Supporting information for article:

The hidden treasure in your data: phasing with unexpected weak anomalous scatterers from routine data sets

Raghurama P. Hegde, Alexander A. Fedorov, J. Michael Sauder, Stephen K. Burley, Steven C. Almo and Udupi A. Ramagopal

Table S1   Anomalous peaks in final structure vs SHELXD peaks for HEWLAll data set.

Peaks were calculated by the program Anode (Thorn & Sheldrick, 2011). SHELXD was run using default parameters with three disulfide bonds resolved and an E-value cut off of 1.7

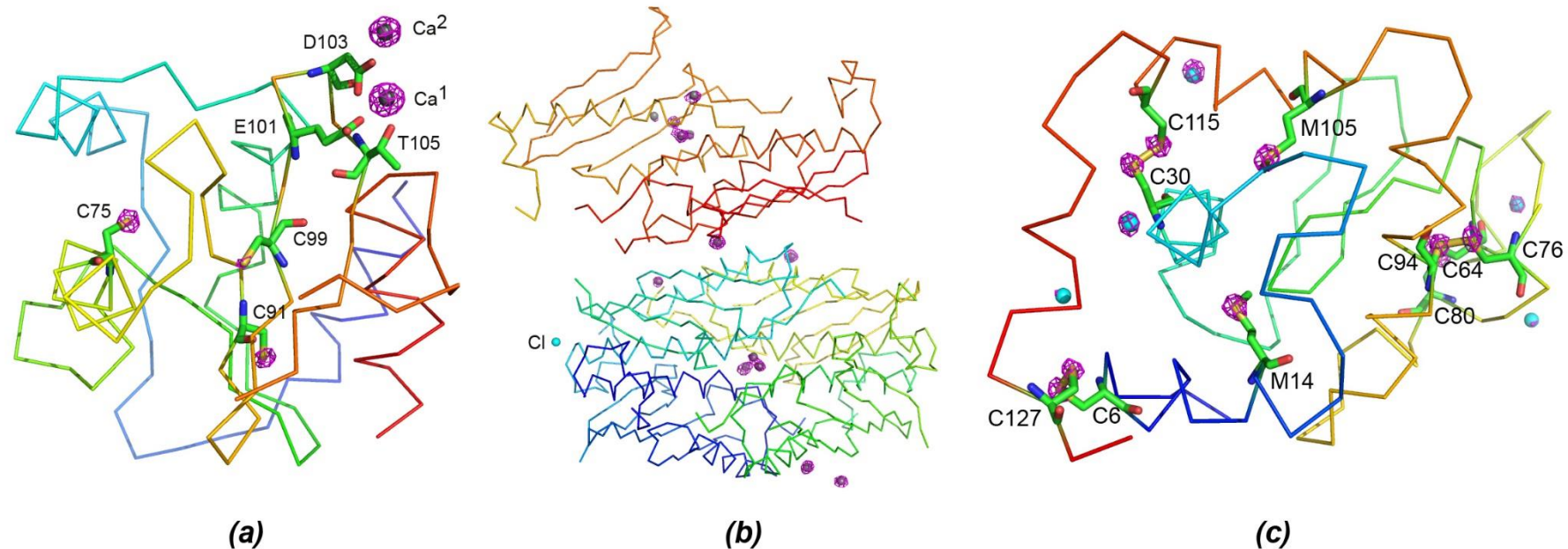| Anomalous peak height (σ)	B-factor (Å^2)	SHELXD peak no.
19.59 (Met105)	11.43	5
18.81 (Cys80)	11.36	2
18.66 (Cys64)	11.96	Not located (part of disulfide with Cys80)
18.17 (**Cl5**)	28.07	7
17.66 (Cys30)	12.2	4
16.73 (Met12)	12.97	1
15.77 (Cys115)	12.73	3 (disulfide with Cys30)
15.49 (Cys6)	17.81	Not located (part of disulfide with Cys127)
15.36 (Cys94)	13.70	8 (part of disulfide with Cys76)
14.51 (Cys76)	14.47	6 (part of disulfide with Cys94)
13.78 (Cys127)	16.11	9 (part of a disulfide with Cys6)
12.60 (Cl1)	20.39	Not located
12.13 (Cl3)	20.54	11
9.22 (Cl4)	26.73	Not located
8.01 (Cl2)	28.92	10

