Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Apparent instability of crystallographic refinement in the presence of disordered model fragments and upon insufficiently restrained model geometry

Edwin Pozharski

Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201, USA

Correspondence e-mail: epozhars@rx.umaryland.edu

Received 14 July 2011
Accepted 23 September 2011

It is demonstrated that the crystallographic models of macromolecules may appear to diverge upon extended refinement against experimental data. Two regimes are identified for this phenomenon. Firstly, at higher resolution the apparent instability of the resulting models is shown to originate from the relatively small fraction of disordered atoms present in the initial model. Secondly, at lower resolution additional refinement instability may arise from insufficiently strong geometry restraints. The convergence of crystallographic refinement is proposed as one of the possible criteria in selecting a specific refinement strategy and in model validation.

1. Introduction

In protein crystallography, experimental data in the form of measured intensities of X-ray diffraction of individual reflections are used to deduce the three-dimensional model that represents the arrangement of protein atoms in space. Inherent to this procedure is a relatively low data-to-parameter ratio. Protein chemistry limits the set of possible atomic models (Engh & Huber, 1991; Evans, 2007; MacArthur & Thornton, 1996; Tronrud *et al.*, 2010), yet even within these confines some 'disordered' elements of the model may become under-determined (in this specific context, the model is defined as the set of equations that provide the relationship between the physical model and the experimental diffraction data). Traditionally, however, such fragments of the structural model (which contain very limited information beyond simple chemical considerations independent of experimental diffraction data) are included together with much more robust elements, leading to potential model (over)interpretation problems. The problem is exacerbated at low resolution, when the underlying data may not even contain the information regarding structural details beyond the approximate conformation of the macromolecular backbone. Several methods have recently been developed to improve the stability of crystallographic refinement at low resolution (Adams *et al.*, 2010; Murshudov *et al.*, 2011; Schröder *et al.*, 2010).

No agreement exists amongst practicing protein crystallographers regarding the most appropriate procedure to be applied to the modeling of structural disorder. It is nearly universally accepted that when large fragments that include backbone atoms (*e.g.* loops connecting secondary-structure elements and polypeptide chain termini) are 'missing from electron density' (also a subjective observation), corresponding atoms are excluded from the model as there is simply not enough information to place them with certainty.

© 2011 International Union of Crystallography
Printed in Singapore – all rights reserved

However, side chains that are 'missing' while the corresponding backbone atoms are well defined are often treated differently by including them in the model and allowing refinement to produce relatively high atomic displacement parameters (ADPs). The argument in support of this approach is that it is known from the protein chemistry that the atoms in question must be located in the vicinity (ignoring the possibility of radiation damage) and that large ADPs make the position of the atoms rather uncertain, reflecting the lack of specific information in the form of a peak in electron density. It is also said that including the disordered atoms results in more 'end-user-friendly' models, while excluding them may lead to misleading derivative results (*e.g.* electrostatic potential calculations). The counterarguments are that high ADPs only indicate increased static and/or dynamic disorder (*i.e.* increased variation in atomic positions across lattice points and in time) and not the precision of the determination of the average atomic position and that, simply put, disordered atoms are assigned average positions that are entirely unsubstantiated. Inclusion of disordered atoms often has negligible effect on the traditional measures of the model quality and may even provide some resemblance of validation by producing weak electron density which is simply a consequence of model bias. As for the distortions of the calculated electrostatic maps owing to the removal of (most often cationic) charged side chains, it is indeed incumbent upon the scientist interpreting the model to understand the limitations of such theoretical predictions.

The presence of poorly defined atoms in the model leads, as we show here, to increased instability of crystallographic refinement. The traditional measure of structural variation, the root-mean-square deviation (r.m.s.d.), tends to be sensitive to the presence of outliers (Pozharski, 2010) and thus the drift of disordered atoms that are not firmly 'anchored' to a peak in the electron-density map will produce what may be interpreted as a lack of refinement convergence. We will demonstrate that, at least at high resolution (1.8 Å), this lack of convergence may be eliminated by restricting the model only to atoms that are actually found in the electron-density map.

