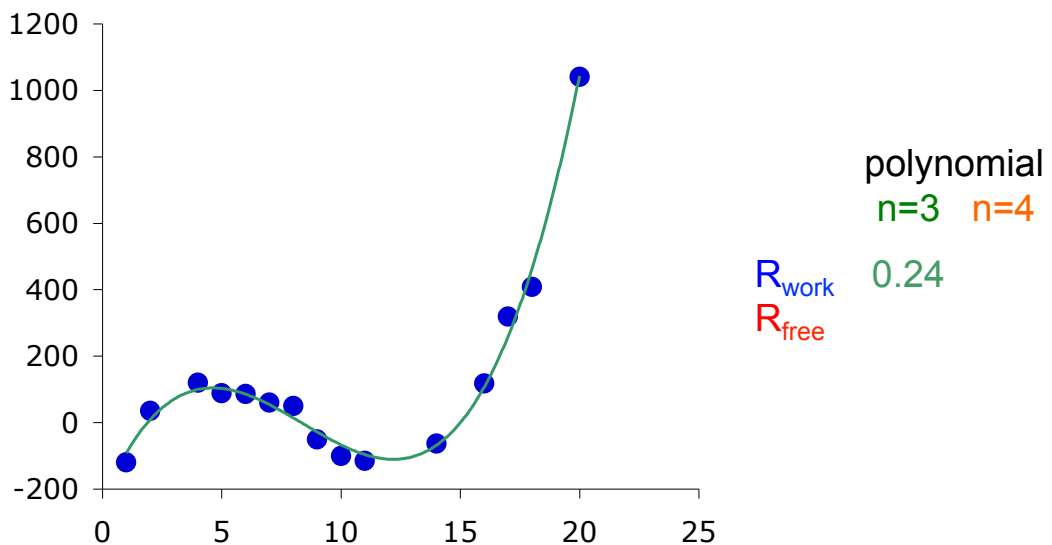$y = ax^3 + bx^2 + cx + d$ **or** $y = zx^4 + ax^3 + bx^2 + cx + d$

The full data set

polynomial
n=3 n=4

R_{work}
R_{free}

To Add Another Variable?

$y = ax^3 + bx^2 + cx + d$ **or** $y = \mathbf{z}x^4 + ax^3 + bx^2 + cx + d$

Only use 90% of data
for curve fitting

polynomial
n=3   n=4

R_{work}
R_{free}

polynomial
n=3   n=4

R_{work} 0.24
R_{free}

To Add Another Variable?

$y = ax^3 + bx^2 + cx + d$ **or** $y = zx^4 + ax^3 + bx^2 + cx + d$

polynomial
n=3   n=4

R_{work} 0.24   0.20
R_{free}

To Add Another Variable?

$y = ax^3 + bx^2 + cx + d$ **or** $y = zx^4 + ax^3 + bx^2 + cx + d$

polynomial
n=3   n=4

R_{work} 0.24   0.20
R_{free} 0.34

To Add Another Variable?

$y = ax^3 + bx^2 + cx + d$ **or** $y = \mathbf{z}\mathbf{x}^4 + ax^3 + bx^2 + cx + d$

polynomial
n=3   n=4

R_{work} 0.24 0.20
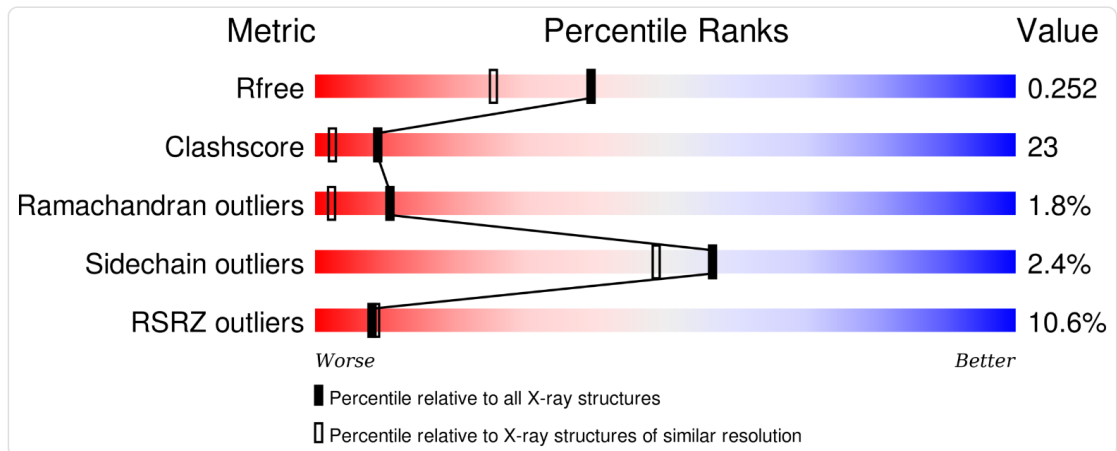R_{free} 0.34 0.43

Is it a "good" model?

X-ray structure of a voltage-dependent K$^+$ channel

Youxing Jiang*, Alice Lee, Jiayun Chen, Vanessa Ruta, Martine Cadene, Brian T. Chait & Roderick MacKinnon

Table 1 **Data and refinement statistics**

Protein	Fab 6E1–KvAP	Fab 33H1–isolated voltage sensor
Space group	I422	C2
Data source*	CHESS A1	CHESS A1
Resolution (Å)	30–3.2	30–1.9
Completeness (%)	98.8 (95.2)	97.1 (92.8)
Redundancy†	5.5	2.9
R_{sym} (%)‡	8.0 (42.4)	5.8 (17.7)
I/σ	18 (2.2)	15.5 (5.8)
Unique reflections	22,476	56,778
Atoms refined	5,046 protein, 6 H$_2$O, 6 K, 7 Cd	4,365 protein, 403 H$_2$O
R_{work}/R_{free} (%)§	25.6/29.9	23.1/25.1
r.m.s.d. of bond‖	1.44°/0.009 Å	1.48°/0.006 Å

Model Validation at PDB

Metric	Percentile Ranks	Value
Rfree		0.252
Clashscore		23
Ramachandran outliers		1.8%
Sidechain outliers		2.4%
RSRZ outliers		10.6%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Take Homes

• The R-factor measures the residual difference between the model and the experimental data. Lower values are better.

• R_{work} (or R_{cryst}) is calculated with the working data set, used in model building and refinement.

• R_{free} is calculated from the test data set, and tells you how well your model predicts the un-used data.

• Typically an R_{free} of **less than** 30% is considered satisfactory.