Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Not so clear on oxygen. Comment on *Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis* by Widboom *et al.* (2007), *Nature (London)*, 447, 342–345

Boguslaw Stec[a]* and Kimberly A. Stieglitz[b]

[a]Burnham Institute for Medical Research, La Jolla, CA 92037, USA, and [b]Department of Chemistry, University of Massachusetts, Boston, MA 02467, USA

Correspondence e-mail: bstec@burnham.org
Received 26 February 2008
Accepted 14 July 2008

Correctly performed crystallographic analysis is a richly rewarding process that has led to significant insights into chemistry and biology in the last 100 years. Establishing the correct crystal lattice and symmetry elements is the first important task of every crystal structure determination. It is not a trivial task, as is testified by the ∼10% of small-molecule structures in the Cambridge Crystallographic Database that have a mis-assigned space group (Marsh, 2004). The same issue became the subject of a recent commentary (Baker *et al.*, 2008) that was precipitated by several documented cases of such a problem among macromolecular structures (Le Trong & Stenkamp, 2007, 2008). Extra attention paid to detail may identify further problems, however, when researchers encounter a nonclassical crystallography problem. We are convinced that a case below may serve as such an example.

In a recently published article, *Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis* (Widboom *et al.*, 2007), the authors claimed to obtain the first clear view of dioxygen binding by a non-metal, non-cofactor-binding oxygenase. Discovering the chemical principles leading to dioxygen activation without using metal ions or cofactors would be an important contribution to chemistry. However, several aspects of the structure determination raise concerns about the structure itself and its mechanistic implications. In particular, the relatively high $R_{\rm merge}$ of 0.13 and high values for the final R factor (0.33) and $R_{\rm free}$ (0.36), together with the rationale offered for the high values of these statistical indicators, pointed to the likelihood of deeper underlying problems.

The Matthews coefficient as well as the unit-cell size suggested the presence of two trimers in the asymmetric unit. The reported structure showed only a single well ordered trimer (forming a biological hexamer by application of the crystallographic symmetry), while the second trimer was interpreted as being completely disordered. If indeed the second trimer (hexamer) was completely disordered, then the contribution to the scattering factors would be random, resulting in a lower $R_{\rm merge}$, R factor and $R_{\rm free}$ in the refinement. Previous papers reporting structures with large disordered fragments show normal values of these parameters (Tulinsky *et al.*, 1988; Robinson *et al.*, 2001; Troffer-Charlier *et al.*, 2007). High numerical values for these terms suggest that the contribution is not random and that the established symmetry of the lattice is incorrect. Additionally, in cases where not all diffracting atoms are included in the refinement, the temperature factors for the refined atoms are inaccurate and the electron density is biased towards explicitly modeled atoms.

Prompted by these observations, we retrieved the model and the diffraction data from the PDB (entry 2np9) and re-refined the model. We calculated electron-density maps from the deposited data and inspected in detail the molecular packing. The electron density was of good to very good quality, slightly better than expected at this resolution and $R_{\rm merge}$ level. Several $F_{\rm o} - F_{\rm c}$ density peaks suggested a modestly under-refined structure, but no major problems were found to explain the high R factor. Examination of the packing provided greater insight. In the original report it was stated that the electron density was of insufficient quality to position the second trimer

(hexamer). In actuality, the packing is so sparse that the second trimer could be docked in several positions. (The trimers located at those positions fail to create the biological hexamer, however.) This suggests that the structure might suffer from a 'lattice-translocation defect' (Wang, Kamtekar *et al.*, 2005; Bragg & Howells, 1954; Cochran & Howells, 1954), a well recognized but insufficiently popularized malady (Zhu *et al.*, 2008). The lattice-translocation defect was first described in the late 1950s, when streaky diffraction data led to the first formulation of its physical and mathematical description. The detailed theory was formulated by Wang, Kamtekar *et al.* (2005). Disregarding such a defect may lead to 'ghostly features' (additional non-existent molecules) and erroneous structural interpretation (Bochtler *et al.*, 2000; Wang, Rho *et al.*, 2005).

