

Chapter 4 Quantitative study of the effects of chemical shift tolerances and rates of SA cooling on structure calculation from automatically assigned NOE data.

4.1 Introduction

NOESY spectra of even small proteins often contain significant signal overlap resulting in multiple assignment options for the majority of cross-peaks. If the average number of NOE assignment options per peak, n_{av} , is too large, the simulated annealing process (§ 1.2.1.2) does not converge. In the previous Chapter, we have seen that n_{av} depends on the setting of two parameters, the chemical shift tolerance window Δ and the cut-off n_{max} (§ 3.2.2.4). Reducing Δ reduces the number of assignment options, but introduces the risk of excluding correct assignments. Hence, Δ can be reduced only to a certain extent (§ 3.3.4). An alternative way of reducing the average number of assignments per peak is to reject all peaks for which there are more than n_{max} assignment options. For good datasets of small proteins (50-130 amino acids), a compromise can usually be found between degree of assignment ambiguity and number of peaks rejected by means of n_{max} . However, with increasing protein size, the increased signal overlap due to the sheer number of protons present results in a dramatic increase in average number of assignments per peak. The same problem affects datasets of smaller proteins where alignment of the frequencies in the list of resonance assignments and the NOESY peak list is

particularly poor (e.g., in cases where peaks are shifted due to experimental effects (such as sample heating in a TOCSY experiment) or where there are small differences between samples used for the collection of different spectra (as in the case of the PB1 domain, § 3.3.1 and 3.3.3)). In these cases, any attempt to reduce the number of assignment options by means of n_{\max} results in the rejection of too many NOESY peaks, leaving insufficient distance restraints to define a high-resolution structure. For this reason, a strategy is needed for handling peaks with highly ambiguous assignments, rather than simply rejecting them.

The success of previous workers in obtaining high-quality structures for the melanoma inhibitory activity (MIA) protein by using a slow cooling phase during simulated annealing (§ 1.4.2.4) to handle highly ambiguous NOE peak lists⁵⁸ showed that the rate of cooling is an important parameter which influences convergence in calculations from very ambiguous data. Allowing more time for thermal equilibration minimises the risk of the structure becoming trapped in local energy minima with insufficient kinetic energy to escape. Slow cooling therefore increases the probability of finding the global minimum of the target function.

This idea was expanded to study, systematically, the effects of slow-cooling on calculations carried out using automatically assigned NOE data from ARIA³²⁻³⁶, starting from peak lists with different chemical shift tolerances Δ applied, according to the quality of the datasets and the sizes of the proteins. The aim of this part of the work is a quantification the benefits of slow-cooling on convergence when very ambiguous or very incomplete peak lists are used.

4.2 Results and discussion

4.2.1 Effect of SA cooling rate on the accuracy of structures obtained with increased Δ

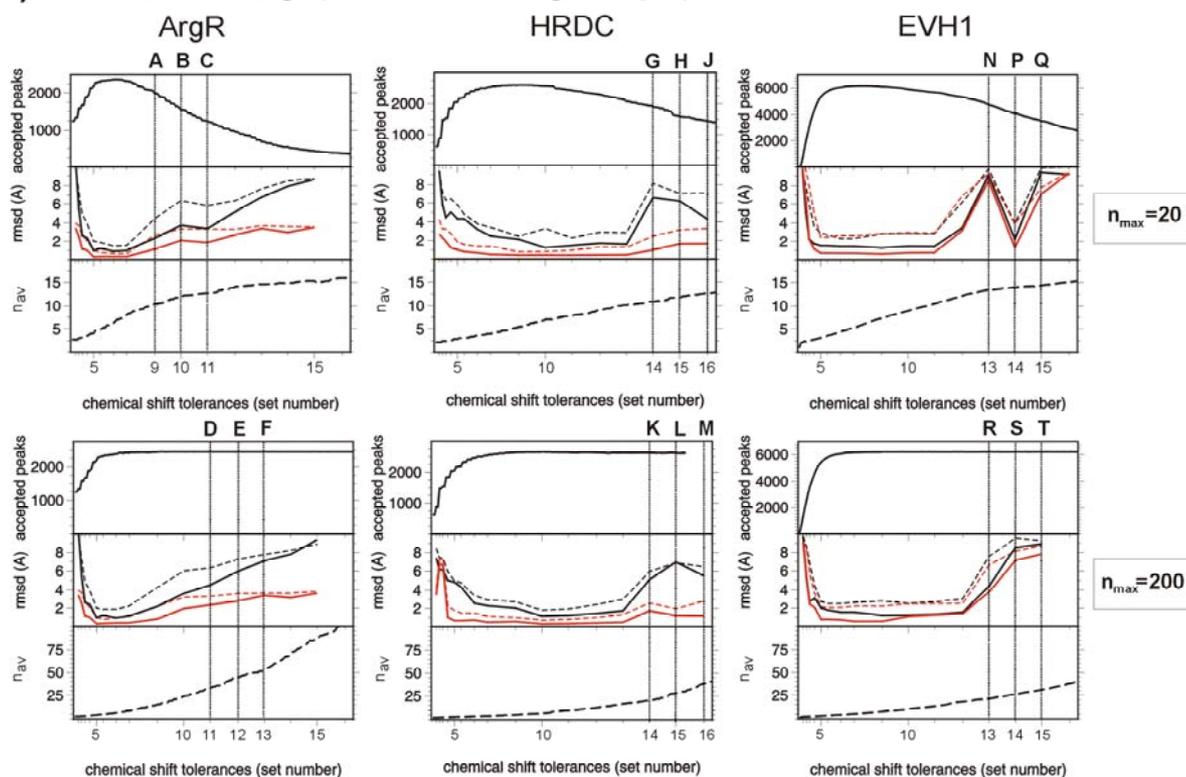
The lists of resonance assignments and the peak lists from three proteins, ArgR, HRDC and

EVH1 were used as test input data (see Table 2.1).

