X–ray Crystallography and NMR: Complementary Views of Structure and Dynamics

Axel T. Brunger

The Howard Hughes Medical Institute and

Department of Molecular Biophysics and Biochemistry,

Yale University, New Haven, CT 06520, USA.

Phone (203) 432-6143

FAX (203) 432-6946

e-mail address: axel.brunger@yale.edu

Nuclear magnetic resonance (NMR) spectroscopy and X–ray crystallography are complementary methods which are routinely used to study macromolecular structure and dynamics. Over the past decade, developments in biotechnology, instrumentation, and computational methods have allowed nearly exponential growth of macromolecular structural and dynamic studies using both techniques. NMR and X–ray crystallography share certain similarities: both methods determine time– and spatial averages of molecular properties, in particular atomic coordinates. For both techniques, average properties are determined from measurements performed on an ensemble of molecules whose population is of the order of Avogadro's number. In addition to providing a description of the ensemble average, diffraction and NMR data reveal information about variability within the ensemble, for example the type and relative population of conformational substates [1, 2] of the molecule.

Beyond these similarities, significant differences between NMR and X–ray crystallography exist, notably in terms of the spatial distribution of the molecules in the sample and time scales accessible to each method. Solution NMR data represent an average over semi–randomly oriented molecules in solution, whereas diffraction data represent an average over molecules arranged in a periodic crystal lattice. Analysis of thermal diffuse (non–Bragg) scattering [3, 4] from crystals, and solid–state NMR techniques [5, 6] are somewhere between these two extremes: the former method provides information about correlated thermal motion within or between crystal unit cells while the latter method allows one to obtain structural information from samples in the solid phase.

NMR data represent averages in the nanosecond to second time regime, while diffraction data typically represent averages in the seconds to hours time regime. Much shorter time scales are sometimes accessible in Laue diffraction experiments which allow the study of time–dependent processes in the nanosecond time scale provided the process can be synchronized in all unit cells of the crystal [7].

X–ray crystallography can be applied to very large macromolecular complexes (> 100 kDa) as long as suitable crystals can be obtained whereas solution NMR is presently limited to significantly smaller systems. There has been, however, a steady increase of the molecular size limit that can be studied by NMR. For example, the solution structure of a 27kDa protein [8] and backbone resonances of a 64 kDa trp repressor–operator complex [9] have been recently reported.

Although limited in terms of the size of molecules that can be studied at present, NMR has a number of advantages over X–ray crystallography in studies of macromolecular dynamics. Different parts of the molecule may exhibit different degrees of flexibility. For example, a multi–domain protein can have a flexible linker region or unstructured domain and NMR can be used to identify these flexible regions [10]. Relaxation measurements provide time–dependent correlation functions that can be interpreted in terms of dynamical models [11] and macromolecular dynamics [12, 13]. NMR has been used to study protein folding: equilibrium hydrogen exchange data provide information about the degree of protection of hydrogen donors due to secondary structure formation [14]. H/D pulse–labeling [15, 16] and real–time NMR experiments [17] allow the time–dependence of structure

formation to be followed.

NMR experiments provide chemically specific, local structural information about macromolecules, such as nuclear Overhauser effect (NOE)–derived inter–atomic distances, J–coupling or chemical shift derived dihedral angle restraints, or orientational information [18, 19, 20]. In contrast, X–ray diffraction data are related to all atomic properties by a Fourier transformation [21, 22]. These fundamental differences between the two types of experimental data have implications for the way atomic models are determined and their quality.

Macromolecules undergo significant motion, including vibrational motion and transitions between conformational substates [1, 2]. Thus, a crystal or NMR "structure" is a model of this physical reality where the coordinates describe the mean of all molecular conformations accessible on the time scale of the experiment. Conformational variability of the macromolecule can be modelled with additional parameters (e.g., thermal factors [21] or multi–conformer models [23]) where the maximum allowable number of these additional parameters is related to the number of observations [25].

How accurate are X–ray crystal and solution NMR structures? Accuracy of a structure is defined as the difference between the determined structure and the conformations of the molecule in its native environment. By definition, it is impossible to obtain an absolute measure of accuracy since no existing method allows the study of single macromolecules *in vivo*. Accuracy is thus a theoretical concept that can only be determined when the "answer" is known, for example, by using synthetic data. Precision (in crystallography more commonly referred to as coordinate error) is defined

as the extent to which the determined structure differs from a hypothetical "error–free" and "complete" atomic model determined from noise–free data. In contrast to accuracy, statistical and/or empirical methods can be used to reliably estimate precision.

Figures 1 and 2 illustrate the effect of decreasing experimental information on the coordinate error or precision of X–ray and solution NMR structures. In case of X–ray crystallography, the lower the resolution, the less precise the structure. For the particular case studied, the coordinate error could be objectively determined by comparison to the solved crystal structure at 1.8 Å resolution. Normally, the coordinate error cannot be determined this way, so it has to be estimated using a statistical description [24] of the expected errors of the model and the diffraction data (Fig. 1). Cross–validation [25, 26] plays a crucial role here since it prevents overfitting the diffraction data which would result in an artificially low coordinate error estimate.