Another factor that may contribute to refinement instability is model bias. Depending on the starting model, the refinement may converge to a different local minimum. Particularly at low resolution, if the geometric restraints imposed in macromolecular refinement are not sufficiently tight then the system of equations that describe the structural model becomes under-determined. As a result, some model degrees of freedom (which in many cases are not simple atomic coordinates but rather combinations of these such as dihedral angles) are not effectively determined by the data, and refinement becomes partially unstable. We show that without proper restraints even the core variation of the model, as defined by the percentile-based spread (Pozharski, 2010), may drift during refinement, reducing model precision.

Crystallographic model refinement instability may potentially be viewed as resulting from the corresponding software algorithms being insufficiently robust. There appears to be an expectation that in the future clever algorithms may completely replace human judgment in the process of refining a crystal structure. Whether such complete automation may be indeed achieved remains to be seen. At present, however, the possibility of careless modeling that is incompletely justified by experimental data remains and monitoring the instability of refinement may be a useful safeguard against creating a flawed structural model.

In conclusion, we propose a test of the adequacy of a refinement protocol, which requires that refinement successfully converges. The resulting solution may still contain systematic error, or model bias owing to a suboptimal choice of parameters and the shortcomings of the standard crystallographic model, but the statistical error is thus minimized with respect to controlled parameters.

2. Methods

2.1. Assessment of refinement convergence

For every test of refinement convergence, 16 randomized models were generated by adding normally distributed noise with a standard deviation of 0.1 Å to every atomic coordinate, resulting in an initial r.m.s.d. of 0.17 Å. ADPs were also randomized, with a target standard deviation of 10% of their values. This was followed by crystallographic model refinement using *REFMAC* (Murshudov *et al.*, 2011) and *phenix.refine* (Adams *et al.*, 2010) under standard protocols. The version of *phenix.refine* used in this work provides real-space and occupancy refinement, which are activated by default. These features were deactivated to assure that both algorithms utilize the same general set of model parameters (coordinates and ADPs).

Refinement was generally performed for up to 64 cycles with *REFMAC* and for up to eight cycles with *phenix.refine*. The intermediate models were stored for analysis and then used as starting points for continued refinement. At least with *REFMAC*, this may provide slightly different results compared with continued refinement for the full number of cycles, in particular owing to updates of the bulk-solvent mask (Jiang & Brünger, 1994). However, our tests showed no detectable change in model ensemble variation between 'continuous' and 'back-to-back' protocols.

The resulting model ensembles were analyzed using two measures: the traditional root-mean-square deviation of atomic coordinates (r.m.s.d.) and the percentile-based spread (p.b.s., Pozharski, 2010). Both measures were used to characterize the variation among the models in an ensemble. Briefly, the p.b.s. is similar to the interquartile distance, which is of wide use in descriptive statistics as a measure that is less sensitive to the presence of outliers in data samples. An important difference is that the spatial distances in three dimensions are always represented by a positive number and follow a Maxwell–Boltzmann rather than a Gaussian distribution (Chambers & Stroud, 1979). Accordingly, the percentile cutoff used for the p.b.s. calculation is set to correspond to the hypothetical standard deviation assuming that no outliers are present. Thus, the p.b.s. represents the variation in atomic

positions displayed by most of the atoms. To calculate the p.b.s. of a model ensemble, the distances were calculated for every atom from the median position of that atom in the ensemble, followed by determination of the corresponding percentile (~60.8%, which corresponds to the standard deviation in the absence of outliers). This analysis was performed for the whole set of atoms as well as several groups: protein atoms, waters, and protein backbone and side-chain atoms.

The algorithm of repetitive refinement of initially randomized models and subsequent analysis of model ensembles is implemented in the *ShakErr* software, which is available from http://shakerr.sourceforge.net.