In Chapter 3, the effects of different chemical shift tolerances Δ on the accuracies of structures obtained using ARIA were systematically investigated. In all these calculations, the ARIA default simulated annealing protocol was used, consisting of 9,000 cooling steps (5,000 for the first and 4,000 for the second cooling phase, respectively). For each of these datasets, two sets of structure calculations with two values of the parameter n_{\max} (20 and 200) were performed, in which only the sizes of the tolerance window Δ were varied (see Table 3.1). With $n_{\max}=200$, practically all peaks were included in the calculation, regardless of their number of assignment options. The results of these previous calculations for ArgR, HRDC and EVH1 are summarised again in Figure 4.1a.

When using large values for Δ , the increased ambiguity led either to the rejection of large numbers of peaks in the case where n_{\max} was limited to 20 (Figure 4.1a, top panel) or a dramatic increase in n_{av} in the case where this limit was relaxed (Figure 4.1a, bottom panel). In all cases, above a certain critical Δ , the number of rejected peaks in calculations with $n_{\max}=20$ and the number of assignment possibilities per peak in calculations with $n_{\max}=200$, eventually prevents convergence to accurate structures. In this part of the work, the effect of slower cooling rates on datasets of different quality was systematically investigated, to assess whether it may be possible to improve convergence at higher Δ . For each domain, three sets of Δ values were selected for which unsatisfactory structures had been obtained using the standard 9,000 SA cooling steps. The selected sets are indicated by the vertical lines (labelled A to T) in Figure 4.1a. The calculations were repeated four times with increasing numbers of cooling steps to a maximum of 144,000 (see Table 2.2). In the case where n_{\max} was limited to 20 (Figure 4.1a, top panel), many peaks were thus rejected and peak lists were very incomplete. Despite this, high quality, well-converged structures could still be obtained, provided that the number of cooling steps was sufficiently increased.

a) Fast cooling (9000 cooling steps)



b) Slower cooling improves convergence

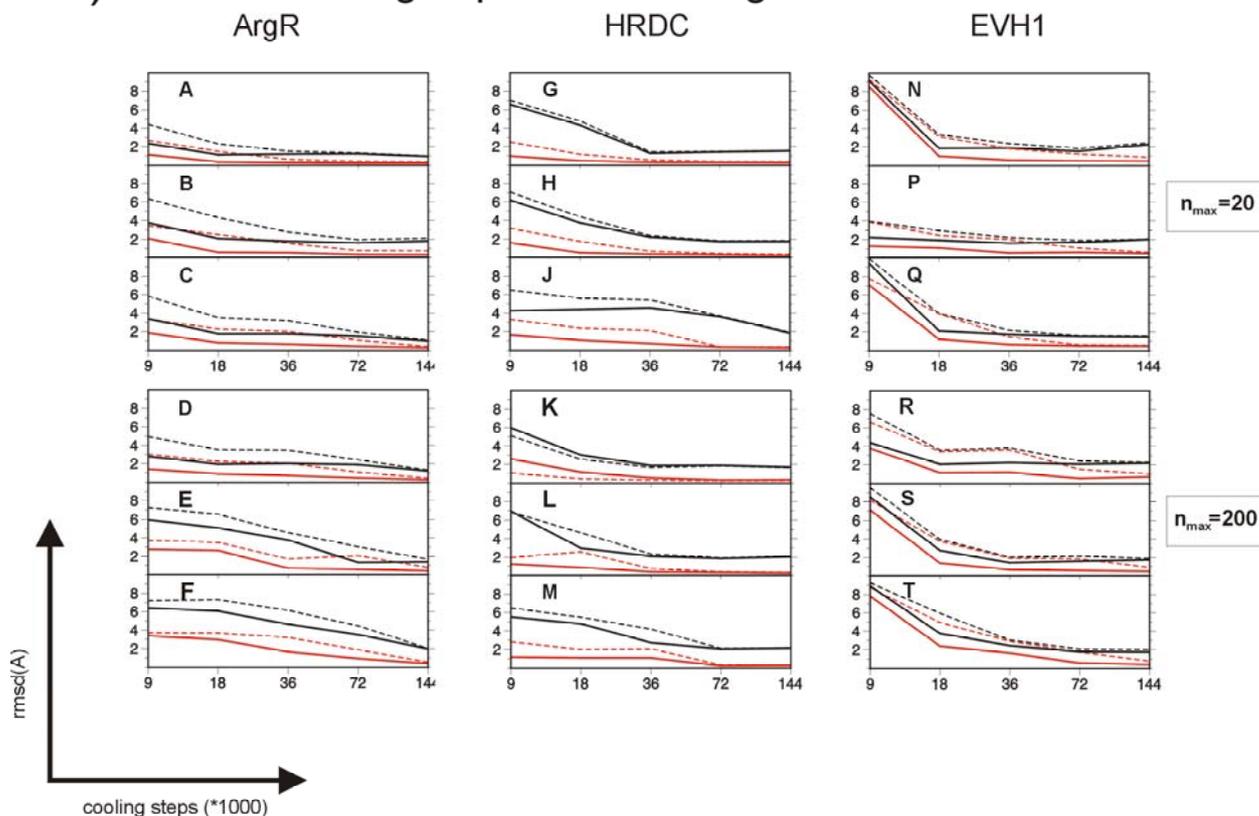


Figure 4.1 (a) Effects of chemical shift tolerances Δ on calculations with 9,000 cooling steps. For each protein, two plots are presented corresponding to $n_{\max}=20$ (upper plots) and $n_{\max}=200$ (lower plots). Each plot contains