In the case of NMR structures, a statistical theory to connect data error to coordinate error has yet to be developed. Thus, at present, coordinate error can only be assessed empirically by determining the root–mean–square (rms) deviation around the mean (or the pairwise rms difference) of an ensemble of independently determined structures that agree with the NMR data to some specified degree (Fig. 2b). The case used here is representative of well–determined NMR structures [30]. As the completeness of the NOE–data is decreased, the precision worsens and the difference to the structure with complete data inreases nearly linearily [31]. However, even with only two randomly selected NOEs per residue, there is enough

information to obtain a low–resolution structure as long as at least some long–range NOEs are included in the data. It is thus conceivable that the inclusion of empirical database information about protein conformations [32] may increase the precision of these structures and enable NMR structure determinations of large macromolecules with sparse sets of NMR–data.

The above examples illustrate that the precision of a high–resolution crystal structure at 1.8 Å resolution (typically 0.2 – 0.3 Å, Fig. 1) exceeds that of a well–determined NMR structure (typically 0.5 – 1.0 Å, Fig. 2). The atomic rms differences between independently solved structures of the same macromolecule by NMR and X–ray crystallography are generally within the precision of the corresponding NMR structures [19]. Are these differences a consequence of the limited quality of the NMR structures or do they reflect genuine differences between crystal and solution structures? This question has been the subject of intense discussions and controversy.

Crystal structures can be affected by inter–molecular packing or crystal lattice packing which may lock the structure into a particular substate of the molecule, whereas in solution, the molecule can adopt many more conformations. Under favorable circumstances, X–ray crystallography can shed some light on the influence of such interactions. High–resolution crystal structures with multiple, independent copies of the same molecule in the asymmetric unit provide different environments for each of the copies, allowing unbiased estimation of the effect of the crystal lattice on the structure. The rms differences between these non–crystallographic–symmetry (NCS)–related molecules range from 0.5 to 3 Å for high–resolution crystal structures [34] (Fig. 3a). These differences are predominantly caused

by inter–domain motions or flexible domains. Figure 3b shows a striking example consisting of two NCS–related molecules which have very different conformations for an N–terminal region [35, 36]. The large change is induced by dimerization of this particular fragment of mannose binding protein A. Yet each conformation is precisely determined as evidenced by the excellent quality of the experimentally phased electron densities which are entirely free of model–bias [36]. In other regions of the molecule, there are indications of local disorder and conformational variability which can be effectively modeled by multi–conformer models [23, 36]. Thus, information about conformational variability can be obtained from X–ray crystal structures in favorable cases.

While the collection of NCS–related molecules in a crystal structure generally represents only a subset of the conformational substates of the macromolecule it would be useful to obtain information about all conformational substates. It is tempting to relate the distribution of NMR structures in a refined ensemble (Fig. 2b) to conformational variability in solution if appropriate precautions have been taken to ensure proper sampling of conformational space. Multi–conformer models can also be employed during NMR–refinement if sufficient data is available [38]. However, the interpretation of these distributions and multi–conformer models is made difficult because apparent variability can be due to a number of unrelated phenomena, including conformational averaging, relaxation, oligomerization, or experimental difficulties during data collection. Until a statistical theory along with advances in instrumentation are developed that relates NMR data to conformational variability, it will be difficult to determine all conformational substates of macromolecules in solution.

The combination of NMR and X–ray diffraction data can be useful for structure determination by either method. For example, an NMR structure can be used to solve a crystal structure by molecular replacement [39, 40, 41, 42, 43], and a crystal structure can be used to obtain assignments of the NMR resonances. Independent structure solution by both methods can be useful to validate the structures. NMR and X–ray diffraction data can be combined to obtain more precise models [44, 45, 46, 10].

NMR and X–ray crystallography are complementary methods for probing a wide range of structural and dynamical properties of macromolecules. Application of both methods to a particular biological problem will likely produce a better understanding of the biological function involved than using each method individually.

Acknowledgements

I thank G.M. Clore and G.J. Kleywegt for stimulating discussions, P.D. Adams, K.M. Fiebig, L.M. Rice, S.C. Stallings, & G.L. Warren for a critical reading of the manuscript, and F.T. Burling for help with Fig. 3b.

References

[1] Frauenfelder, H., Petsko, G.A., & Tsernoglu, D. Temperature–dependent x–ray diffraction as a probe of protein structural dynamics. *Nature (Lond.)* **280**, 558–563 (1979).

[2] Hartmann, H., Parak, F., Steigemann, W., Petsko, G.A., Ringe, D., & Frauenfelder, H. Conformational substates in a protein: structure and dynamics of metmyoglobin at 80K. *Proc. Natl. Acad. Sci. USA* **79**, 4967–4971 (1982).

[3] Caspar, D.L.D., Clarage, J., Salunke, D.M., & Clarage, D. Liquid–like movements in crystalline insulin. *Nature* **332**, 659–662 (1988).

[4] Phillips, G.N.Jr., Fillers, J.P., & Cohen, C. Motions of tropomyosin. *Biophys. J.* **10**, 485–502 (1980).

[5] Cross, T.A., Opella, S.J. Protein structure by solid-state NMR. *J. Am. Chem. Soc.* **105**, 306–308 (1983).

[6] Raleigh, D.P., Levitt, M.H., Griffin, R.G. Rotational resonance in solid state NMR. *Chem. Phys. Lett.* **146**, 71–76 (1988).

[7] Moffat, K., Szebenyi, D., Bilderback, D. X–ray Laue diffraction from protein crystals. *Science* **223**, 1423–1425 (1984).

