

Chapter 1

Perspectives and Pitfalls in Nucleic Acids Crystallography

Eric Westhof

Abstract

X-ray crystallography offers precious and striking knowledge on biomolecular architectures. Although safeguards do exist to guarantee the accuracy of the structures deposited in databases, they are not always applied, leading to the spread of inaccurate data. The importance of validation reports in the publication process is emphasized.

Key words Nucleic acids, Crystallography, Model quality, RNA, DNA, B factors, Validation

1 Introduction

The vast majority of our structural knowledge of nucleic acids, as it is for any biological molecule, is derived from X-ray diffraction data of crystals. The accuracy of the final structural model results from a complex set of factors. *First* and foremost, the preparation and purification of chemically pure RNA molecules constitute major experimental bottlenecks. Except for transfer RNAs or ribosomes, not too many RNA molecules or RNA–protein complexes can be obtained in sufficient quantities by extraction and purification from cultivated cells. Fortunately, chemical techniques and methods based on molecular biology tools have progressed enormously in the last 15 years. *Secondly*, the preparation of crystals diffracting at the highest possible resolution, in order to derive reliable structural models, is a major experimental challenge extremely demanding in patient manpower. *Thirdly*, data collection, treatment, the solution of the phasing problem, and the refinement form the core of the crystallographic work. From the early times of visual estimation of diffracted data obtained using copper anodes to present-day “zero-noise photon counting detectors” installed at novel generation synchrotron lines, the quality of the measured data and the proper choice of their mathematical treatment were both essential to derive accurate and cogent crystallographic structures that are biologically pertinent.

Finally, the biological relevance of the derived structural models has to be assessed and discussed. Highly chemically precise and accurate structures can still be of doubtful biological relevance. In this respect, the telling saga of the hammerhead ribozyme should be kept in mind. Those four main aspects of nucleic acid crystallography are addressed in this book.

The first eight articles are dedicated to RNA preparation and crystallogensis. *Alexander Serganov*, who crystallized and solved so many beautiful structures of riboswitches, starts with a protocol for the preparation of short oligonucleotides with modified 5'-ends often a must for RNA-protein complexes (with *Nikita Vasilyev*). It follows with a detailed description on how to prepare crystals of riboswitches with and without ligands (with *Alla Peselis* and *Ang Gao*). Transfer RNA was the first large RNA molecule to be crystallized (tRNA-Phe) from yeast because it could easily be separated from a mixture of tRNAs owing to the large modified base at position 37 in the anticodon loop (the wybutine or Y base). *Clément Dégut*, *Alexandre Monod*, *Franck Brachet*, *Thibaut Crépin*, and *Carine Tisné* describe original methods for producing and purifying large quantities of tRNAs for crystallization. In a next chapter, *Mélanie Meyer* and *Benoît Masquida* describe how to use polyacrylamide gel electrophoresis for preparing milligram amounts of large (around 200 nt) RNAs. Over the years, lots of efforts have been put in the development of systematic approaches for assisting both crystallogensis of RNA molecules and the solution to the phase problem of RNA crystals. A very successful method has been developed by *Adrian Ferré-D'Amaré*, while he was in *Jennifer Doudna's* laboratory, the use of the U1A protein. Here, *Adrian Ferré-D'Amaré* describes the improved and very practical method. This chapter is followed by a more recent method developed by *Jing-Dong Ye* while he was in *Piccirilli's* laboratory, the use of Fab fragments from phage display against a target RNA. With *Eileen Sherman* and *Jennifer Archer*, *Jing-Dong Ye* outlines very precisely the successive steps leading to the selection and co-crystallization of RNA molecules with a Fab assisting the crystallization and facilitating the resolution of the crystal structure. *Martin Egli* and *Pradeep S. Pallan* present the sole chapter on a DNA system, clearly a sign of present RNA-dominated times. They describe how the co-crystallization of DNA dodecamers with non-cleaving RNase H enzymes allows for the determination of sequences that do not crystallize by themselves. Finally in this first set of chapters, *Cyrielle da Veiga*, *Joelle Mezher*, *Philippe Dumas* and *Eric Ennifar* explain how a powerful biophysical method, isothermal titration calorimetry, by monitoring RNA/ligand binding parameters can increase the rate of success of crystallization experiments.

