Supplementary Material (ESI) for CrystEngComm

Supplementary Information for: Revised Electrostatics from Invariom Refinement of the 18-Residue Peptaibol Antibiotic Trichotoxin A50E†

[a,b]B. Dittrich[1], [b]C. S. Bond, [c]R. Kalinowski, [b]M. A. Spackman and [b]D. Jayatilaka

[a]Institut für Anorganische Chemie der Universität Göttingen, Tammannstr. 4, D-37077 Göttingen, Germany

[b]Chemistry M313, School of Biomedical, Biomolecular and Chemical Sciences, University of Western Australia, WA 6009 Crawley, Australia

[c]Institut für Chemie und Biochemie - Anorganische Chemie, Freie Universität Berlin, Fabeckstr. 36a, 14195 Berlin, Germany.

What is high resolution?

It remains unclear what high resolution in protein crystallography actually means. The term 'ultra-high' resolution in protein crystallography was introduced in the study of crambin [1] and used subsequently [2, 3, 4, 5, 6, 7, 8], albeit with data sets of crystals that diffracted to substantially lower resolution than crambin. We think that it is useful to relate the term 'high' resolution or even 'ultra-high' resolution, to empirical findings. Analogous to the definition of atomic resolution by Sheldrick [9], which was based on the applicability of direct methods for macromolecular structure solution, ultra-high resolution could be defined as requiring observations for more than half the number of theoretically measurable reflections around —or better than— a resolution threshold of $d \approx 0.9$ Å $[(\sin\theta/\lambda)_{max} \approx 0.56$ Å$^{-1}]$. Such a resolution —or beyond— is recommended for the application of invarioms in macromolecular structure refinement because this is the resolution threshold at which a physical significance of atomic displacement parameters (ADPs) emerges [10] when the non-spherical electron density is taken into account.

[1]bdittri@gwdg.de

Irrespective of the meaning of 'ultra-high' resolution the papers cited above provide a good overview of what can be achieved with high-resolution data on macromolecules as discussed in the main paper. A list of approximately 100 structures with a resolution better than 1 Å had been compiled in 2004 [11] and the number has already grown to more than 300 structures in the PDB [12]. These structures can provide a wealth of additional information beyond primary, secondary, tertiary and quaternary structure when compared to resolutions conventionally reached in protein crystallography.

Detail of least-squares refinement for trichotoxin A50E

- Atom notation: An individual notation of protein atoms was used to facilitate practical aspects of XDLSM refinements and the analysis of the structure. Residues were numbered from 01 to 19 in molecule one and from 20 to 38 in molecule two. A systematic notation for protein atoms was used, based on IUPAC notation [13]. One purpose of such a precise notation for individual atoms rather than non-specific notation for atoms belonging to residues is to facilitate identification of 'non-behaving' atomic parameters in the least-squares procedure. A further aim was to be able to automatically decode the residue information from this notation —including the position in the main or side chain— for generating a PDB file from cif-output (see below). The notation consists of the element symbol followed by the residue number, the residue one- or three-letter code (non-standard amino acids are assigned the letter X or UNK, terminal groups the letter Z or TER), a letter describing the position of the atom in the amino acid (' for atoms belonging to the protein backbone, a for α, b for β, c [or g[1]] for γ, d for δ, e for ϵ, n for ν and z for ζ) and finally a number, if necessary, if the previous letter is not specific enough; in case of disorder an extra letter

[1] c is preferred as the alphabetical order is perceived that way.

can be added. An example would be atom H09Qc1 (H009GLNc1 in three letter notation), which is part of glutamine residue 09 at position gamma. The small-molecule program PLATON [14] accepts this one-letter residue notation. Obviously this notation is limited to 99 (999) residues; the one letter residue notation does not allow to distinguish between non-standard amino acids.

- To smoothen the introduction of monopole populations, initially a refinement of only the scale-factor was performed, including the database monopole 'charges' and the aspherical valence scattering (option 'model 4' with individual $lmax = 0$ in XD) in the model.

- For the twenty proteogenic amino acids only 73 invarioms occur including different possible protonation states and mesomeric (delocalized electron density) structures [15]. The non-standard amino acids Aib (α-aminoisobutyric acid) and L-valinol [16] that occur in trichotoxin A50E are also included in the published subset of the invariom database. To each atom in the structure, initially except for the solvent water molecules, an invariom was assigned based on IAM molecular geometry. 30 out of 529 atoms had to be manually assigned, mainly in the disordered regions of the molecule and solvent region. For a full list of invarioms assigned please contact the author.

- Bond distances to hydrogen atoms were set to theoretically calculated values from model compounds. Those model compounds were the same as used for invariom assignment; X—H distances are also included in the invariom database. When these distances were also used in the IAM refinement, the R(F) increases to 6.60 % rather than 6.19 % using a weighting scheme of $1/\sigma^2$ (see Table 1).