3. Results and discussion

3.1. High-resolution scenario

PDB entry 1bmb represents the crystal structure of the SH2 domain of Grb2(1) in complex with a phosphorylated peptide (Ettmayer *et al.*, 1999) at 1.8 Å resolution and was chosen because it is a relatively small protein refined against data of excellent quality. When randomized models were refined using *REFMAC*, the r.m.s.d. rapidly (within ten steps) declined from its starting value of ~0.17 Å (dictated by the amplitude of initial randomization) to the lower value of 0.06 Å (see Fig. 1). However, further refinement appeared to become somewhat unstable and the overall r.m.s.d. kept drifting upwards without reaching any apparent plateau after 64 refinement steps. To determine whether the refined model would stabilize during an extra long refinement it was extended to 1024 steps and it appeared that ~100 steps were required for convergence.

The r.m.s.d. of atoms grouped by type is shown in Fig. 2. These results obviously indicate that the instability 'resides' in the side chains and water molecules. The backbone atoms do not experience any significant increase in r.m.s.d., converging to the low value of 0.01 Å. The side-chain atoms, in contrast, significantly diverge as refinement progresses. Further analysis of the variation of individual residues showed that the increase in r.m.s.d. is associated with a limited number of side-chain atoms that may be characterized as 'disordered', *i.e.* the electron density associated with them is weak and often appears to be influenced by the random conformation assumed by the side chain, in a clear indication of model bias. When 44 of these disordered atoms were omitted from the model (together with 23 disordered water molecules described below), the side-chain and overall r.m.s.d. no longer increased upon continued refinement (see Fig. 3). These atoms constitute ~6.6% of all the protein atoms in the structural model and clearly dominate the higher end of the r.m.s.d. distribution, as shown in Fig. 4.

Water molecules present in the structure behaved similarly. The r.m.s.d. of the water molecules in the model ensemble was

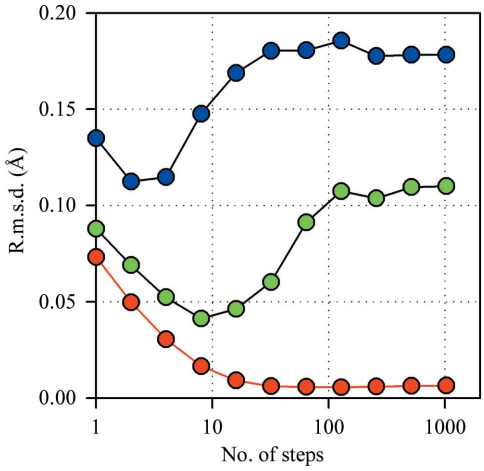

Figure 2
The r.m.s.d. of groups of atoms in the 1bmb model ensembles. Protein backbone (red), protein side chain (green) and water molecules (blue) are shown. The backbone atoms do not experience any instability and rapidly converge to the final solution. Water molecules diverge and reach the plateau within ~30 steps of refinement, while the side-chain atoms experience a much slower drift and converge in ~100 steps.

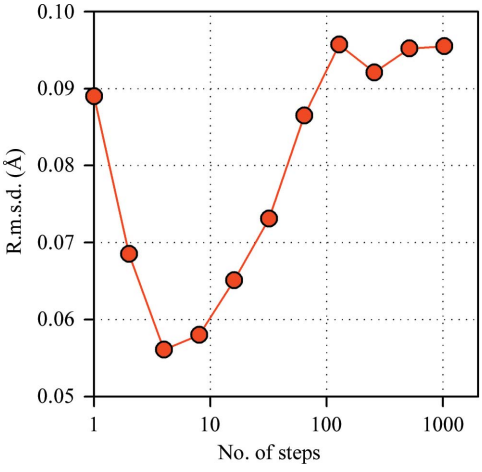

Figure 1
The all-atom r.m.s.d. of the 1bmb model ensembles *versus* the number of refinement steps. The initial convergence (~10 steps) of the randomized models is followed by a loss of stability until the disordered elements completely diverge within ~100 steps.