[8] Martin, J.R., Mulder, F.A.A., Karimi-Nejad, Y., van der Zwan, J., Mariani, M., Schipper, D., & Boelens, R. The solution structure of serine protease PB92 from *Bacillus alcalophilus* presents a rigid fold with a flexible substrate–binding site. *Structure* **5**, 521–532 (1997).

[9] Shan, X., Gardner, K.H., Muhandiram, D.R., Rao, N.S., Arrowsmith, C.H. & Kay, L.E. Assignment of ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and HN resonances in an ^{15}N, ^{13}C, ^{2}H labeled 64 kDa trp repressor–operator complex using triple resonance NMR spectroscopy and ^{2}H-decoupling. *J. Am. Soc.* **118**, 6570–6579 (1996).

[10] Muchmore, S.W., Sattler, M., Liang, H., Meadows, R.P., Harlan, J.E., Yoon, H.S., Nettesheim, D., Chang, B.S., Thompson, C.B., Wong, S.-K., Ng, S.-C., & Fesik, S.W. X–ray and NMR structure of human Bcl-X_L, an inhibitor of programmed cell death. *Science* **381**, 335–341 (1996).

[11] Lipari, G. & Szabo, A. Model–free approach to the interpretation of nuclear magnetic resonance in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).

[12] Bax, A. Multidomensional nuclear magnetic resonance methods for protein studies. *Curr. Opin. Struct. Biol.* **4**, 738–744 (1994).

[13] Kay, L.E. NMR methods for study of protein structure and dynamics. *Biochem. Cell. Biol.* **75**, 1–15 (1997).

[14] Wagner, G. & Wüthrich, K. Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two–dimensional nuclear magnetic resonance. *J. Mol. Biol.* **160**, 343–361 (1982).

[15] Roder, H. & Wüthrich, K. Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins* **1**, 34–42 (1986).

[16] Roder, H., Elove, G.A., & Englander, S.W. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and protein NMR. *Nature* **335**, 700–704 (1988).

[17] Balbach, J., Forge, V., Lau, W.S., van Nuland, N.A.J., Brew, K., & Dobson, C.M. Protein folding monitored at individual residues during a two–dimensional NMR experiment. *Science* **274**, 1161–1163 (1996).

[18] Wüthrich, NMR of proteins and nucleic acids, John Wiley & Sons, New York, 1986.

[19] Gronenborn,A.M., Clore, G.M. Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. *Crit. Rev. Biochem. Mol. Biol.* **30**, 351–385 (1995).

[20] Wüthrich, K. NMR – this other method for protein and nucleic acid structure determination. *Acta Cryst. D* **51**, 249–270 (1995).

[21] Stout, G.H. & Jensen, L.H., eds. *X-ray Structure Determination, A Practical Guide*, John Wiley & Sons, New York (1989).

[22] Drenth, J., Principles of protein X–ray crystallography, Springer Verlag, New York (1994).

[23] Burling, F.T., & Brunger, A.T. Thermal motion and conformational disorder in protein crystal structures: Comparison of multi–conformer and time–averaging models. *Israel Journal of Chemistry* **34** 165–175 (1994).

[24] Read, R.J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140–149 (1986).

[25] Brunger, A.T. The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–474 (1992).

[26] Kleywegt, G.J., Brunger, A.T. Checking your imagination: applications of the free R value. *Structure*, **4**, 897-904 (1996).

[27] James, M.N.G. & Sielecki, A. R. (1983). Structure and refinement of penicillopepsin at 1.8 Å resolution. *J. Mol. Biol.* **163**, 299–361.

[28] Brunger, A.T., Krukowski, A., Erickson, J. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst.* **A46**, 585–593 (1990).

[29] Hendrickson, W.A. (1985). Stereochemically restrained refinement of macromolecular structures. *Meth. Enzymol.* **115**, 252–270.

[30] Gronenborn, A.M., Filpula, D.R., Essig, N.Z., Achari, A., Whitlow, M., Wingfield, P.T., & Clore, G.M. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657–661 (1991).

[31] Brunger, A.T. Clore, G.M., Gronenborn, A.M., Saffrich, R., & Nilges, M. Assessing the quality of solution nuclear magnetic resonance structures by complete cross–validation. *Science* **261**, 328–331 (1993).

[32] Kuszewski, J., Gronenborn, A.M., & Clore, G.M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Science* **5**, 1067–1080 (1996).

[33] Nilges, M., Clore, G.M., Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324 (1988).

[34] Kleywegt, G.J. Use of non–crystallographic symmetry in protein structure refinement. *Acta Cryst. D***42**, 842–857 (1996).

[35] Weis, W.I., Kahn, R., Fourme, R., Drickamer, K., & Hendrickson, W.A. Structure of the calcium–dependent lectin domain from a rat mannose–binding protein determined by MAD phasing. *Science* **254** 1608–1615 (1991).

[36] Burling, F.T., Weis, W.I., Flaherty, K.M., & Brunger, A.T. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **271**, 72–77 (1996).

[37] Kleywegt, G.J., Jones, T.A. Where freedom is given, liberties are taken. *Structure* **3**, 535–540 (1995).

[38] Bonvin, A.M.J.J. & Brunger, A.T. Conformational Variability of solution nuclear magnetic resonance structures. *J. Mol. Biol.* **250**, 80–93 (1995).

[39] Brunger, A.T., Campbell,R.L., Clore,G.M., Gronenborn, A.M., Karplus,M., Petsko, G.A., & Teeter, M.M. Solution of a protein crystal structure with a model obtained from NMR interproton distance restraints. *Science* **235**, 1049–1053 (1987).