- Isotropic temperature parameters of hydrogen atoms were constrained to be 2 times the U_{iso} of the non-hydrogen atoms to which they were

attached. A new feature in InvariomTool [17] is the ability to automatically generate SHELXL-like riding constraints for positional parameters of hydrogen atoms, that can be constrained to their parent atom. This procedure was used for all hydrogen atoms.

Programming

In a multipole refinement simultaneous adjustment of occupancy and multipole parameters of disordered atoms against the experimental structure factors can not yield reasonable results due to pronounced correlations. Conversely, when fixed scattering factors of database multipole populations are used, additional information (taking into account the aspherical electron density distribution) can be obtained for 'stabilizing' the refinement of occupancy parameters. We have attempted to implement and optimize occupancy parameters in a locally modified version of XDLSM. These enabled determination of the occupancies of solvent molecules and rotamers in trichotoxin A50E. However, since the occupancies are related to the monopole charges, which in turn influence the overall charge of the molecule, we kept the occupancies at the SHELX-result in final refinements.

Additional code modifications of the local version of XDLSM used include parallisation of the inversion of the least-square matrix for i686 multi-processor machines by Peof. E. Weckert (DESY, Hamburg), employing the math kernel library routines of the intel fortran compiler. On an 8-processor machine the inversion of the least-squares matrix is now as fast as using SHELXL. In order to monitor the progress of the refinement and to validate that the model was free of model bias, an R_{free} regime [18] was also implemented. For visualisation of Fourier maps the local XDLSM version was additionally interfaced to XFIT (from XTALVIEW [19]) or alternatively to COOT [20] *via* output of calculated structure factors and phases in a FCF file. The most convenient way to use COOT is to generate a SHELX-res file (that can then

Figure of merit	promolecule model	invariom model
R(F) /%	7.15	6.19
$R_w(F)$	3.53	2.93
$R_{free}(F)$	5.68	4.81
GoF	2.97	2.44
max. residual density	0.81	0.78
min. residual density	−0.46	−0.50

Table 1: Figures of merit for promolecule and invariom structure refinement using $w = 1/\sigma^2$.

be imported into COOT) with Platon [14] from XDGEOM CIF output.
CIF files can also be converted to a PDB file using a number of programs, *e.g.* MERCURY [21] or PLATON [14], which can also be imported into COOT. However, application of these programs typically results in loss of residue information and anisotropic temperature parameters. We have therefore implemented a conversion feature into the TONTO program [22] that is able to interpret the atom notation mentioned above (see PDB-file).

Figures of merit and weighting scheme

Table 1 shows additional figures of merit for IAM and the invariom models for a refinement using a weighting scheme of $1/\sigma^2$, comparing only refinements performed in XDLSM. Here the introduction of invarioms improves R(F) more significantly by 0.96% to a value of 6.19% when compared to the refinement reported in the main paper. The significance of the reduction in the R-factor is supported by the fact that no additional parameters were refined compared to the IAM; a validation of the significance with the test by Hamilton [23] is therefore not possible. Using this different weighting scheme, which is commonly used in multipole refinements, the weighted R(F) reduces to below 3%, which is remarkable for such a large molecule. Another result of the different $1/\sigma^2$ weighting scheme is that the Goodness of fit for XDLSM IAM refinement gives a value of 2.97, which can no longer be compared to

the SHELXL result (0.98), where the weighting scheme is adjusted to give approximately unity. Hence, in the paper we chose to use a SHELX-type weighting scheme to allow a comparison to the original refinement by Chugh *et al.* [24].

Refinement on F or on F^2?

It is recommended to perform refinements on F^2 in general for for various reasons [25, 26]. While SHELXL only allows refinement on F only for 'backwards compatibility', and the default is refinement on F^2 using all reflections, in XDLSM the default is still refinement on F with a sigma cutoff. Unfortunately the locally modified version of XDLSM used does not allow refinement on F^2 without a cutoff value. This amounts to refinement on F, since weak and negative intensities as present in the original data are still not included.

Aspherical scattering factors, restraints and disorder

We wish to point out that although aspherical scattering factors usually improve the agreement with small molecule refinements, such improvements may be absent when disorder occurs. For the L-valinol residues in trichotoxin A50E, where ADPs are comparably large, inaccurate bond distances of one of the Cβ–Cγ and the C'—O bond can be observed in unrestrained refinements, which become too short also in the invariom refinement. Badly defined bond distances indicate that even for subatomic-resolution pseudoatom refinements restraints are still necessary at least occasionally. Since restraints are not implemented in XDLSM, we intend to perform future refinements either with other software (*e.g.* MOPRO [27]) or PhenixRefine [28]) or to implement these necessary features for protein refinement. The commonly used distance restraints [29], while perfectly appropriate for IAM refinements, should be supplemented for standard and most common non-standard amino acids when the non-spherical electron density is taken into account, since

systematic differences in bond lengths occur [30]. However, it will need to be shown that this difference is detectable with using high-quality high-resolution data. We plan to extend the functionality of the preprocessor program InvariomTool [17] to facilitate such tasks in the future.