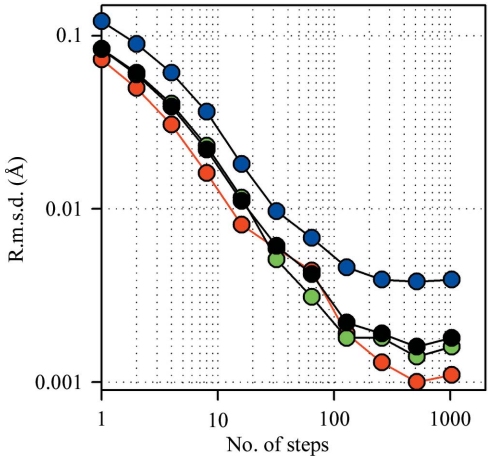

Figure 3
The r.m.s.d. of groups of atoms in the 1bmb model ensembles after removal of 'disordered' protein and solvent atoms. Overall values (black) and those for protein backbone (red), protein side chain (green) and water molecules (blue) are shown. In the absence of disordered model elements, the refinement essentially converges within ~10 steps.

as high as 0.18 Å after prolonged refinement and upon closer inspection this increase arises from 23 water molecules (∼20% of all waters) that are associated with electron-density peaks that are lower than 1σ. Two additional waters exhibited greater variation owing to being located at or near special positions and these were excluded from the analysis. When all of these waters were removed from the model, the r.m.s.d. for this group of atoms was reduced to 0.002 Å after 1024 steps of refinement.

Interestingly, the water molecules appeared to converge much faster (∼30 steps) than the side chains (∼100 steps). This is likely to be a consequence of geometric restraints that would slow down the atomic drift in the case when a specific side-chain conformation is not supported by the experimental data.

When the percentile-based spread was used to characterize the 'core' model divergence, the refinement appeared to be much more stable (see Fig. 5). This is expected since the disordered side chains may be considered to be outliers, to which the p.b.s. is not sensitive by design. An interesting observation was that a small but detectable decrease in the p.b.s. (from ∼0.004 to ∼0.001 Å after 1024 steps of refinement) occurred upon the removal of disordered side chains. The observed drop significantly exceeded the reduction in the p.b.s. prior to re-refinement (which results directly from the percentile cutoff being applied to a smaller number of atoms). This is important because the disordered side chains do not contribute to the p.b.s. directly as they consist of atoms that exhibit positional variation above the ∼60% percentile that is used for the calculation. Thus, the observed effect arises from the increased variation of the well defined elements of the model in the presence of the atoms unsubstantiated by electron density. Specifically, this means that the error associated with the poorly defined positions of these disordered atoms propagates into better defined parts of the model as refinement attempts to compensate for elements not supported by experimental data.

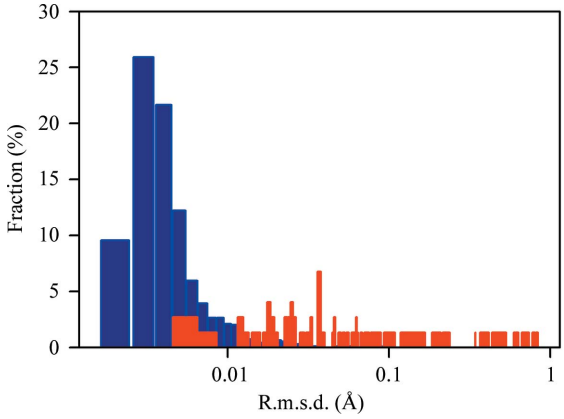

Figure 4
The distribution of r.m.s.d. values among the 'disordered' atoms in 1bmb that were removed from the model to stabilize the refinement (red). The rest of the 'stable' atoms are shown in blue. While a certain overlap between the two groups is observed, the majority of the disordered atoms are characterized by higher r.m.s.d.

An increased amplitude of initial randomization resulted in a sharp increase of the overall r.m.s.d. even for the model devoid of disordered elements (see Fig. 6). Inspection of the resulting model ensembles confirmed that the main source of increased variation is the systematic error introduced when atoms are moved outside of the convergence radius, similarly to a previously discussed (Pozharski, 2010) case of automated rebuilding (Terwilliger *et al.*, 2007). The model errors appear in all structural elements (*i.e.* backbone, side chains and water molecules). If the amplitude of initial randomization was low the resulting ensembles also have a slightly lower variation, perhaps owing to the model remaining closer to the same local minimum.