The presence of off-origin peaks in the native Patterson map was proposed as a conclusive test for identifying such a defect (Wang, Kamtekar *et al.*, 2005). We calculated the native Patterson map from the deposited data and the extra-origin peaks were found at ±0.085, ±0.1, 0.5, with a height $\sim$0.22 of the origin peak (Fig. 1). This indicates the likely existence of a lattice-translocation defect, which basically entails that all the crystal lattice layers containing the second trimer (hexamer) undergo a displacement that is random but precisely defined by the Patterson peaks. This is equivalent to a lower symmetry space group of the real crystal lattice. The higher $P2_12_12$ symmetry is an artifact of an uncorrected lattice-translocation defect. Although the possibility remains that a lower symmetry lattice does apply, we think that a lattice dislocation defect is the most likely explanation for the Patterson peak.

This complex three-dimensional defect is unlike those of previously published cases, which were limited to one dimension. The inspection of raw data can show clear signs of the translocation

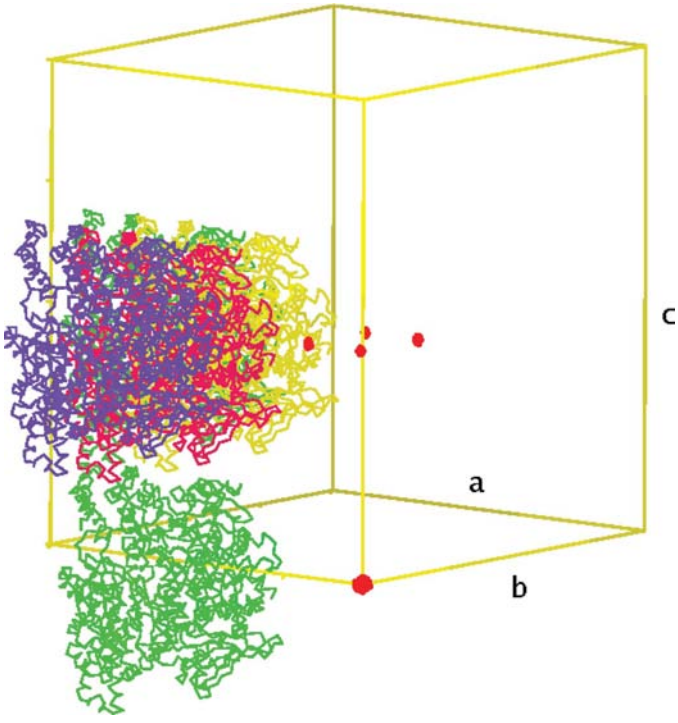

Figure 1
Contents of the asymmetric unit of the DpgC oxygenase crystal structure. A single molecule (green) represents the original fully ordered position while the remaining overlapping molecules (blue, red, gold and green) located near *c*/2 represent the disordered unmodeled layer discussed in this work. The Patterson map is represented by the red dots with dimensions proportional to their map height located at the cell origin and at positions centered around *c*/2.