three panels, representing three different parameters as a function of Δ . The top panels show the total number of peaks accepted by ARIA in the calculations. The middle panels show (i) the accuracy (black lines) and (ii) the precision (red lines) of the calculated structures relative to the reference structures. These rmsds were calculated for the 20 lowest-energy structures (solid lines) and for all the 100 final structures (dashed lines). The lower panels show the average number of assignment options per peak (n_{av}). Numbers on the x -axis represent the Δ -sets given in Table 3.1, with the trend from left to right being from narrower to wider tolerances. The three capital letters on the top of each plot indicate calculations using larger chemical shift tolerances, for which poor quality structures were obtained. (b) Effect of slow cooling on the eighteen calculations indicated by capital letters in (a). Black and red, solid and dashed lines are as in (a). The top and bottom panels show calculations performed with $n_{max}=20$ and $n_{max}=200$ as in (a). The plots show consistently that, from the same NOE input data and Δ -sets, greatly improved structures can be obtained if the number of cooling steps is increased. In most cases, 72,000 cooling steps were sufficient to obtain accurate and precise high-resolution structures. In addition, with slower SA cooling, the dashed and solid lines converge, i.e. the statistics for the entire ensemble become comparable to those for the best 20 structures. This reflects the higher percentage success rate of the algorithm when slower cooling is used.

The results of calculations carried out with $n_{max}=20$, for Δ values labelled: A, B, C (ArgR); G, H, J (HRDC); and N, P, Q (EVH1) are summarised in the upper panel of Figure 4.1b. It can be seen that simply doubling the default number of 9,000 cooling steps can lead to dramatic improvements in rmsd to the reference structure, as in the cases of ArgR (2D dataset) and the EVH1 domain (3D dataset). The HRDC domain (3D dataset) showed comparable improvements with 36,000 steps. In the case where n_{max} was effectively unlimited (i.e. set to 200), (Figure 4.1a, bottom panel), similarly dramatic improvements in the final structures were obtained. In this case, increasing the number of cooling steps enabled the ARIA software to cope with the larger number of assignment options per peak, rather than simply excluding large numbers of highly ambiguous peaks from the calculations. The results of calculations carried out with $n_{max}=200$, for Δ values labelled: D, E, F (ArgR); K, L, M (HRDC); and R, S, T (EVH1) are summarised in the lower panel of Figure 4.1b. For all 18

sets of calculations, regardless of whether n_{\max} was restricted or not, the same input data which gave badly defined structures with 9,000 cooling steps produced structures with greatly improved rmsd to the published, reference structures ($< 2 \text{ \AA}$), when the number of cooling steps was increased.

A closer look at Figure 4.1a (bottom panel) allows quantification of these benefits. When $n_{\max}=200$, no peak is rejected due to n_{\max} . When Δ was increased over a certain critical value, peak lists became too ambiguous and calculations produced incorrect results. With the default value of 9,000 cooling steps, ARIA produced unsatisfying results already with Δ -set 9 for ArgR ($n_{\text{av}}=15.4$), Δ -set 14 for HRDC ($n_{\text{av}}=20.8$) and Δ -set 13 for EVH1 ($n_{\text{av}}=22$) (see Table 3.1 and the bottom panel in Figure 4.1a). Nevertheless, with sufficiently slow cooling, ARIA handled much larger values of n_{av} , such as 52.0, 39.3 and 32.1, for ArgR, HRDC and EVH1, respectively, without detrimental effects (see Figure 4.1b, calculations F, M and T). Notable is the case of ArgR, where a simple increase in the number of cooling steps from 9,000 to 144,000 enabled ARIA to handle 3.5 times more average ambiguities per peak. In calculations with limited ambiguity ($n_{\max}=20$) (Figure 4.1a, top panel), incorrect structures were obtained when Δ was increased over a critical value because too many peaks were rejected due to the imposed limitation on n_{\max} . With 9,000 cooling steps, poor results were obtained already with Δ -set 9 for ArgR ($n_{\text{av}}=10.4$, 18.7 % rejected peaks), Δ -set 14 for HRDC ($n_{\text{av}}=10.8$, 27.4 % rejected peaks) and with Δ -set 12 for EVH1 ($n_{\text{av}}=12.7$, 14.7 % rejected peaks). When a sufficient number of cooling steps was provided, good results were obtained even with Δ -set 11 for ArgR ($n_{\text{av}}=12.7$, 48 % rejected peaks), with Δ -set 16 for HRDC ($n_{\text{av}}=12.5$, 45 % rejected peaks) and Δ -set 15 for EVH1 ($n_{\text{av}}=14.3$, 43 % rejected peaks). Taken together, the results show that slower rate of cooling during the simulated annealing process increases the general robustness of the calculation procedure, allowing for the use of more ambiguous and less complete peak lists.

Slower cooling increases the probability of obtaining the correct fold, which corresponds theoretically to the global minimum of the target function. The solid lines in Figure 4.1b show the accuracy (black lines) and precision (red lines) of the 20 lowest-energy structures from the 100 calculated in each ensemble. For comparison, the dashed lines indicate the accuracy and precision of the total ensemble of 100 structures, respectively. The general trend is that the solid and dashed lines tend to become closer and sometimes even overlay, as the number of cooling steps increases. This shows that, in general, the larger the number of cooling steps, the higher the percentage of accurate calculated structures (with backbone rmsd values within 2 Å of the published reference structures). The boxed numbers in Figure 4.2 indicate, for calculations E, L and S (representing the datasets for ArgR, HRDC and EVH1, respectively), the number of structures, out of the total 100 calculated, which were accurate to within 2 Å backbone rmsd of the respective reference structures, when calculations were performed using different numbers of cooling steps. All 100 structures in each ensemble calculated are superimposed, clearly illustrating that slowing down the cooling rate dramatically increases the robustness of the calculation procedure and hence the probability of obtaining the correct structure.