Refinement in *phenix.refine* showed very similar results that support the conclusions regarding the role of disordered

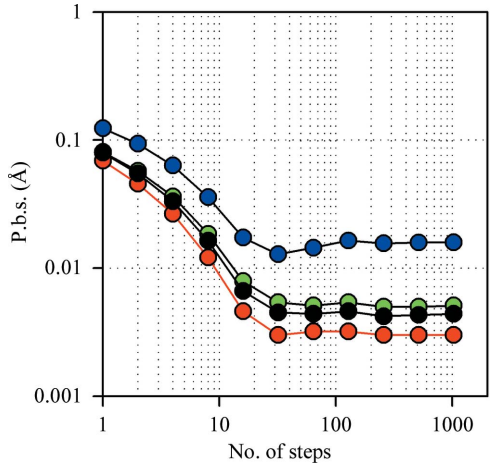

Figure 5
The p.b.s. of groups of atoms in the 1bmb model ensembles. The same groups as in Fig. 3 are shown. Since the measure used here is not sensitive to the outliers, it behaves similarly to the r.m.s.d. after the removal of disordered atoms, indicating that the more stable elements of the structural model converge within ∼10 steps.

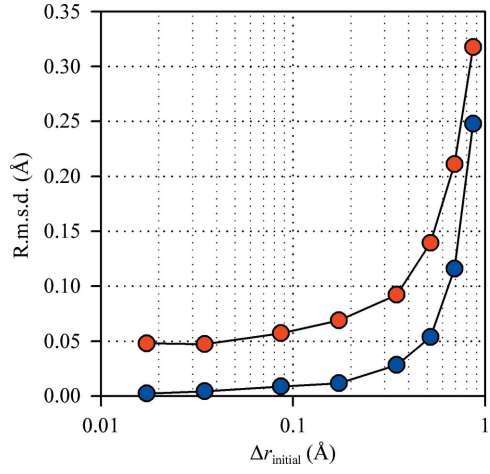

Figure 6
The r.m.s.d. of the 1bmb model ensembles after 16 steps of refinement using different amplitudes of the initial randomization. The results are shown for the full model (red) and after the removal of disordered atoms (blue). The sharp increase observed in both cases at $\Delta r_{initial} \simeq 0.5$ Å arises from the systematic model errors introduced by initial randomization.

atoms in refinement instability and improved behavior of the p.b.s. compared with the r.m.s.d. Interestingly, the 'magnitude' of the instability is somewhat lower for *phenix.refine*, despite the adjustment of the weight of geometry restraints to match *REFMAC*'s overall r.m.s. deviation of the covalent bond lengths from their ideal values. Certainly, there are sufficient differences between the implementations of crystallographic refinement in the two programs that may result in *phenix.refine* producing slower divergence of the disordered side chains over refinement progression. For instance, while *REFMAC* refines coordinates and ADPs simultaneously, *phenix.refine* utilizes separate steps for these groups of parameters. The two programs also use different approaches to formulate their maximum-likelihood targets (Murshudov *et al.*, 2011; Lunin *et al.*, 2002).

11 other structural models of small proteins at 1.8 Å resolution were analyzed in order to determine how common the behavior exhibited by the 1bmb model is. Several of these showed an identical pattern of refinement instability which can be rectified by removal of the disordered atoms [PDB entries 3lh4 (Kotsyfakis *et al.*, 2010), 2eff (Pastore *et al.*, 2007) and 1ubq (Vijay-Kumar *et al.*, 1987)]. In some cases no instability was observed for the protein and the source of the increased r.m.s.d. over refinement progression only included water molecules [PDB entries 1mku (Sekar *et al.*, 1997), 3ons (Huang *et al.*, 2011), 3gwg (Lam *et al.*, 2010) and 3lo4 (Wei *et al.*, 2010)]. Yet in some instances the refinement was entirely stable even when using the model as deposited in the Protein Data Bank [PDB entries 3ag7 (Takano *et al.*, 2010), 3ls0 (Jackson *et al.*, 2010) and 3idw (Di Pietro *et al.*, 2010)].