Figure S1 Representation of the structures of the three proteins with anomalous maps displayed around the anomalous scatterers in the proteins. The protein chains are shown as C^{α} traces and the anomalous maps are shown as purple meshes contoured at 5σ level. (a) PSPTO, the two calcium atoms are shown as grey spheres, the residues coordinating with the calcium atoms and the cysteines in the protein chain are shown and labelled; (b) PTO, the calcium atoms are shown as grey spheres and the chlorine atom as a cyan sphere. Each of the six chains in the asymmetric unit are colored differently; (c) HEWL, the cysteines and methionines in the protein chain are shown and labelled, the chlorine atoms are shown as cyan spheres. Made using PyMOL (Schrodinger, 2010).
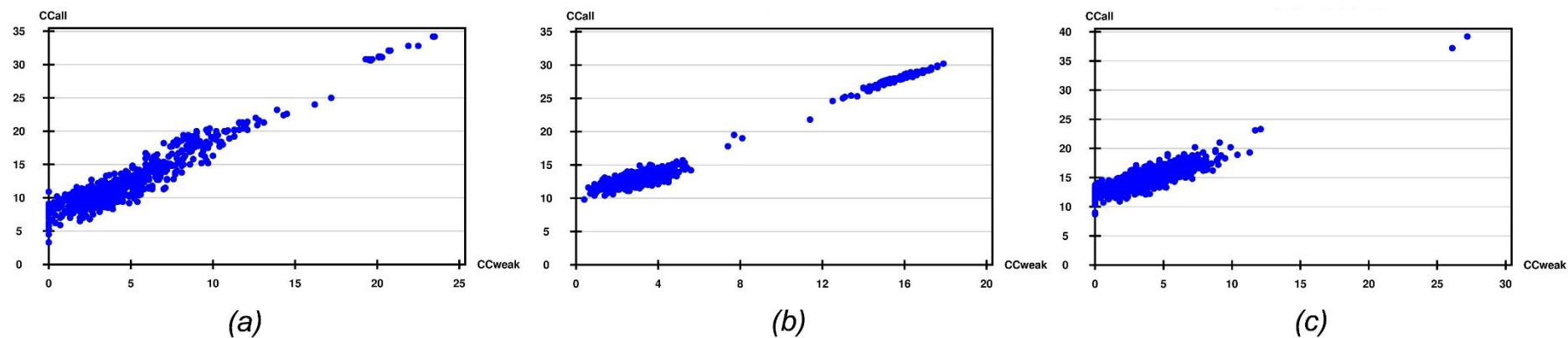
Figure S2 CC_{all} vs CC_{weak} from SHELXD run of 1000 cycles for each of the three cases: *(a)* PSPTO, *(b)*PTO and *(c)* HEWL (corresponding to data set HEWLAll) showing a clear bimodal distribution of solutions, separating the correct and incorrect substructure solutions.
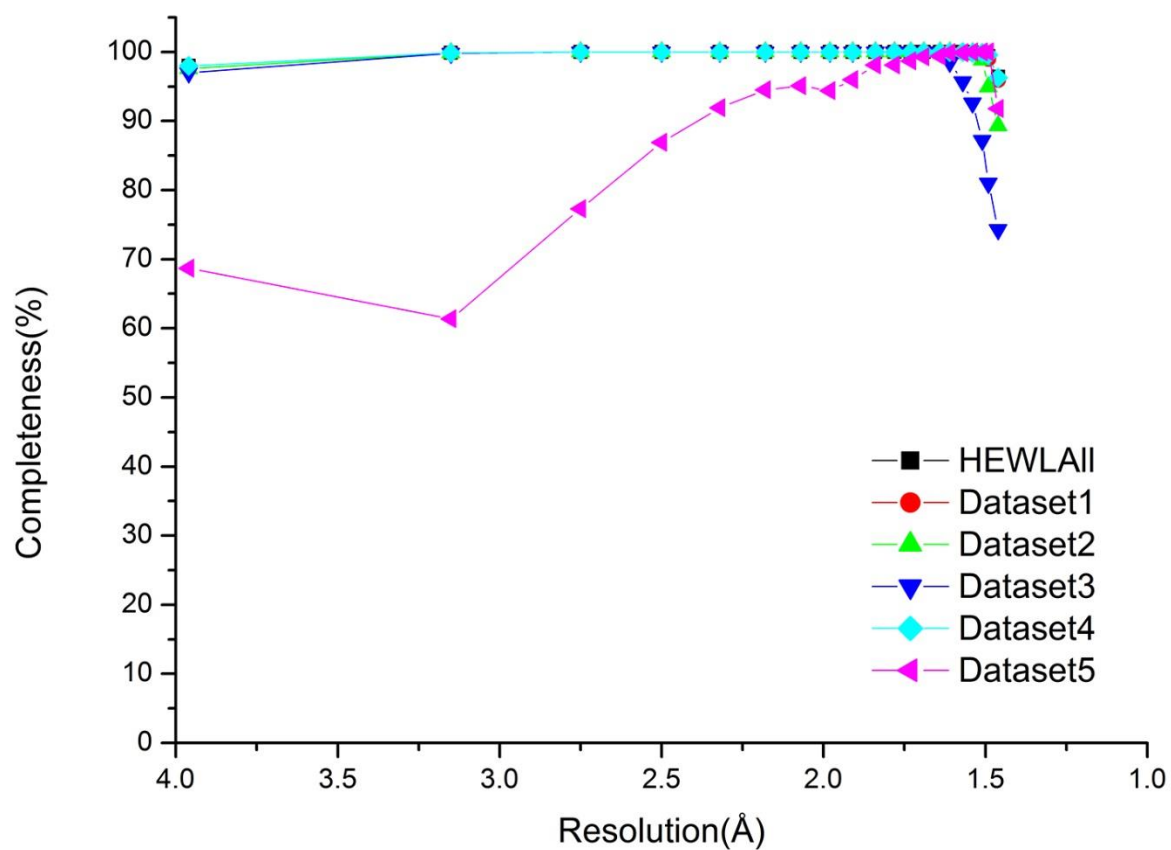
Figure S3 A plot of completeness vs resolution for the lysozyme (HEWL) data sets.
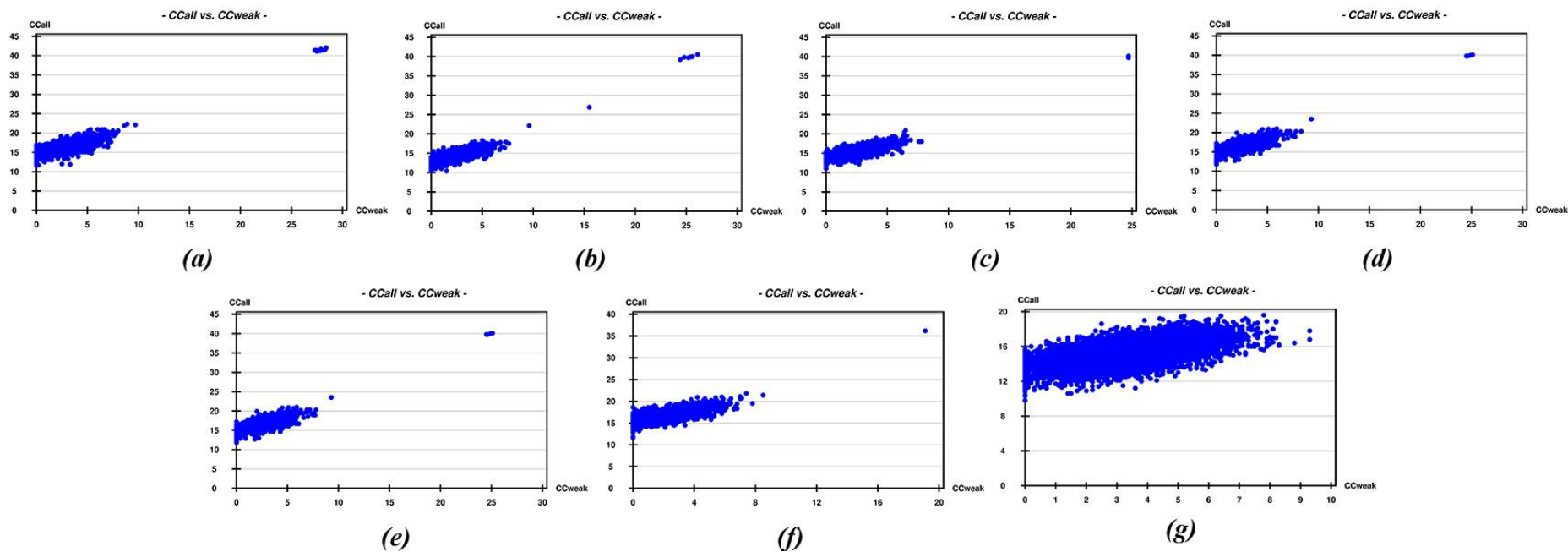
Figure S4 CC_{all} vs CC_{weak} from SHELXD run for HEWLAll data set with *(a)* top 5% intense reflections removed *(b)* top 10% intense reflections removed *(c)* top 15% intense reflections removed *(d)* top 20% intense reflections