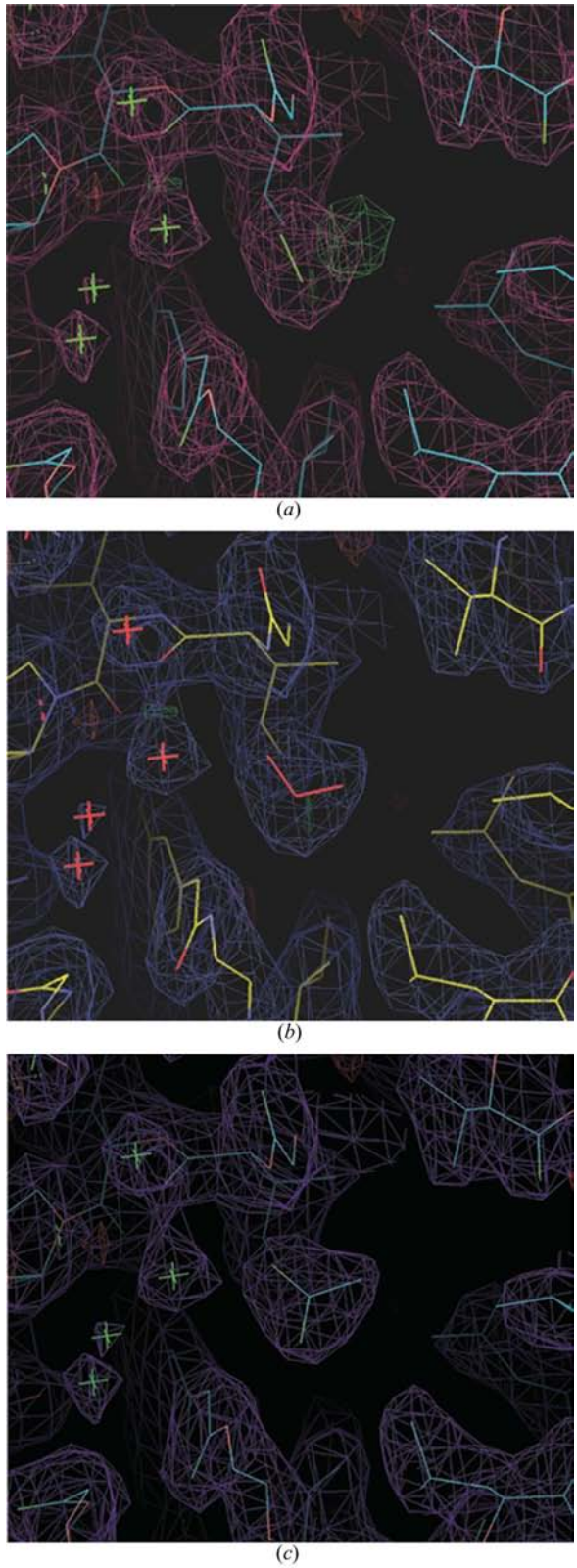

Figure 2
The active site of DpgC. (*a*) The model with dioxygen covered with $2F_o - F_c$ electron density contoured at 1.5σ (purple) and with $F_o - F_c$ electron density (green positive and red negative contours), shown at 3σ and -3σ levels, phased by the final model at 2.4 Å in the original space group $P2_12_12$. (*b*) The model with ozone at the active site. (*c*) The model with acetate at the active site. Note the disappearance of the difference electron density in (*b*) and (*c*) (same contouring as in *a*).

defect. The streakiness of most of the reflections and the presence of well defined layers (every tenth layer in this particular case) would be signs of the defect and would constitute a major obstacle in high-quality data reduction. Without access to the raw data it is impossible to determine which alternative unit cell might better describe this specific case. The experience with a similar case (Hwang *et al.*, 2006), however, suggests that the only simplification might be expected from reducing the data in the lower symmetry space group. The defect can be reduced from three dimensions to two dimensions by selecting a monoclinic or triclinic unit cell. If one of the unit-cell edges coincides with the main direction of the translocation, a further reduction to a one-dimensional defect can occur. Subsequent numerical correction along the lines suggested by Wang and coworkers should lead to an asymmetric unit containing two well defined hexamers. However, the translocation defect is equivalent to a global (correlated) disorder. Once the lattice is established, the discreet full molecule disorder is capable (to a certain extent) of mimicking the effect of the translocation.