[40] Braun, W., Epp, O., Wüthrich, K., Huber, R. Solution of the phase problem in the X–ray diffraction method for proteins with the nuclear magnetic resonance solution structure as initial model. Patterson search and refinement for the α–amylase inhibitor Tendamistat. *J. Mol. Biol.* **206**, 669–676 (1989).

[41] Baldwin, E.T., Weber, I.T., St. Charles, R., Xuan, J.C., Appella, E., Yamada, M., Matsushima, K., Edwards, B.F.P., Clore, G.M., Gronenborn, A.M., & Wlodawer, A. Crystal structure of interleukin 8: symbiosis of NMR and crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 502–506 (1991).

[42] Müller, T., Oehlenschlaeger, F., & Buehner, M. Human interleukin 4 and variant R88Q: phasing X-ray diffraction data by molecular replacement using X-ray and nuclear magnetic resonance models. *J. Mol. Biol.* **247**, 360–372.

[43] Wilmanns, M. & Nilges, M. Molecular replacement with NMR models using distance–derived pseudo B–factors. *Acta Cryst. D* **52**, 973–982 (1996).

[44] Shaanan, B., Gronenborn, A.M., Cohen, G.H., Gilliland, G.L., Veerapandian, B., Davies, D.R., & Clore, G.M. Combining experimetal information from crystal and solution studies: joint X–ray and NMR refinement. *Science* **257**, 961–964 (1992).

[45] Schiffer, C.A., Huber, R., Wüthrich, K., & van Gunsteren, W.F. Simultaneous refinement of the structure of BPTI against NMR data measured in solution and X–ray diffraction data measured in single crystals. *J. Mol. Biol.* **241**, 588–599 (1994).

[46] Hoffman, D.W., Cameron, C.S., Davies, C., White, S.W., & Ramakrishnan, V. Ribosomal protein L9: a structure determination by the combined use of X–ray crystallography and NMR spectroscopy. *J. Mol. Biol.* **264**, 1058–1071 (1996).

Figure 1: Effect of resolution on coordinate error in X–ray crystallography. Refinements were begun with the crystal structure of penicillopepsin [27] with water molecules omitted and uniform B–factors. The low resolution limit was set to 6 Å and the high resolution limit was truncated to the specified limit. Each refinement consisted of simulated annealing using a Cartesian–space slow–cooling protocol [28], and individual restrained B–factor refinement [29]. Refinements were carried out with 10% of the data randomly omitted for cross–validation [25]. The coordinate error is measured by the atomic r.m.s. difference between all non–hydrogen atoms of the refined structure and the original crystal structure. Coordinate–error was estimated using the σ_A method [24] in combination with cross–validation [26]. The estimated coordinate error agrees well with the actual coordinate error.

Figure 2: (a) Precision the solution NMR structures of the 56 residue IgG binding domain of protein G [30] versus completeness of the NOE-derived distance restraints. The observed distance data set consisted of 854 distance restraints derived from NOE experiments and 68 distance restraints for 34 hydrogen bonds involving slowly exchanging backbone amide protons [30]. The former contained 291 long–range interresidue restraints ($|i-j|>5$ for residue numbers i,j) that determine the tertiary fold of the structure. In addition, 105 torsion angle restraints derived from experimental NOE and three–bond coupling constant data were included in the calculations. The completeness of the distance data set was reduced by randomly eliminating distances from the full data set. Ensembles of 80 structures were obtained by a hybrid distance geometry and SA protocol [33]. Details are described in ref. [31]. The 12.5% partial data set corresponds to about two distances per residue. The precision was estimated by the rms difference for all non–hydrogen atoms between the ensemble and the mean structure. For the incomplete data sets the precision is compared with the rms difference between the mean structures of the partially complete and the complete data set. There is reasonable agreement between the two quantities. (b) View of the β–sheet region of the NMR structure of protein G obtained from the complete NMR data.

Figure 3: (a) Global statistics concerning differences between NCS–related molecules in a sample of high–resolution structures from the PDB (quality data base [34]) using the qdb program (Gerard J. Kleijwegt, Uppsala, gerard@xray.bmc.uu.se). Rms differences between all non–hydrogen atoms of the two least similar NCS–related molecules are shown. In the selected high–resolution range, are more likely to be genuine than to reflect refinement artifacts [37]. (b) Superposition of the two NCS–related molecules for the 115–residue fragment of rat mannose–binding protein [35, 36]. Shown are C^{α}–traces for the entire molecules, and the sidechain atoms and corresponding electron densities for the N–terminal region. The superposition was performed with the conserved core of the molecules. The electron density map was obtained from highly accurate MAD phases [35, 36] and contoured at 1.2 standard deviations above the mean; the electron densities for both NCS–related molecules are superposed. The large conformational differences occur in a dimerization interface wich makes an extensive contacts between the two NCS–related molecules in this particular crystal form. Note, that the *experimental* electron densities for both conformations are extremely well resolved.