Figure 4.3 shows, with histograms, the effect of the SA cooling rate on the percentage of accurate structures obtained for each of the datasets studied. To exemplify this, three calculations carried out with $n_{\max}=200$ (calculations F (ArgR), L (HRDC) and S (EVH1)) were chosen, where the benefits of slow-cooling were particularly dramatic.

Using the standard 9,000 cooling steps, the structure calculations showed an almost zero probability of success in converging to within 2 Å of the reference structure (white bars). However, increasing the number of cooling steps resulted in a dramatic increase in the percentage of highly accurate structures.

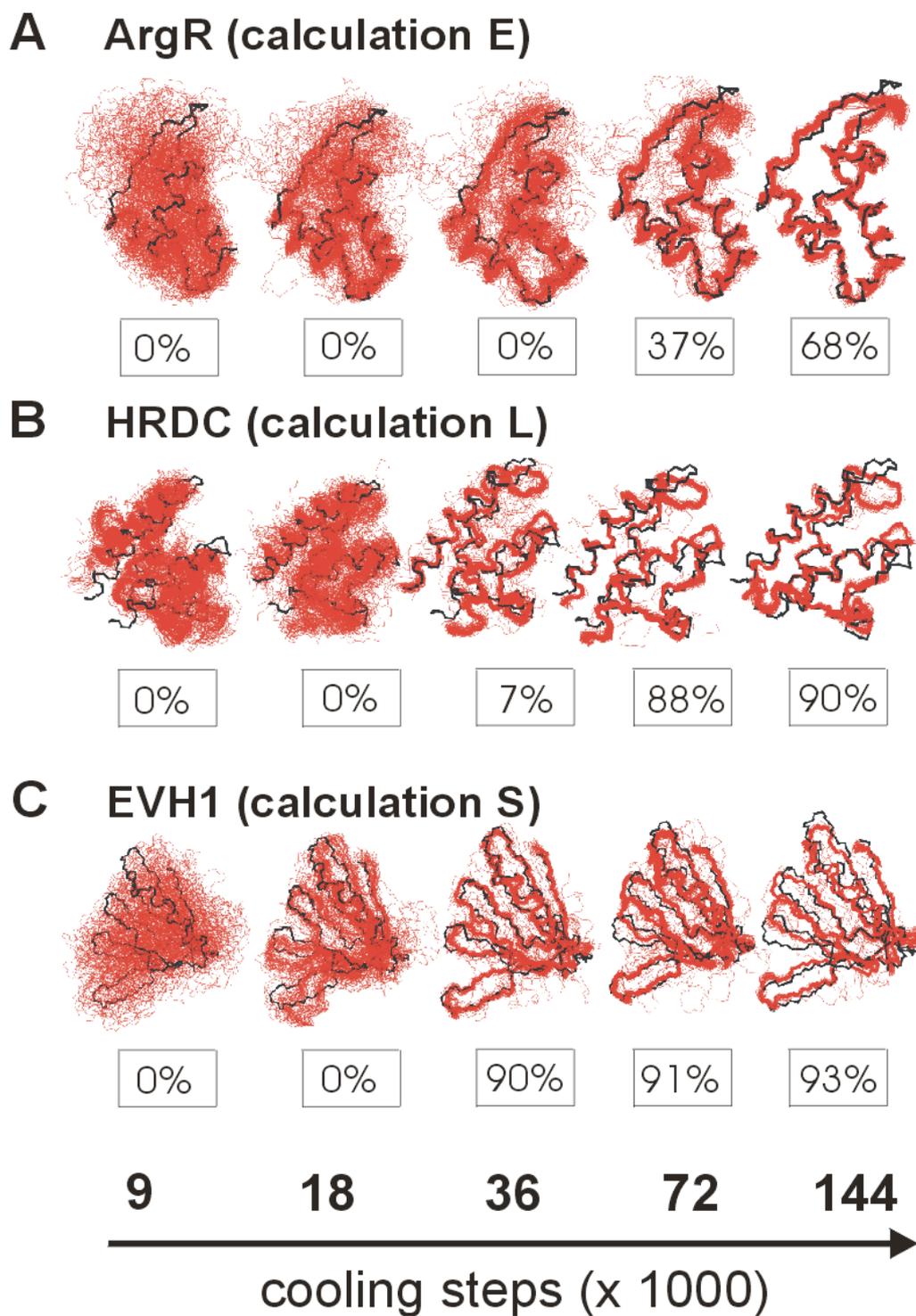


Figure 4.2 Ensembles of 100 structures obtained from calculations E (ArgR), L (HRDC) and S (EVH1) with different numbers of cooling steps. The number in the box below each ensemble represents the percentage of structures with accuracy within 2 Å of the reference structure, and shows that slower cooling dramatically increases the probability of each individual SA run leading to the correct fold.

For both 3D datasets tested, almost all the final structures were accurate to within 2 Å of the reference structure when using 36,000 cooling steps or more (mid-grey, dark grey and black bars). Even with the extremely ambiguous calculation F ($n_{av}=52.0$, highly degenerate 2D ArgR dataset), similar accuracies could still be achieved when the cooling was slowed to 144,000 steps.

It should be noted that structures calculated using very slow cooling show a tendency to appear over-restrained in regions where there are few or no experimental restraints, such as loops or terminal regions. In the absence of experimental restraints, slower equilibration leads to preferences between almost equivalent minima in the potential surface created by the *a priori* component of the target function alone (E_{chem} in Equation 1.12). Therefore, ^{15}N T_1 , ^{15}N T_2 and heteronuclear NOE measurements are particularly important to identify any genuinely flexible regions within the structure.