In order to address the question of how the faulty structure determination might influence the main conclusions of this study, in regard to the presence of dioxygen (notoriously difficult to identify, especially at 2.4 Å resolution) and the proposed details of the catalytic mechanism, we carried out a more detailed refinement. First, molecular replacement was used to place the second trimer in an asymmetric unit. Refinement with two trimers in the asymmetric unit with full occupancy resulted in a significant lowering of the crystallographic R factor (∼3% drop) without increasing $R_{\rm free}$ and a model with comparable stereochemical quality. Encouraged, we modeled four overlapping trimers (in addition to the original trimer) with translations suggested by the Patterson vectors and with occupancies of 0.25. Because the dislocation disobeys the higher $P2_12_12$ symmetry assumed by Widboom and coworkers, this procedure results in a very crude approximation of the general defect. The additional trimers do not reproduce hexamers, but they do pack properly with the original trimer. Despite this rather crude modeling, which increases the number of refined parameters by fivefold, the refinement of the model with five trimers lowered the R factor by 9% and $R_{\rm free}$ by 3% and converged close to expected R and $R_{\rm free}$ levels (0.24 and 0.32, respectively). Even though our modeling procedure only roughly imitated the general distribution of the electron density in the lattice, the electron-density maps were improved and we were able to model more than 100 additional water molecules and some missing side chains. Notably, we could model two water molecules and two charged residues, Lys428 and Arg224, in the vicinity of the substrate.

Curiously enough, excess difference electron density (>3σ level) was apparent at the two active sites in the vicinity of the tentative dioxygen, suggesting that a larger molecule should occupy this site (Fig. 2). We modeled and refined an ozone molecule and acetate ion as possible alternative models. Both refined well, with B factors comparable to those of the protein atoms and eliminated the difference density. While the three-atom molecule (isoelectronic with ozone) was sufficient to eliminate the difference density, the acetate is probably the most likely to be present at this binding site, originating from the crystallization media (160 mM) and low apparent binding constant for dioxygen (Widboom *et al.*, 2007).

In summary, the general outline of the molecule in the published structure is most likely correct. Nevertheless, the crystallographic problems highlighted here render this work questionable. Without having clarified the crystallographic details and the role of added charged residues and water molecules in the active site, the proposed interpretation and mechanism remain speculative. Special attention should be focused on several factors: (i) identification of the correct space group, (ii) correction of the data for the translocation defect and (iii) identification of the entity bound at the active site, including evaluation of its contacts with water molecules, amide N 324, NZ of Lys428 and to the CB methyl group of Ala319.

References

Baker, E. N., Dauter, Z., Guss, M. & Einspahr, H. (2008). *Acta Cryst.* D**64**, 337–338.

Bochtler, M., Hartmann, C., Song, H. K., Bourenkov, G. P., Bartunik, H. D. & Huber, R. (2000). *Nature (London)*, **403**, 800–805.

Bragg, W. L. & Howells, E. R. (1954). *Acta Cryst.* **7**, 409–411.

Cochran, W. & Howells, E. R. (1954). *Acta Cryst.* **7**, 412–415.

Hwang, W. C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W. A. & Liddington, R. C. (2006). *J. Biol. Chem.* **281**, 34610–34616.

Le Trong, I. & Stenkamp, R. E. (2007). *Acta Cryst.* D**63**, 548–549.

Le Trong, I. & Stenkamp, R. E. (2008). *Acta Cryst.* D**64**, 219–220.

Marsh, R. E. (2004). *Acta Cryst.* B**60**, 252–253.

Robinson, R. C., Turbedsky, K., Kaiser, D. A., Marchand, J. B., Higgs, H. N., Choe, S. & Pollard, T. D. (2001). *Science*, **294**, 1679–1684.

Troffer-Charlier, N., Cura, V., Hassenboehler, P., Moras, D. & Cavarelli, J. (2007). *EMBO J.* **26**, 4391–4401.

Tulinsky, A., Park, C. H. & Skrzypczak-Jankun, E. (1988). *J. Mol. Biol.* **202**, 885–901.

Wang, J., Kamtekar, S., Berman, A. J. & Steitz, T. A. (2005). *Acta Cryst.* D**61**, 67–74.

Wang, J., Rho, S.-H., Park, H. H. & Eom, S. H. (2005). *Acta Cryst.* D**61**, 932–941.

Widboom, P. F., Fielding, E. N., Liu, Y. & Bruner, S. D. (2007). *Nature (London)*, **447**, 342–345.

Zhu, X., Xu, X. & Wilson, I. A. (2008). *Acta Cryst.* D**64**, 843–850.