Fig. 1

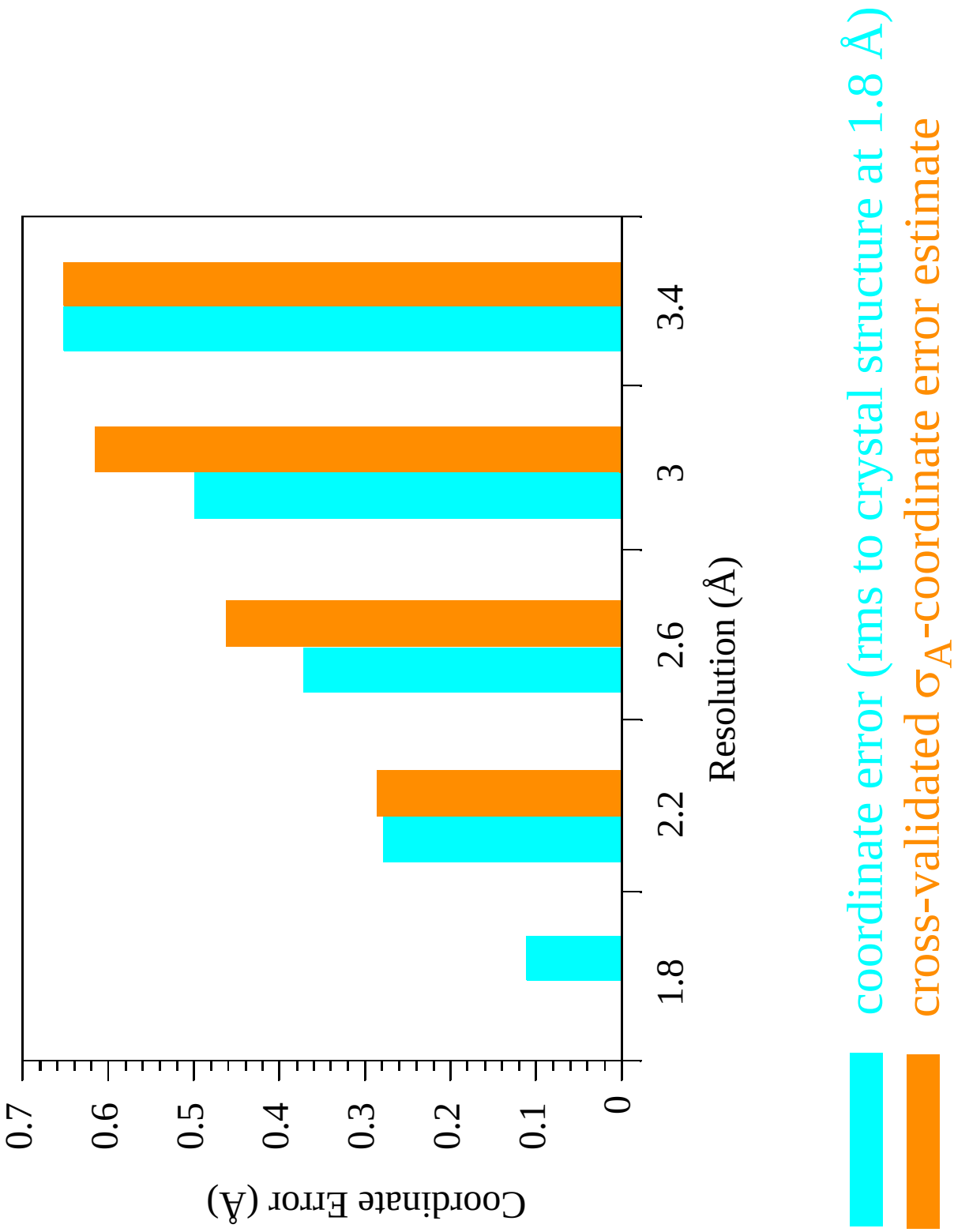

Coordinate Error (Å)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1.8
2.2
2.6
3
3.4
Resolution (Å)
coordinate error (rms to crystal structure at 1.8 Å)
cross-validated σA-coordinate error estimate