The results for the 12 models are summarized in the Table 1. In most cases the introduction of disordered atoms was accompanied by an increase in the p.b.s. value, suggesting that the model bias may become delocalized and affect the atomic coordinates throughout the structural model, likely resulting in slightly reduced model quality. In addition, in most cases it also resulted in a slight increase in $R_{\rm free}$. While changes in each individual case were small, averaging at ∼0.7%, the general trend is obvious. It must be pointed out that the corresponding shifts in atomic positions are small compared with the overall model error [as defined by the Cruickshank diffraction-component precision index (DPI; Cruickshank, 1999) or its maximum-likelihood-based analog (Murshudov & Dodson, 1997)].

Table 1
Statistical characteristics of the model ensembles refined under a high-resolution scenario.

All models were refined for 64 steps in *REFMAC* after initial randomization. ΔN designates the fraction of atoms deemed to be disordered and removed from every model. The $R_{\rm free}$-based Cruickshank DPI (Cruickshank, 1999), maximum likelihood-based estimate of coordinate uncertainty (Murshudov & Dodson, 1997), overall model ensemble r.m.s.d., p.b.s. (Pozharski, 2010) and $R_{\rm free}$ (Brünger, 1992) are shown. The two values in these columns represent the corresponding measure for the original protein model as deposited in the Protein Data Bank and after removing the disordered atoms.

PDB code	$N_{\rm atom}$	ΔN (%)	$\mathrm{DPI}_{\rm free}$ (10^{-3} Å)	$\sigma_{\rm ML}$ (10^{-3} Å)	R.m.s.d. (10^{-3} Å)	P.b.s. (10^{-3} Å)	$R_{\rm free}$ (%)
3lh4	1016	2.3	111/109	72.0/70.0	51.8/13.6	4.4/2.1	21.5/21.4
2eff	997	3.7	146/140	86.4/84.0	113/5.2	6.5/1.4	27.3/26.7
1cm2	926	27.2	177/170	123/121	45.2/20.4	8.1/4.6	27.0/26.1
1mku	1036	9.2	151/132	90/82	96.3/7.7	12.3/2.5	23.1/21.2
1ubq	660	10.8	158/141	92/84	73.3/18.2	6.1/1.7	25.2/23.9
1bmb	1018	6.6	111/103	63/59	88.3/15.3	4.5/2.7	19.1/18.2
3ons	666	2.7	118/114	74/74	35.2/4.9	2.2/1.4	20.8/20.5
3gwg	1120	0.5	120/121	70/71	25.6/9.8	4.6/5.1	18.2/18.4
3lo4	538	1.9	152/150	103/100	20.9/2.4	2.1/1.0	27.6/27.5
3ag7	942		145	90	12.5	5.6	23.3
3ls0	1014		116	71	20.7	7.4	21.4
3idw	576		113	86	6.6	4.1	22.2

3.2. Low-resolution scenario

The structure of the bacterial lipocalin Blc (PDB entry 3mbt; 2.6 Å resolution; Schiefner *et al.*, 2010) was used to study the stability of crystallographic refinement at low resolution. The model itself is well refined and represents a relatively small protein, simplifying the analysis. Moreover, initial inspection of the electron density indicated that no potentially disordered elements were included in the model by the structure's authors and thus the source of instability characteristic of the high-resolution scenario discussed above has already been eliminated. In a few other test cases that were investigated (data not shown) the initial models did contain disordered elements, which, as expected, behaved similarly to the high-resolution case by inflating the r.m.s.d. and making refinement unstable. Notably, in several instances structural elements were identified that could fit into the electron density in alternate conformations that cannot be distinguished given the underlying resolution. This leads to a generally increased r.m.s.d./p.b.s. of the final model ensemble, but the speed with which refinement converges is not affected.