4.2.2 Slower cooling *versus* increased number of structure calculations

The above data shows that slower cooling increases dramatically the probability of success even when n_{av} is high and/or the NOESY input peak lists are incomplete or poorly aligned with one another. In contrast, fast cooling frequently leads to incorrect structures, regardless of the number of structures calculated. This suggests that, where high ambiguity or incompleteness of NOE peak lists prevent calculations from converging, a slow-cooling protocol may be a far more productive way to spend CPU time than simply increasing the number of calculated structures per iteration, as previously suggested¹³.

In order to verify this, an NMR dataset from the PB1 domain was analysed. These data suffered from particularly poor alignment of the frequencies in the NOESY peak lists and the list of resonance assignments, a problem that often arises when working with samples which are unstable over the period of time needed to acquire all of the experiments.

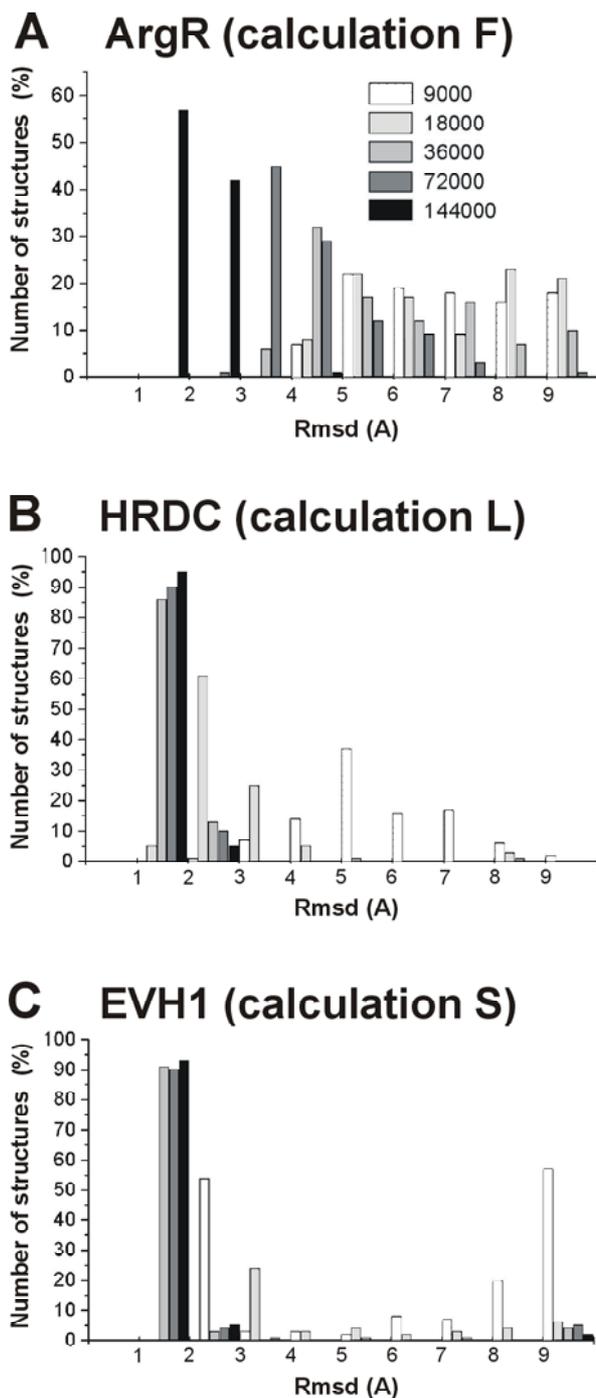


Figure 4.3 The percentage of accurate structures obtained improves with the number of cooling steps used. The probability of success of calculations increases dramatically with slower cooling. When a sufficiently large number of cooling steps were used, almost all the final structures were accurate to within an rmsd of 2 Å to the reference structure (dark bars). This is in direct contrast to similar calculations which used 9,000 cooling steps (white bars), where few, if any, of the final structures were close to the reference structure. (A) Calculation F (ArgR). (B) Calculation L (HRDC). (C) Calculation S (EVH1) (see Figure 4.1).

Any attempt to obtain *de novo* structures of this domain using a protocol with 9,000 cooling steps failed completely, even when a limited number of unambiguous, manually assigned, long-range NOEs were added as input to the calculations (§ 3.3.3). In cases like this, there is no choice but to use larger chemical shift tolerances Δ . The analysis by Cesta.py (§ 3.3.1) suggested the generous values of Δ -set 13 ($\delta^{\text{het1}}=0.88$, $\delta^{\text{pro1}}=0.07$, $\delta^{\text{pro2}}=0.035$) as a lower-limit for Δ . However, the use of these unusually large values led to excessive ambiguity (>16 assignment possibilities per peak) and very cumbersome, non-convergent calculations.

Increasing the number of structures calculated in each iteration did not result in satisfactory numbers of correct structures. 100, 200 and 500 structures per iteration (far in excess of the typical default value of 20 calculated structures) were calculated. The results are summarised in Table 4.1 and show clearly that a better strategy is needed to calculate structures from such highly ambiguous data than simply increasing the number of calculated structures. Therefore, a slow-cooling SA protocol was applied also on this dataset in order to rescue the ARIA structure calculations.