removed *(e)* top 25% intense reflections removed *(f)* top 30% intense reflections removed *(g)* top 35% intense reflections removed. All the above are for 1000 runs in SHELXD, except in *(g)* which represents the results after 10,000 runs.
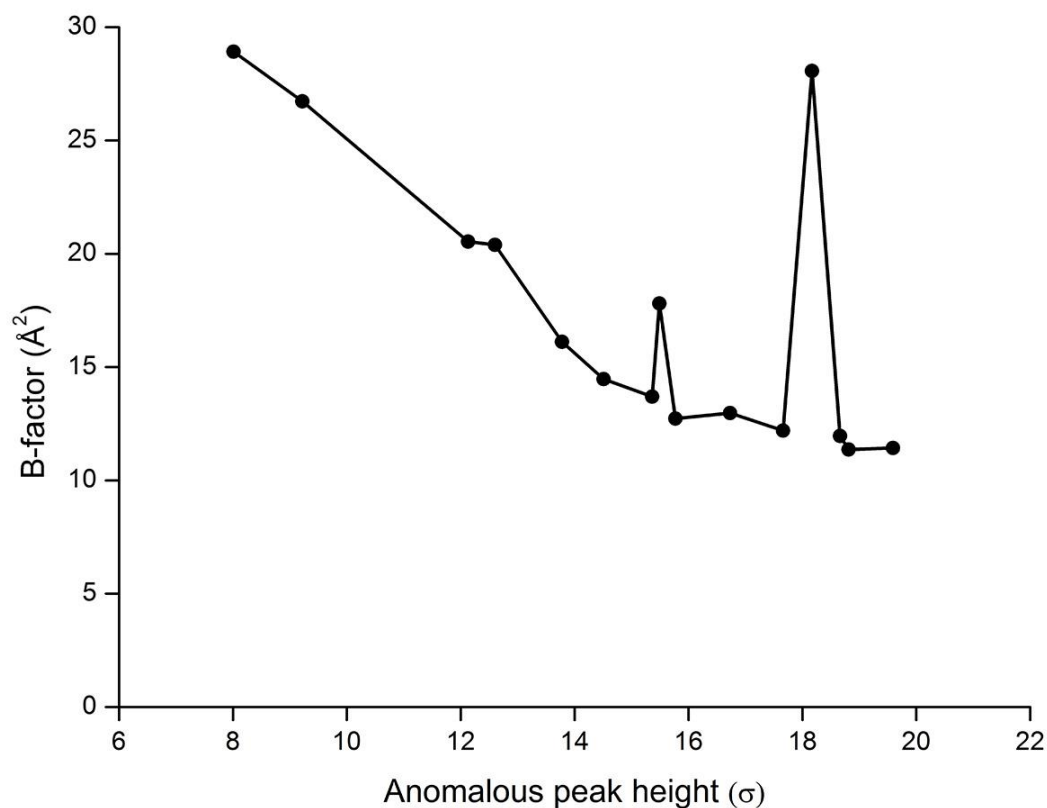
Figure S5 A scatter plot of the distribution of B-factor vs anomalous peak height in HEWL structure (generated from Table S1). The large spike around 18 σ corresponds to a surface bound chlorine which is expected to have a high B-factor and the smaller spike around 15.5σ corresponds to the sulfur of Cys6. Other than these outliers the general trend observed for the intrinsic sulfurs is a decrease in anomalous peak height with increase in B-factor.