Unlike the high-resolution scenario, in this case both of the measures of the ensemble divergence, the r.m.s.d. and the p.b.s., showed significant instability as refinement progresses (see Fig. 7). Interestingly, *phenix.refine* again showed much more stable behavior, unless the weights of the geometric

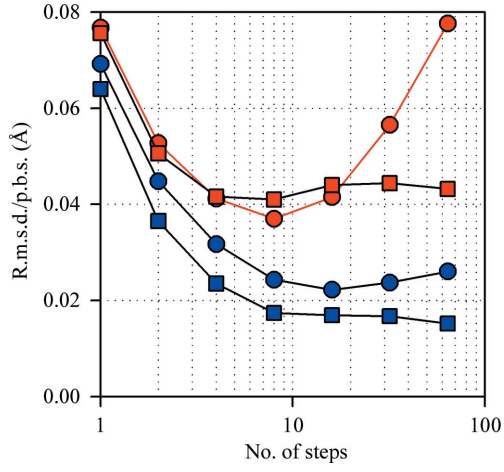

Figure 7
The overall r.m.s.d. (red) and p.b.s. (blue) in the 3mbt model ensembles with loose (circles) and tight (squares) restraints. Initially, the r.m.s.d. drops to approximately the same level, but when the restraints are not sufficiently tight the refinement becomes unstable.

restraints were adjusted to produce an overall r.m.s. deviation of covalent bond lengths from ideal values similar to that produced by *REFMAC*. This observation clearly indicated that at low resolution the algorithm for the automatic determination of the weights as implemented in *REFMAC* tends to under-restrain the model, while producing a bond r.m.s.d. that corresponds to the variation found in the high-resolution structures (∼0.02 Å). When the weighting was adjusted to produce a lower bond r.m.s.d. at ∼0.006 Å the refinement stabilized, as shown in Fig. 7.

4. Conclusions

The present study of the stability of macromolecular crystallographic refinement demonstrates the difficulty of using the conventional measure of structural variation, the r.m.s.d., in the characterization of the divergence of the model ensembles produced from initially randomized models. For instance, it appears as if the models at high resolution diverge significantly; however, most of the atoms retain their positions much better than the overall r.m.s.d. seems to suggest. In fact, all of the refinement instability as defined by the r.m.s.d. can be assigned to a small fraction of disordered atoms. The p.b.s., on the other hand, shows a clear distinction between this case and the under-restrained model at low resolution, where atomic positions converge poorly throughout the structure.

The process of crystallographic refinement allows a multitude of decisions that are often left to the discretion of the person refining the model. Philosophically, this is justified by the notion that a presented model is an interpretation of the experimental data. Structural biologists should be free to interpret their data to their liking (within certain boundaries as supported by the available data). Also, since experimental data deposition is mandatory, others can subsequently reinterpret the data and verify to what extent the presented conclusions were justified. Here, we propose using the convergence of the refinement process as one criterion that can guide some of these decisions.

For example, we have shown that under-restraining a structural model at lower resolution results in increased variation among alternatively refined models. This means that the statistical errors have increased and stronger restraints may be in order. It must be emphasized that while refinement convergence does not assure the absence of systematic errors, lack of convergence clearly indicates that the refinement protocol is somewhat inadequate. The importance of the proper choice of the weight of geometric restraints has been well established and what we propose here is an objective criterion to help guide this and similar decisions. Certainly, it should be used only in combination with other well established approaches, such as improvement in the R values and resolution-dependent expectation of the model adherence to ideal geometry.

In addition to guiding various aspects of refinement protocols, model ensembles can be useful in identifying problematic regions of a structural model. An atom which is systematically placed in a different spatial position when refined from a different starting point clearly has its coordinates determined with lower confidence. While elevated variation alone does not necessitate an atom's removal from the model, obvious outliers need to be inspected. Importantly, the p.b.s. provides a much better estimate of the characteristic level of model variation than the r.m.s.d., as the latter may be significantly influenced by the presence of outliers. For instance, the r.m.s.d. of water molecules in the full model of the SH2 domain of Grb2(1) (high-resolution scenario) was ∼0.18 Å but was reduced by an order of magnitude once the disordered waters were removed. This significant inflation of the r.m.s.d. is produced by a minority of 'rogue' water molecules and may lead to an obviously erroneous conclusion that the limit imposed on model precision by instability of the refinement process is very high (*i.e.* the inaccuracy of the positions of the water molecules in the structure is more than 0.2 Å). Notably, the p.b.s. indicates that the majority of water molecules are refined within 0.015 Å precision, a value that is not significantly affected by a few outliers. It is important to emphasize that instability of the refinement is definitely not the only source of model error (and may in fact provide only a minor contribution when refinement is performed properly).