The next nine chapters form the core of the crystallographic work, once crystals have been obtained successfully, and they are devoted to data collection, data treatment, the solution of the

phase problem, and the final structural refinement. The first two chapters are of utmost importance because they deal with how the diffraction data are collected and computationally treated in order to obtain the best data for deriving accurate atomic models and precise phases by single- or multi-wavelength anomalous dispersion. Both chapters should constitute required reading by any crystallographer collecting data, either student or more senior. Indeed, evolving modern techniques and tools certainly allow for very rapid and simplified data collection protocols, but are still very demanding in high-level knowledge of data collection strategies in order to prevent model errors and loss of data resolution. The chapter written by *Kay Diederichs*, although focused on the widely used package XDS, contains a very general and rich presentation of data collection and data treatment. The second chapter, by *Aaron D. Finke, Ezequiel Panepucci, Meitian Wang, Vincent Olieric, Clemens Vonrhein, and Gérard Bricogne*, reviews all the fundamental aspects covering collection of anomalous data for phasing. In addition, this chapter provides advice helping quick and appropriate decision making while at the synchrotron where time is counted. The replacement of an oxygen atom by a sulfur atom (as in selenomethionine) has been extremely successful for solving the phase problem in protein crystallography. *Huiyan Sun, Sibó Jiang, and Zhen Huang* describe the synthesis of 2-Se-uridine containing RNAs to help in phasing. *Chloe Zubieta* and *Max H. Nanao* detail the process, called radiation damage induced phasing (or RIP), a surprising approach of transforming a drawback of X-ray irradiation into a method for solving the phase problem. In a next chapter, *Robert T. Batey* and *Jeffrey S. Kieft* describe the method they pioneered, the directed soaking of hexamine cations (iridium or cobalt) into RNA crystals containing a GoU wobble motif. *Marco Marcia* expounds very clearly and pragmatically the molecular replacement method for solving the phase problem. This very powerful method is going to take more and more importance owing to the increasing number of RNA crystal structures being regularly solved. The chapter by *Alexandre Urzhumtsev, Ludmila Urzhumtseva, and Ulrich Baumann* addresses a characteristic feature of nucleic acids, helical symmetry. The crystal packing is generally such that pseudo infinite helices appear, resulting in severe difficulties by standard molecular replacement methods. The authors describe various stratagems to turn a nucleic acid structural drawback into a positive contribution to the phase problem. A first crystallographic model fitting in the density map is rarely perfect and error-free in geometry, stereochemistry and without steric clashes. At this stage, the long process of refinement starts under the control of various tools monitoring geometry and fit into the density maps. *Fang-Chieh Chou, Nathaniel Echols, Thomas C. Terwilliger, and Rhiju Das* describe the very useful pipeline ERRASER (Enumerative Real-Space Refinement ASsisted by

Electron density under Rosetta) that couples reciprocal space in Phenix with real space refinement using Phenix. In a final chapter of this set on data collection, phasing, and refinement, *Toshiyuki Chatake* unfolds the steps necessary to perform neutron diffraction analysis of nucleic acid crystals. Neutron crystallography is a powerful tool for determining fine details in the hydration shells around nucleic acids, but also establishing correct enzymatic reactions.

In a last set of four chapters, various aspects related to the biology of nucleic acids are exposed. Three chapters are concerned with ribosomal translation and, among those, two are concerned with drug binding to RNA fragments. *Sultan Agalarov*, *Marat Yusupov*, and *Gulnara Yusupova* describe the in vitro reconstitution from free ribosomal RNA and ribosomal proteins into functional 30S subunit particles for structural and functional studies. In a following chapter, *Jiro Kondo* convincingly argues for the use of model RNA oligomers for systematic and thorough analysis of antibiotic binding to ribosomal fragments. In a similar vein, *Sergey M. Dibrov* and *Thomas Hermann* show how an important viral translation inhibitor of HCV could be structurally characterized by crystallizing it with a subdomain of the Internal Ribosome Entry Site of HCV. Finally, *Luigi D'Ascenzo* and *Pascal Auffinger* discuss the identification of ions in crystal structures, focusing especially on anions around the negatively charged nucleic acids. Several of these anions binding to nucleic acids are clearly due to the buffers used in crystallization conditions and, thus, not necessarily biologically relevant. But, at least, they should not be confused with other types of ions usually more linked to nucleic acids (like, for example, magnesium ions). They end up their chapter with a set of precious guidelines for avoiding solvent identification errors.