References

Thorn, A. & Sheldrick, G. M. (2011). J. Appl. Crystallogr. 44, 1285-1287.

Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.5.0.3.

PSPTO _fa.res file

```
REM Best SHELXD solution:   CC 34.24   CC(weak) 23.47   CFOM  57.70
REM
TITL 10412i_S_fa.ins SAD in P41212
CELL  0.98000   47.43   47.43 122.49   90.00   90.00   90.00
LATT  -1
SYMM 1/2-Y, 1/2+X, 1/4+Z
SYMM -X, -Y, 1/2+Z
SYMM 1/2+Y, 1/2-X, 3/4+Z
SYMM 1/2-X, 1/2+Y, 1/4-Z
SYMM Y, X, -Z
SYMM 1/2+X, 1/2-Y, 3/4-Z
SYMM -Y, -X, 1/2-Z
SFAC S
UNIT  256
S001  1  0.003403  0.426918  0.022643   1.0000  0.2
S002  1  0.117233  0.408012 -0.059802   0.8380  0.2
S003  1  0.080734  0.395813  0.103959   0.6848  0.2
S004  1  0.105286  0.356133 -0.035196   0.5262  0.2
S005  1  0.330635  0.424683  0.135382   0.3388  0.2
S006  1  0.532372  0.296120  0.192777   0.3233  0.2
S007  1  0.046387  0.391068 -0.095932   0.1191  0.2
HKLF 3
END
```

PTO _fa.res file

```
REM Best SHELXD solution:   CC 30.19   CC(weak) 17.94   CFOM  48.13
REM
TITL 10163d_360_fa.ins SAD in C2221
CELL  0.98000   90.76 143.49 129.80   90.00   90.00   90.00
LATT  -7
SYMM -X, -Y, 1/2+Z
SYMM -X, Y, 1/2-Z
SYMM X, -Y, -Z
SFAC S
UNIT  256
S001  1  0.286736  0.135002  0.057452   1.0000  0.2
S002  1  0.416985  0.240723  0.127700   0.7908  0.2
S003  1  0.349785  0.005997  0.007505   0.5983  0.2
S004  1  0.180847  0.282982  0.107185   0.5293  0.2
S005  1 -0.012398  0.258965  0.233591   0.5050  0.2
S006  1  0.296280  0.315948  0.067201   0.4804  0.2
S007  1  0.467339 -0.012825  0.203872   0.4767  0.2
S008  1 -0.046928  0.337044  0.231282   0.4726  0.2
S009  1  0.030632  0.380539  0.060239   0.4725  0.2
S010  1 -0.134888  0.380630  0.103190   0.4388  0.2
S011  1  0.322906  0.001480  0.104827   0.4371  0.2
S012  1  0.183197  0.151321  0.143926   0.4284  0.2
```

```
S013   1  0.570572  0.235069  0.079985  0.4273  0.2
S014   1  0.458809 -0.008781  0.099924  0.4164  0.2
S015   1  0.149506  0.239586  0.136381  0.4105  0.2
S016   1  0.161339  0.364388  0.098257  0.4092  0.2
S017   1  0.238289  0.051743  0.004251  0.4066  0.2
S018   1  0.265968  0.314857  0.195208  0.4042  0.2
S019   1  0.041702  0.448143 -0.002794  0.4004  0.2
S020   1  0.516685  0.324188  0.039774  0.3964  0.2
S021   1  0.048965  0.205032  0.178017  0.3922  0.2
S022   1  0.122383  0.277672  0.016079  0.3907  0.2
S023   1  0.342438  0.201790  0.061948  0.3877  0.2
S024   1  0.327911  0.338631  0.286071  0.3762  0.2
S025   1  0.261909  0.254570  0.325043  0.3759  0.2
S026   1  0.203514  0.242760  0.231148  0.3618  0.2
S027   1  0.318268 -0.005173  0.208022  0.3564  0.2
S028   1  0.281044  0.233078 -0.032458  0.3347  0.2
S029   1  0.088036  0.201172  0.048840  0.3297  0.2
S030   1 -0.212364  0.386841  0.117061  0.3273  0.2
S031   1  0.018074  0.370079  0.034250  0.2872  0.2
S032   1  0.385262  0.214951  0.188541  0.2840  0.2
S033   1  0.317459  0.308258  0.346873  0.2829  0.2
S034   1  0.575638  0.283142  0.126433  0.2727  0.2
S035   1 -0.096169  0.348991  0.252597  0.2707  0.2
S036   1 -0.125580  0.314468  0.233224  0.2310  0.2
S037   1  0.346344 -0.021973  0.311114  0.1765  0.2
S038   1 -0.085876  0.175034  0.127747  0.1701  0.2
S039   1  0.170486  0.216293  0.103913  0.1521  0.2
S040   1  0.222885  0.313072  0.114861  0.1233  0.2
S041   1  0.357903  0.175400  0.200828  0.1228  0.2
S042   1 -0.124092  0.367493  0.194925  0.1187  0.2
HKLF 3
END
```