Fig. 2a

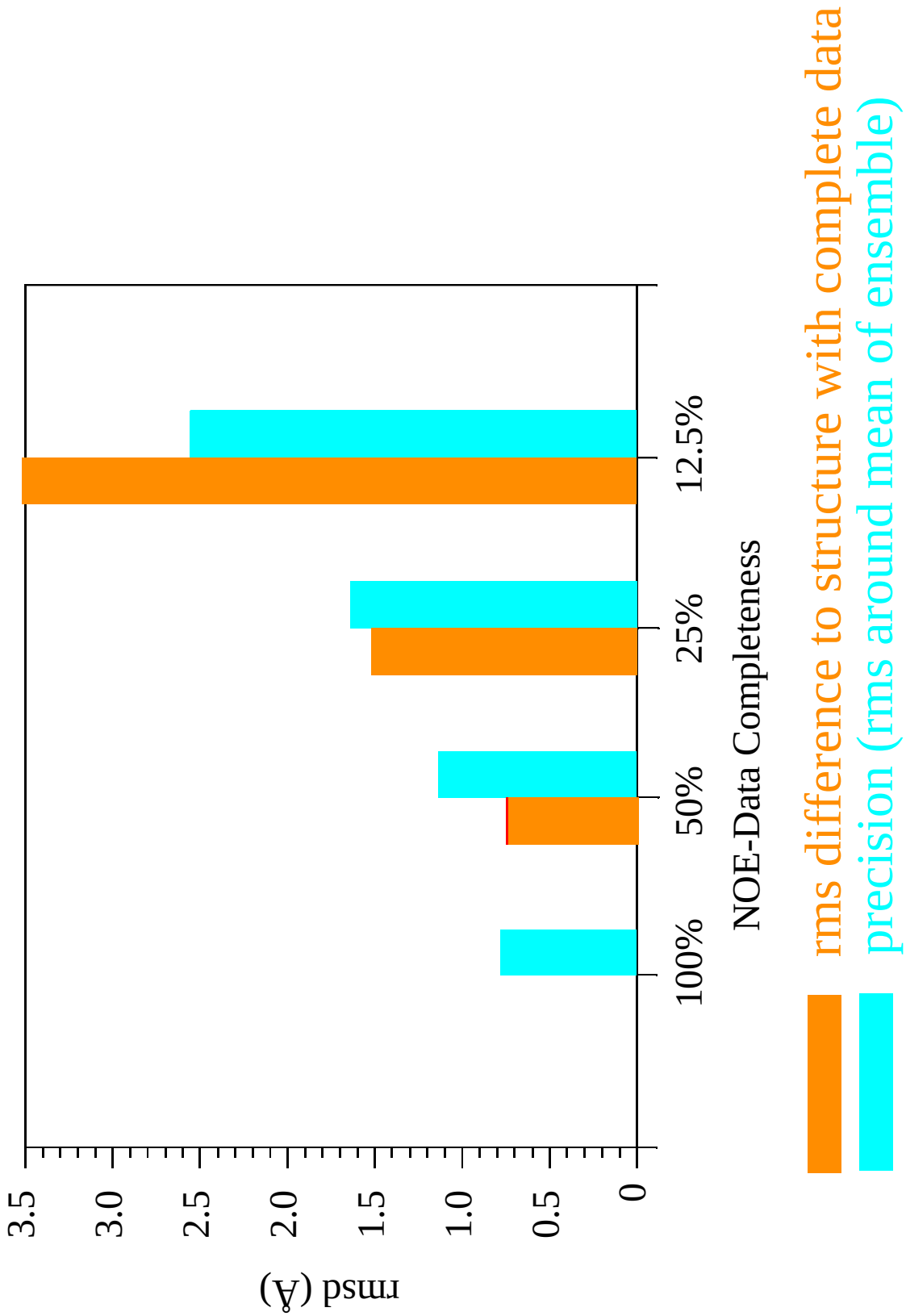

rmsd (Å)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
100%
50%
25%
12.5%
NOE-Data Completeness
rms difference to structure with complete data
precision (rms around mean of ensemble)