X-ray crystallography is a beautiful and successful tool for determining biomolecular structures and architectures. Its success is, undoubtedly, due to the fact that crystallography is firmly and solidly grounded in physics and chemistry. In addition, the extraordinary developments in data collection and computer power allow for complete mathematical and statistical treatments at all stages of the crystallographic process. However, despite the numerous mathematical or chemical safeguards and warnings available, errors in stereochemistry, in RNA folds or in solvent or ligand identifications do occur and are, unfortunately, rarely corrected afterwards. In the well-established protein crystallography field, several papers have appeared describing fatal mistakes or mishandling of data reporting, together with ways how to prevent and correct such errors [1–6]. Almost all, if not all, remarks, recommendations and suggestions made in those papers apply also to the field of nucleic acids crystallography. Very valuable tools for validations [7] and corrections [3] of nucleic acid structures do exist. The Protein Data Bank offers also various validation tools with compelling metrics [2]. As a personal note, while a post-doctoral fellow in

Madison, Wisconsin, in the laboratory of M. Sundaralingam (who with pencil and paper could draw in pre-ORTEP times nucleotides in a unit cell [8], but who unfortunately disappeared during the devastating South-East Asian tsunami 10 years ago), like other fellows in the laboratory, I had to type in all the crystal parameters and coordinates in order to check for stereochemistry, chirality and contacts for every single crystal structure of a nucleic acid component being submitted or published. Those were the days when computer-based databases, like the Cambridge Data Base [9] or the Protein Data Base [10], started to emerge fully. Interestingly, in the latter paper, a footnote warns “The Bank will assume no responsibility for checkout or correction of errors in such deposited programs.” Owing to the amazing interest and usefulness of structural databanks, a similar stand is not held anymore. And, in 1990, Brändén and Jones [5] wrote “It is the crystallographer’s responsibility to make sure that incorrect protein structures do not reach the literature.” Should the problems with bad protein structures be now duplicated for the structures of RNA? One can certainly argue that there is information in data with low signal to noise. But, can one really be confident in a RNA structure at 3.7 Å resolution with average B-factors around 200 Å², bad clash score, and poor PDB validation metrics? Only a systematic analysis of the various regions of the RNA molecule in electron density maps would allow a knowledgeable interested person to reach a personal opinion. Several of the chapters in this book address this difficult question and various protocols, programs and tools for assessing the quality of the data, the refinement, and the final structure are presented in detail. However, as set forth by several authors [1, 2], referees and journal editors have also a major role to play in order to prevent incorrect structures to reach the literature and, almost worse, depository databases. Referees should request complete and detailed statistics tables, validation reports and quality indicators, coordinates, and electron-density maps and editors as well as authors should comply with such requests despite fierce competition for publication.

References

1. Brown EN, Ramaswamy S (2007) Quality of protein crystal structures. *Acta Cryst D* 63:941–950
2. Read RJ, Adams PD, Arendall WB 3rd, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Luttko T, Otwinowski Z, Perrakis A, Richardson JS, Sheffler WH, Smith JL, Tickle IJ, Vriend G, Zwart PH (2011) A new generation of crystallographic validation tools for the protein data bank. *Structure* 19:1395–1412
3. Chou FC, Sripakdeevong P, Dibrov SM, Hermann T, Das R (2013) Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods* 10:74–76
4. Wlodawer A, Minor W, Dauter Z, Jaskolski M (2013) Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J* 280:5705–5736

5. Brändén C-I, Jones TA (1990) Between objectivity and subjectivity. *Nature* 343:687–689
6. Dauter Z, Wlodawer A, Minor W, Jaskolski M, Rupp B (2014) Avoidable errors in deposited macromolecular structures: an impediment to efficient data mining. *IUCrJ* 1:179–193
7. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383
8. Sundaralingam M, Jensen LH (1965) Stereochemistry of nucleic acid constituents. I. Refinement of the structure of cytidylic acid b. *J Mol Biol* 13:914–929
9. Kennard O, Allen FH, Brice MD, Hummelink TWA, Motherwell WDS, Roidgers JR, Watson DG (1977) Computer based systems for the retrieval of data:crystallography. *Pure Appl Chem* 49:1807–1816
10. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 112: 535–542