Number of structures per iteration	20	100	200	500
<i>Accuracy</i>	7.3	7.2	7.0	7.0
<i>Precision</i>	4.2	4.1	3.41	3.35

Table 4.1 Accuracy and precision of the 20 lowest-energy structures calculated for PB1. All calculations were performed using the Δ values of set 13 (Table 3.1). No significant improvement in the results was achieved by simply increasing the number of structures calculated per iteration, even when calculating as many as 500 structures. This shows that the probability of success for each SA calculation is very low when the standard fast-cooling protocol is used.

The result was a successful *de novo* automatic structure determination of the PB1 domain when the large tolerance values of Δ -set 13 were combined with 72,000 cooling steps. Using this combination, an average of 16.4 possible assignments per peak in the first iteration occurred but did not prevent convergence. The results are represented in Figure 4.4. For 9,000 cooling steps, the convergence was extremely poor (Figure 4.4A). The backbone rmsd of any individual calculated structure to the reference structure was never less than 7 Å. However, when 72,000 cooling steps were used with this same set of tolerance values, the 20 lowest-energy structures showed a precision of < 1 Å and an accuracy of < 3 Å to the reference structure (Figure 4.4B). This result is encouraging as the calculations in Figure 4.4B were performed in the absence of any additional structural information such as hydrogen bonds, dihedral angle restraints and manually assigned NOE-based distance restraints. Significantly, it shows that the correct fold can be obtained based entirely on automatically assigned NOE data even in the presence of inconsistencies within the datasets. In Figure 4.4C, the results obtained when supplementing the dataset used in Figure 4.4B with a set of experimental hydrogen bond restraints are shown. This led to further improvements in precision and accuracy (rmsd to reference of 1.45 Å). The plots in Figures 4.4B and 4.4C show clearly that correct results can be obtained only when sufficiently large values for Δ are chosen, as this is the only way to avoid exclusion of correct NOE assignments for this dataset. The wider tolerance windows led to a larger number of assignment possibilities per peak, i.e., a higher degree of ambiguity, which then requires much slower cooling to avoid the structures being trapped in local minima in the target function.

4.2.3 Time costs of the method

The use of a slow cooling phase during simulated annealing has the obvious disadvantage of increasing the CPU time needed for structure calculations. The time needed to calculate one

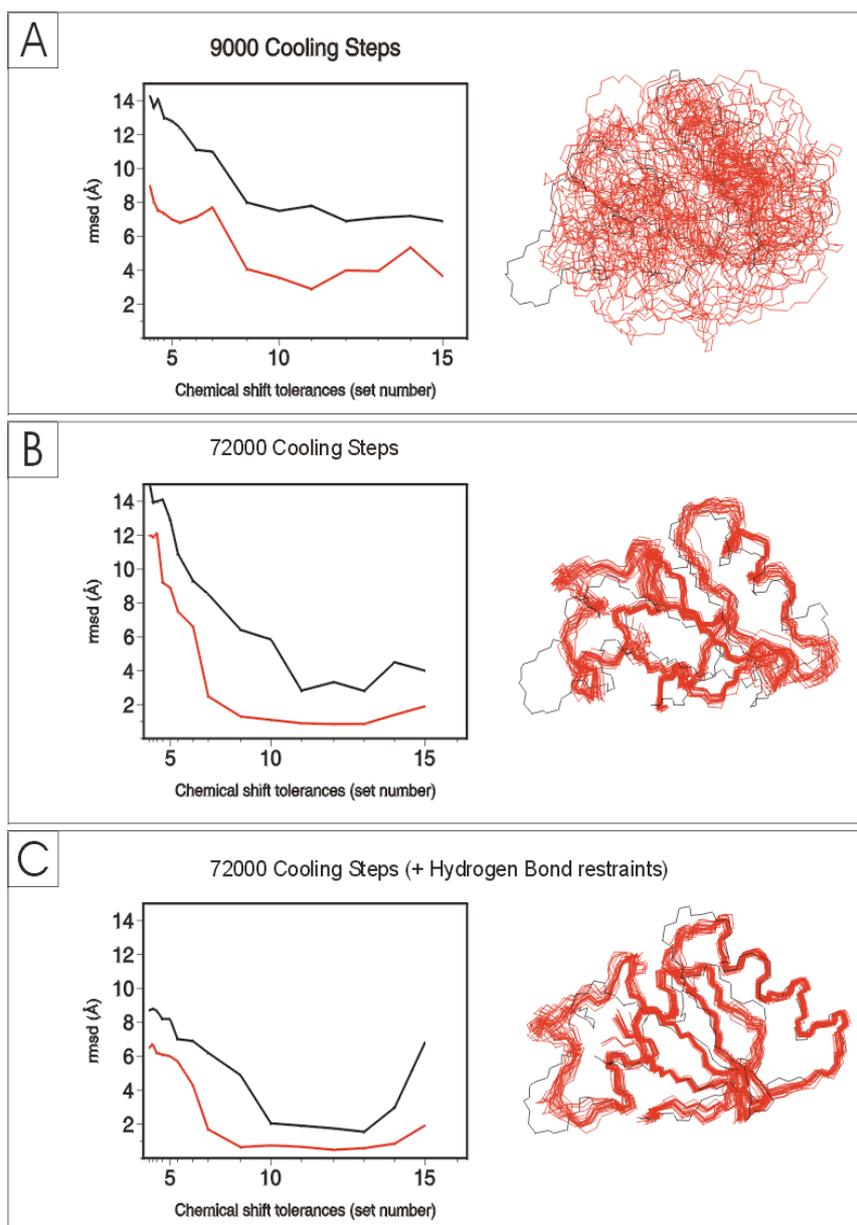


Figure 4.4 *De novo* automatic structure determination of PB1 domain of CDC24p. Dependence of the accuracy (black line) and precision (red line) of the 20 lowest-energy structures on Δ . The structure ensembles calculated with Δ -set 13 ($\delta^{\text{het1}}=0.88$, $\delta^{\text{pro1}}=0.07$, $\delta^{\text{pro2}}=0.035$) are shown on the right. (A) Calculations with 9,000 cooling steps. (B) Calculations with 72,000 cooling steps. When Δ is sufficiently large, the information contained in the dataset is very ambiguous but self-consistent. The slow-cooling yields convergent results and a correct overall fold. (C) Calculations with 72,000 cooling steps with hydrogen bond restraints added. The presence of these additional restraints restricts the conformational space to be searched and yielded very well converging and accurate results (accuracy = 1.55 for Δ -set 13) (see Table 3.1).