References

[1] Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C. *Proc. Nat. Acad. Sci. USA* **2000,** *97,* 3171–3176.

[2] Kang, B. S.; Cooper, D. R.; Jelen, F.; Devedjiev, Y.; Derewenda, U.; Dauter, Z.; Otlewski, J.; Derewenda, Z. S. *Structure* **2002,** *11,* 459–468.

[3] Afonine, P. V.; Lunin, V. Y.; Muzet, N.; Urzhumtsev, A. *Acta Cryst. D* **2004,** *60,* 260–274.

[4] Esposito, L.; Vitagliano, L.; Sica, F.; Sorrentino, G.; Zagari, A.; Mazzarella, L. *The Journal of Molecular Biology* **2000,** *297,* 713–732.

[5] Addlagatta, A.; Krzywda, S.; Czapinska, H.; Otlewskic, J.; Jaskolski, M. *Acta Cryst. D* **2001,** *57,* 649–663.

[6] Howard, E. I.; Sanishvili, R.; Cachau, R. E.; Mitschler, A.; Chevrier, B.; Barth, P.; Lamour, V.; Van Zandt, M.; Sibley, E.; Bon, C.; Moras, D.; Schneider, T. R.; Joachimiak, A.; Podjarny, A. *Proteins: Struct., Func. Bioinf.* **2004,** *55,* 792–804.

[7] Bönisch, H.; Schmidt, C. L.; Bianco, P.; Ladenstein, R. *Acta Cryst. D* **2005,** *61,* 5794–5799.

[8] Guillot, B.; Jelsch, C.; Podjarny, A.; Lecomte, C. *Acta Cryst. D* **2008,** *64,* 567–588.

[9] Sheldrick, G. M. *Acta Cryst. A* **1990,** *46,* 467–473.

[10] Dittrich, B.; Hübschle, C. B.; Messerschmidt, M.; Kalinowski, R.; Girnt, D.; Luger, P. *Acta Cryst. A* **2005,** *61,* 314–320.

[11] Afonine, P. V.; Urzhumtsev, A. *Acta Cryst. A* **2004,** *60,* 19–32.

[12] Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Research* **2000,** *28,* 235–242.

[13] IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry* **1970,** *9,* 3471–3479.

[14] Spek, A. L. *J. Appl. Cryst.* **2003,** *36,* 7–13.

[15] Dittrich, B.; Hübschle, C. B.; Luger, P.; Spackman, M. A. *Acta Cryst. D* **2006,** *62,* 1325–1335.

[16] Dittrich, B.; Munshi, P.; Spackman, M. A. *Acta Cryst. C* **2006,** *62,* o633-o635.

[17] Hübschle, C. B.; Luger, P.; Dittrich, B. *J. Appl. Cryst.* **2007,** *40,* 623–627.

[18] Brünger, A. T. *Nature* **1992,** *355,* 472–475.

[19] McRee, D. E. *Journal of Structural Biology* **1999,** *125,* 156–165.

[20] Emsley, P.; Cowtan, K. *Acta Cryst. D* **2004,** *60,* 2126–2132.

[21] Macrae, C. F.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Shields, G. P.; Taylor, R.; Towler, M.; van de Streek, J. *J. Appl. Cryst.* **2006,** *39,* 453–457.

[22] Jayatilaka, D.; Grimwood, D. J. *Computational Science - ICCS 2003* **2003,** *2660,* 142–151.

[23] Hamilton, W. C. *Acta Cryst. A* **1965,** *18,* 502–510.

[24] Chugh, J. K.; Brückner, H.; Wallace, B. A. *Biochemistry* **2002,** *41,* 12934–12941.

[25] Hirshfeld, F. L. *Acta Cryst. A* **1973,** *A29,* 510–513.

[26] Arnberg, L.; Hovmöller, S.; Westman, S. *Acta Cryst. A* **1979,** *35,* 497–499.

[27] Jelsch, C.; Guillot, B.; Lagoutte, A.; Lecomte, C. *J. Appl. Cryst.* **2005,** *38,* 38–54.

[28] Adams, P. D.; Gopal, K.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Pai, R. K.; Read, R. J.; Romo, T. D.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C. *Journal of Synchrotron Radiation* **2004,** *11,* 53–55.

[29] Engh, R. A.; Huber, R. *Acta Cryst. A* **1991,** *47,* 392–400.

[30] Coppens, P.; Sabine, T. M.; Delaplane, R. G.; Ibers, J. A. *Acta Cryst. B* **1969,** *25,* 2451–2458.