Another important observation is that with a proper refinement protocol the level of variation among the converged models is much lower than the model error as estimated by global measures of model uncertainty such as the Cruickshank DPI or its maximum-likelihood-based analog. The variation among the converged models may be ascribed to 'model bias', *i.e.* the tendency of the minimum of the target function to be affected by the current model. Our observations indicate that this model bias is not the dominant source of model error in protein crystallography.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chambers, J. L. & Stroud, R. M. (1979). *Acta Cryst.* B**35**, 1861–1874.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Di Pietro, S. M., Cascio, D., Feliciano, D., Bowie, J. U. & Payne, G. S. (2010). *EMBO J.* **29**, 1033–1044.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Ettmayer, P., France, D., Gounarides, J., Jarosinski, M., Martin, M. S., Rondeau, J. M., Sabio, M., Topiol, S., Weidmann, B., Zurini, M. & Bair, K. W. (1999). *J. Med. Chem.* **42**, 971–980.
Evans, P. R. (2007). *Acta Cryst.* D**63**, 58–61.
Huang, K. Y., Amodeo, G. A., Tong, L. & McDermott, A. (2011). *Protein Sci.* **20**, 630–639.
Jackson, S. A., Fagerlund, R. D., Wilbanks, S. M. & Eaton-Rye, J. J. (2010). *Biochemistry*, **49**, 2765–2767.
Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
Kotsyfakis, M., Horka, H., Salat, J. & Andersen, J. F. (2010). *Mol. Microbiol.* **77**, 456–470.
Lam, K. H., Ling, T. K. & Au, S. W. (2010). *J. Bacteriol.* **192**, 2324–2334.
Lunin, V. Y., Afonine, P. V. & Urzhumtsev, A. G. (2002). *Acta Cryst.* A**58**, 270–282.
MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.

Murshudov, G. N. & Dodson, E. J. (1997). *CCP4 Newsl. Protein Crystallogr.* **33**, 31–39.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Pastore, C., Franzese, M., Sica, F., Temussi, P. & Pastore, A. (2007). *FEBS J.* **274**, 4199–4210.
Pozharski, E. (2010). *Acta Cryst.* D**66**, 970–978.
Schiefner, A., Chatwell, L., Breustedt, D. A. & Skerra, A. (2010). *Acta Cryst.* D**66**, 1308–1315.
Schröder, G. F., Levitt, M. & Brunger, A. T. (2010). *Nature (London)*, **464**, 1218–1222.
Sekar, K., Yu, B.-Z., Rogers, J., Lutton, J., Liu, X., Chen, X., Tsai, M.-D., Jain, M. K. & Sundaralingam, M. (1997). *Biochemistry*, **36**, 3104–3114.
Takano, A., Suetsugu, N., Wada, M. & Kohda, D. (2010). *Plant Cell Physiol.* **51**, 1372–1376.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Adams, P. D., Moriarty, N. W., Zwart, P., Read, R. J., Turk, D. & Hung, L.-W. (2007). *Acta Cryst.* D**63**, 597–610.
Tronrud, D. E., Berkholz, D. S. & Karplus, P. A. (2010). *Acta Cryst.* D**66**, 834–842.
Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). *J. Mol. Biol.* **194**, 531–544.
Wei, G., Pazgier, M., de Leeuw, E., Rajabi, M., Li, J., Zou, G., Jung, G., Yuan, W., Lu, W.-Y., Lehrer, R. I. & Lu, W. (2010). *J. Biol. Chem.* **285**, 16275–16285.