single structure in each of the successive ARIA iterations, when using different numbers of cooling steps, is shown in Figure 4.5A (with the example of calculation L of the HRDC domain). The initial iterations, in which the average number of assignment options per peak n_{av} is still large, are clearly the slowest, but the time needed to calculate a single structure decreases as the iterations progress, as a consequence of the decrease of n_{av} . This is illustrated by the three examples shown in Figure 4.5B. In addition, convergence to the correct overall fold is already achieved in the earlier iterations (in iteration 4 in the case of HRDC and EVH1 and in iteration 6 in the case of the highly ambiguous dataset of ArgR), as shown in Figure 4.5C.

These trends suggest that, although slow-cooling is of benefit in the early iterations, it is probably unnecessary to cool slowly in all ARIA iterations. The calculations shown in Figure 4.5B were therefore repeated using 72,000 cooling steps up until iteration 4 for HRDC and EVH1 domains, and until iteration 6 for ArgR (i.e. the points at which good folds were obtained) and then 9,000 cooling steps for the remaining iterations, during which the structures were refined. The structures showed no notable differences to those obtained when using 72,000 cooling steps for all iterations (Table 4.2).

Such modified versions of the strategy allowed substantial savings in CPU time, as five times more structures are calculated in the final iteration (iteration 8) than in all earlier iterations. Figure 4.6 shows the time required to accomplish calculation L of the HRDC domain with the different cooling schemes. The time taken for the entire calculation using a constant large number of cooling steps for each iteration (grey) is compared with the time taken using a reduced number of 9,000 cooling steps for: (i) iterations 5-8 (white); and (ii) iteration 8 only (black). The reduced-time protocols allow time savings of about 33% and 18%, respectively. On four 1.8 GHz processors, the CPU time required with 72,000 cooling steps in all iterations was 37 h, and the reduced-time version of this, using 9,000 steps for iterations 5 to 8 was 31 h.

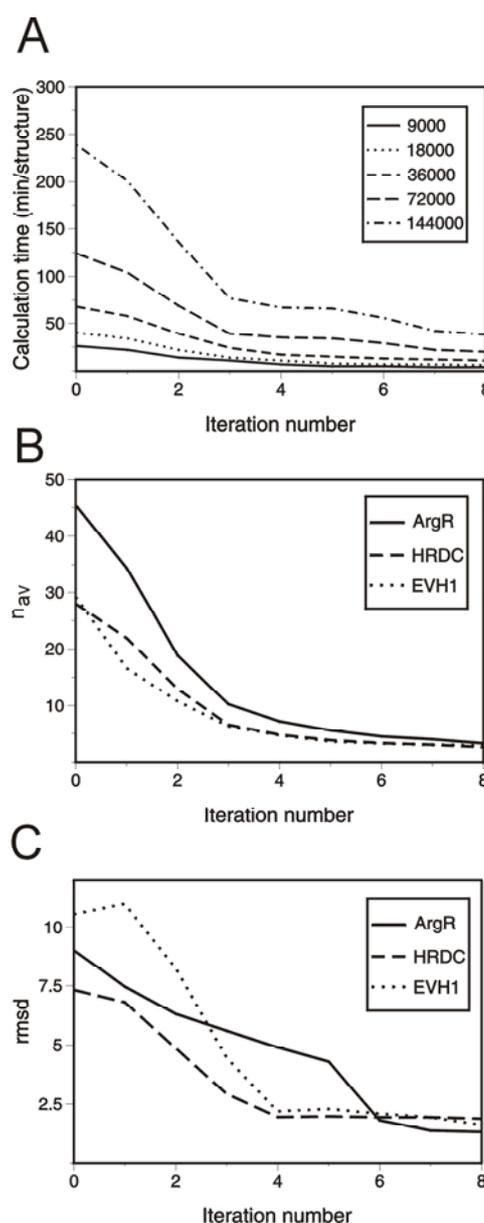


Figure 4.5 Variations of structure calculation time, n_{av} and accuracy during the ARIA iterations.

(A) Computational time required per structure at each iteration, using calculation L of the HRDC domain as an example. The differently dashed and dotted lines represent different numbers of cooling steps ranging from 9,000 to 144,000. (B) Reduction in n_{av} with successive iterations (shown for three calculations: E (ArgR), L (HRDC) and S (EVH1), corresponding to E, L, S in Figure 4.1). (C) Reduction in backbone rmsd to the reference structure with successive iterations. Rmsd were calculated using the eight lowest-energy structures (out of 20) for iterations 0-7 and the twenty lowest-energy structures (out of 100) for the final iteration 8. The correct fold is already obtained by iteration 3 for calculation L, by iteration 4 for calculation S and by iteration 6 for calculation E, due to the higher initial value of n_{av} in the latter case.

This makes the slow-cooling protocols approximately 3-4 times slower than the fast (9 h) protocol which uses 9,000 steps throughout. However, given that the former yielded high-quality, accurate structures, while the latter frequently fails with imperfect datasets, the investment of CPU time in slow-cooling is clearly time well spent. Hence, if computational time available is limited (rarely a problem with the recent generations of computers), protocols B or C are recommended. However, if time is not limiting, the slower, most robust, protocol A is to prefer.