Fig. 2b

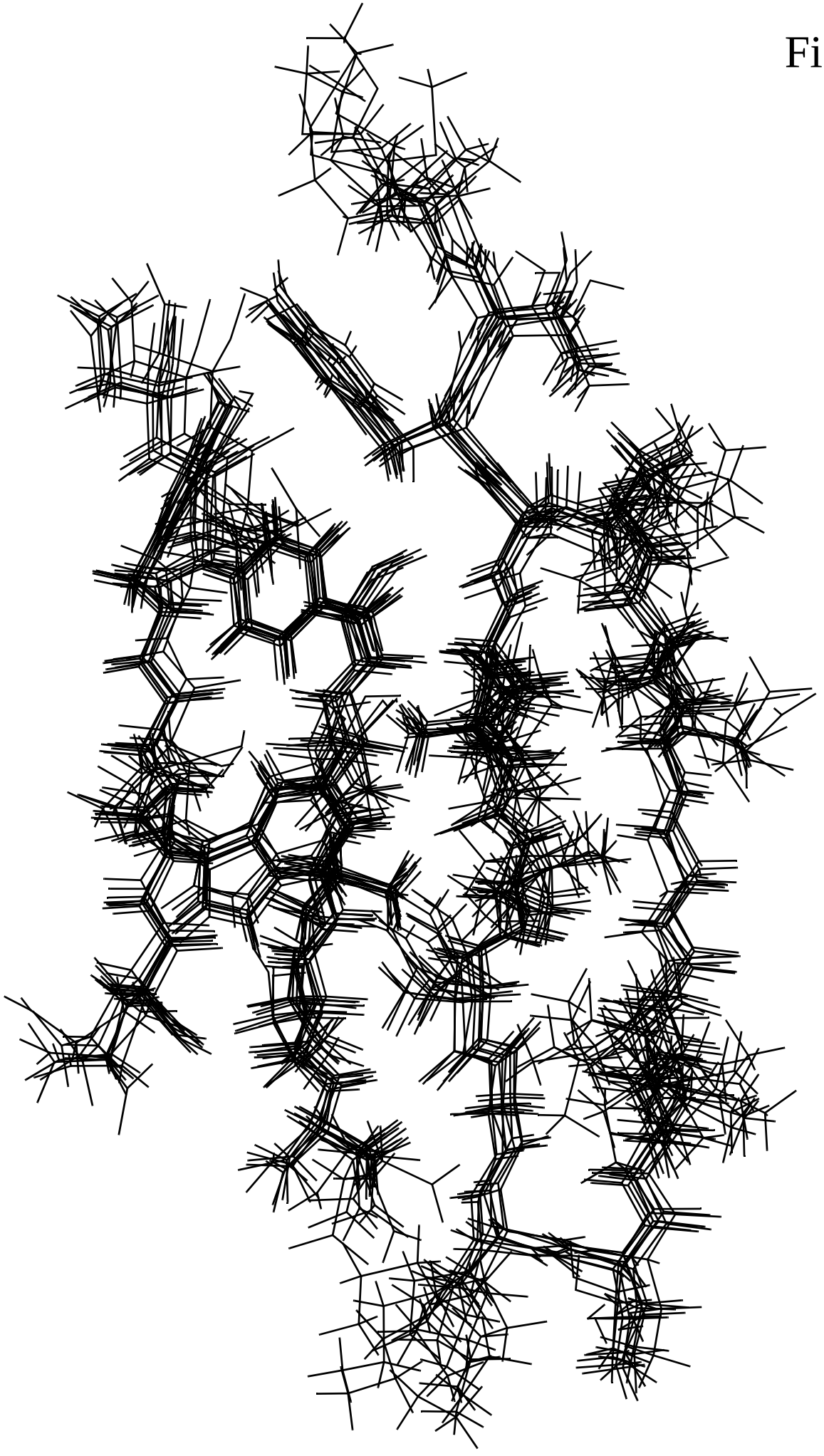


Fig. 3a

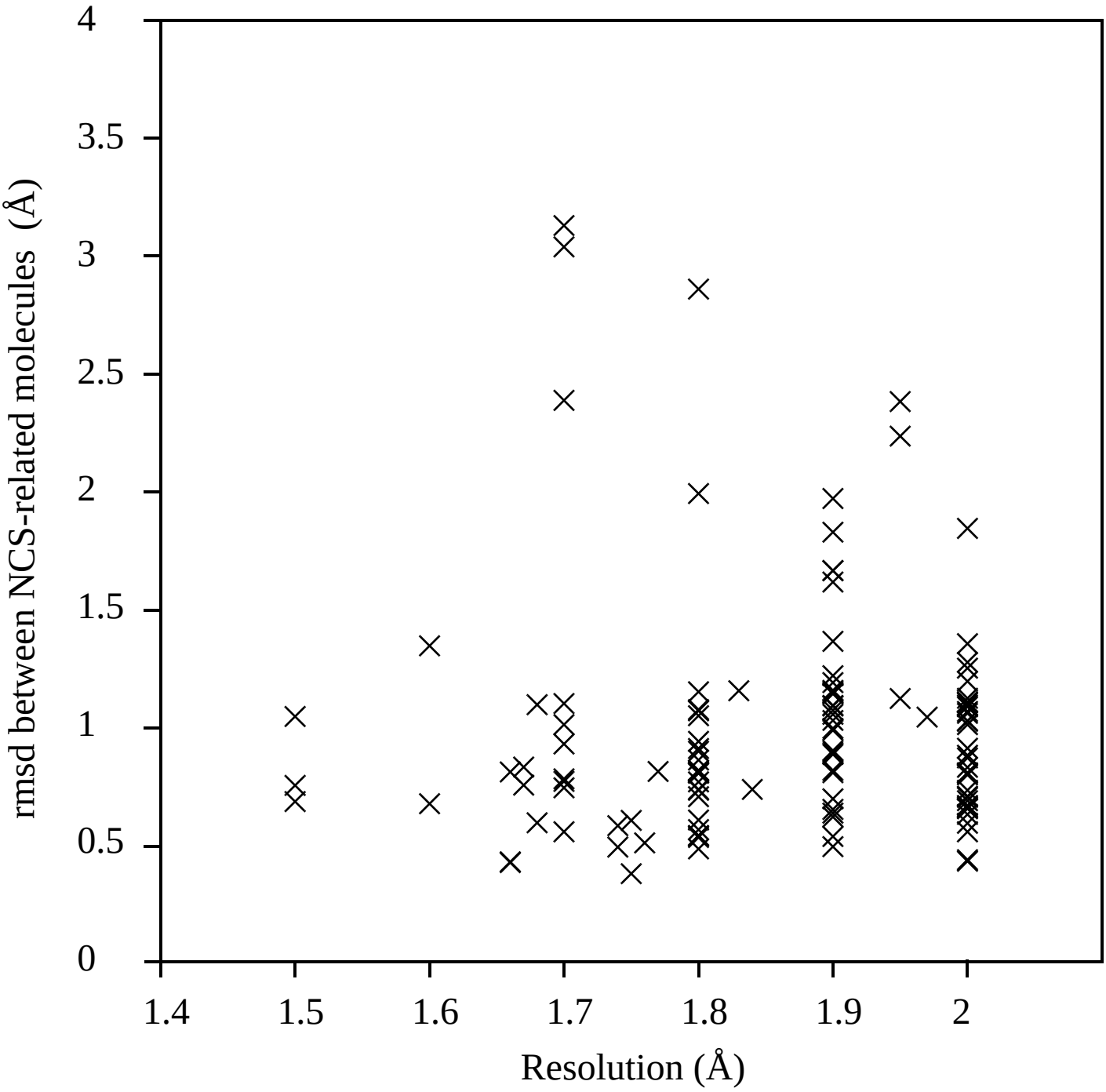
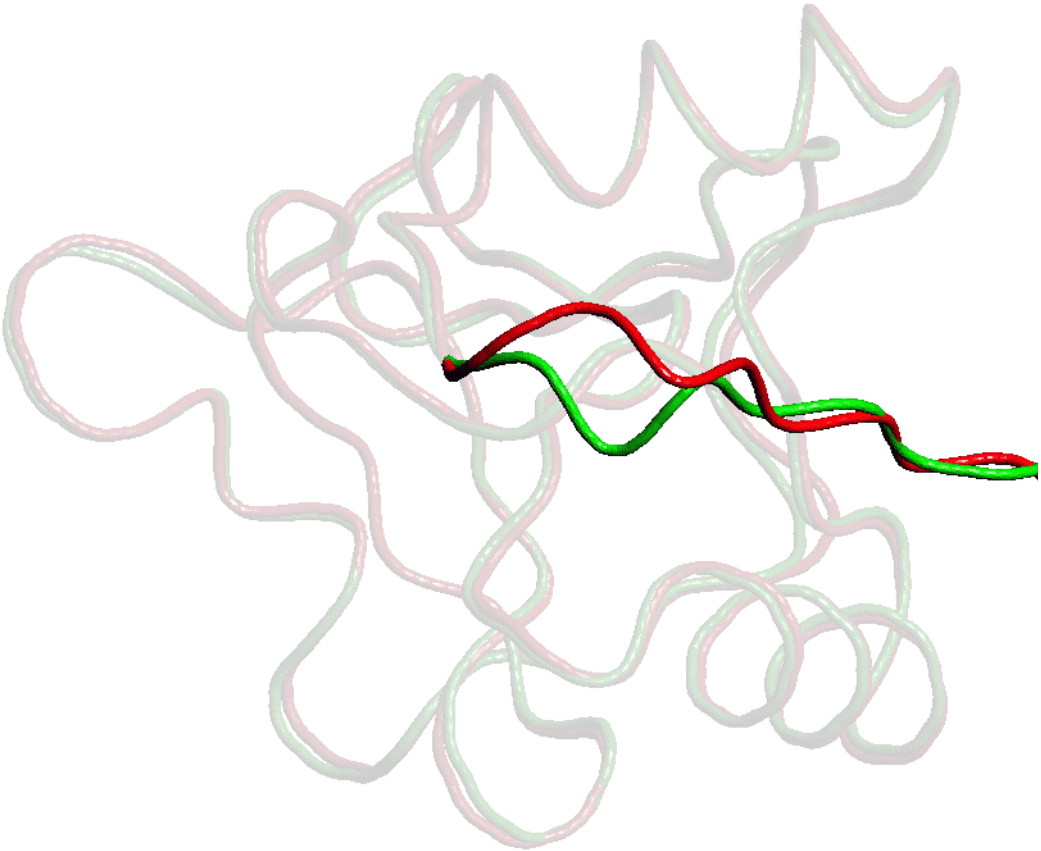