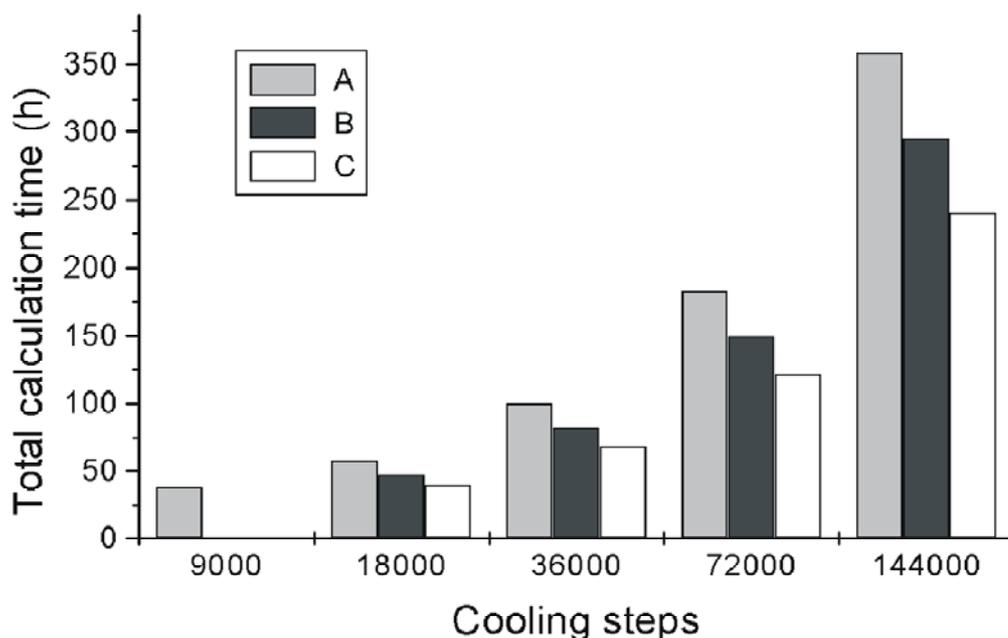


Figure 4.6 Total computational time for different cooling schemes using calculation L of HRDC as an example (see Figure 4.1): (A) constant slow-cooling (in all iterations); (B) slow-cooling for the first seven iterations, then reducing the number of cooling steps to 9,000 in iteration 8; (C) slow-cooling in early iterations until reasonable convergence was observed. The latter allows considerable time savings without noticeably affecting the quality of the results (results for the case of 72,000 cooling steps, for all NMR datasets used in these studies, are summarised in Table 4.2).

Protein		Method		
		A	B	C
ArgR (calculation E)	Accuracy	1.34	1.56	1.51
	Precision	0.56	0.62	0.61
HRDC (calculation L)	Accuracy	1.87	1.96	1.92
	Precision	0.31	0.73	0.67
EVH1 (calculation S)	Accuracy	1.60	1.62	1.62
	Precision	0.55	0.58	0.57

Table 4.2 Comparison of accuracy and precision for three different calculation strategies. Once the correct fold is obtained, slow-cooling can be replaced by faster cooling. Using three test calculations corresponding to each NMR dataset, the quality of the structures in terms of accuracy and precision are compared for three different calculation strategies: (A) 72,000 cooling steps used in each iteration of ARIA; (B) 72,000 cooling steps used in the early iterations until the calculation converges to a consistent global fold, then 9,000 in the following iterations. Good folds were observed already in iteration 3 for HRDC, in iteration 4 for EVH1 and in iteration 6 for ArgR; (C) 72,000 cooling steps used in iterations 1-7 and 9,000 in iteration 8. The structures obtained with the faster strategies B and C were only marginally less accurate and precise than those obtained with the slower strategy A (see Figure 4.6).

4.3 Conclusions

In the previous Chapter, it is shown that over-restricting the chemical shift tolerance windows Δ during automated NOE assignment can result in the exclusion of the correct assignment from the available options for a given peak and consequently induce distortions in the calculated structures. Here, it is shown that even highly ambiguous and incomplete NMR datasets can be handled by the automatic NOE assignment software ARIA if sufficient time for thermal equilibration is allowed during the SA cooling phase. For the ArgR, HDRC and EVH1 datasets tested, using larger values of Δ produced accurate and precise structures when the number of cooling steps was increased (from the default 9,000 to 72,000 steps). The

combined use of large Δ values and slow-cooling dramatically increased the robustness of the SA protocol towards assignment ambiguity. Using this approach, reliable structures were successfully calculated from data which suffered from poor peak alignment, uncertainty in chemical shift positions, or extensive resonance overlap; problems that usually plague the spectra of large proteins and complexes. The method described in this work could thus extend the role of NMR to solving the structures of larger proteins, and enhance the role of NMR in structural genomics.

In addition, it is clearly shown that, whenever the ambiguity or the incompleteness of the peak list prevent calculations from converging, a slow-cooling strategy is far more productive than previous suggestions simply to calculate many more structures in the computational time available¹³. When the input NMR data are largely complete and unambiguous, fast cooling may indeed produce accurate structures in a short time. However, with incomplete or highly ambiguous input data, the percentage success level of the calculations drops rapidly, and increases only when more generous chemical shift tolerances are coupled with slower SA cooling protocols. The higher demand on CPU time of slow-cooling protocols is well compensated by a dramatic rise of the probability of success in finding the global minimum in the overall potential during simulated annealing. As demonstrated by the PB1 domain, CPU time well spent on slow-cooling can save months of painstaking manual assignment.