rmsd between NCS-related molecules (Å)
4
3.5
3
2.5
2
1.5
1
0.5
0
1.4
1.5
1.6
1.7
1.8
1.9
2
Resolution (Å)

Fig. 3b



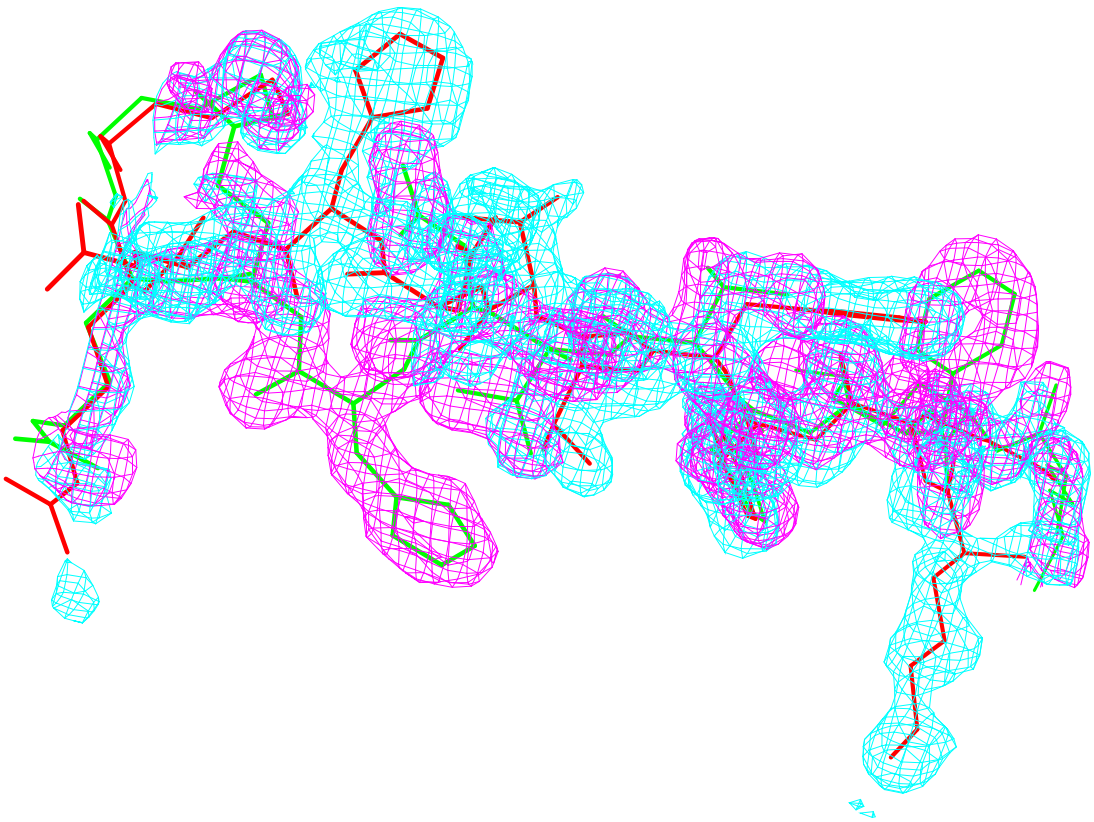


Fig. 3b