Chapter 2 Material and Methods

2.1 ARIA calculations

2.1.1 Protein datasets used

Five different protein NMR datasets were used for the ARIA calculations described in Chapters 3 and 4: the C-terminal domain of Lac Repressor (Lac)[48]; the N-terminal domain of Arginine Repressor (ArgR)[49,50]; the C-terminal PB1 domain of yeast CDC24p (PB1)[1,51]; the HRDC domain of RecQ (HRDC)[52] and the EVH1 domain of human VASP (EVH1)[53,54]. Only 2D NOESY data were used for ArgR and Lac; for PB1, HRDC and EVH1, also ^{13}C- and ^{15}N-edited 3D NOESY data were employed. In all calculations, hydrogen bond and dihedral angle restraints were supplied. The main characteristics of these proteins and their spectra, relevant to this work, are summarised in Table 2.1.

2.1.2 Pre-calculation analysis by Cesta.py

The pre-calculation analysis by the script Cesta.py is described in § 3.2. The script created 165 different Δ-sets, ranging from very small (δ^{het1}=0.00062, δ^{pro1}=0.00005, δ^{pro2}=0.000025) to very large values (δ^{het1}=1.25, δ^{pro1}=0.1, δ^{pro2}=0.05) and evaluated the four diagnostic functions $N^{rej}_{noassig}(\Delta)$, $N^{rej}_{n_{max}}(\Delta, n_{max})$, $N^{acc}(\Delta, n_{max})$ and $n_{av}(\Delta, n_{max})$ (§ 3.2.2.1-4) for all five protein datasets described above. On a 1.8 GHz processor, the pre-calculation analysis

required 8-12 h of CPU time, depending on the dataset. The calculation time can be reduced by lowering the number of values to be evaluated. For example, evaluating the diagnostic functions using only 30 Δ-sets required just 90-150 minutes.

Name	**PDB entries**	**Chain length**	**Secondary Structure**	**Type of data**	**Number of peaks**
Lac	1JWL	56	3α	2D NOESY	2D (H_2O): 2122 2D (D_2O): 2106
PB1	1PQS	77	4β+2α	3D NOESY	^{13}C-edited 3D: 1909 ^{15}N-edited 3D: 766
ArgR	1AOY	78	3α+2β	2D NOESY	2D (H_2O): 1403 2D (D_2O): 1245
HRDC	1D8B	91	3α	3D NOESY	^{13}C-edited 3D: 2455 ^{15}N-edited 3D: 824
EVH1	1QC6	115	7β+1α	3D NOESY	^{13}C-edited 3D: 3506 ^{15}N-edited 3D: 2772

Table 2.1 The five different datasets used for the ARIA calculations. The NOESY peak lists of Lac, ArgR, HRDC and EVH1 were obtained with manual peak-picking, while those of PB1 were generated automatically with the internal peak-picking algorithm of Sparky v.3.1 (http://www.cgl.ucsf.edu/home/sparky/). Diagonal and obvious noise peaks were removed manually. No manual NOE assignments were included into the peak list.

2.1.3 Structure calculations

The structures were calculated by ARIA v.1.2 on a Dual Athlon M1800+ cluster at the Pasteur Institute. The number of calculated structures was 20 for iterations 0-7 and 100 for the

final iteration 8. The number of cooling steps used in each of the five cooling protocols tested in Chapter 4 is summarised in Table 2.2.

Cool 1	Cool 2	Total
5,000	4,000	9,000
10,000	8,000	18,000
20,000	16,000	32,000
40,000	32,000	64,000
80,000	64,000	144,000

Table 2.2 Number of cooling steps for the first and second cooling stage in the five SA protocols tested in the work described in Chapter 4 (Figure 1.17). The number of steps for the high temperature (2,000 K) conformational search was in all cases 10,000.

The calculations were evaluated by computing a pair-wise rmsd (precision) and an rmsd to a reference structure (accuracy) of the 20 lowest-energy structures. As a reference, the X-ray structure for Lac and EVH1 and the averaged NMR solution structure for ArgR, HRDC and PB1 were used.

2.2 SOLARIA calculations

2.2.1 Dataset used and solid-state MAS NMR spectroscopy

SOLARIA was tested on the α-spectrin SH3 domain. The main characteristics of the SH3 domain and of the spectra used in the work described in Chapter 5 are summarised in Table 2.3. The sample preparation is described in detail elsewhere[55]. Peak lists for the structure calculations were generated from solid-state 2D ^{13}C-^{13}C and 3D ^{15}N-^{13}C-^{13}C PDSD correlation spectra[20,21].

Name	PDB entry	Chain length	Secondary Structure
SH3	1SHG	62	β–sandwich

Spectrum	Dimensionality	Labelling	Observed nuclei	Number of peaks
1,3-CC	2D	[2-^{13}C]-glycerol	^{13}C-^{13}CO	589 (566)
2-CC	2D	[2-^{13}C]-glycerol	^{13}C-^{13}CA	498 (461)
1,3-NCOCX	3D	[2-^{13}C]-glycerol	^{13}C-^{13}CO	481 (477)
2-NCACX	3D	[2-^{13}C]-glycerol	^{13}C-^{13}CA	391 (377)

Table 2.3 Protein characteristics and spectra used in the calculations performed with SOLARIA. The PDB entry refers to the X-ray reference structure. Peak lists were generated by manual peak-picking. The numbers in parentheses refer to the total number of peak list entries after the manual removal of inter-molecular cross-peaks. No manual assignment was included in the lists.

2.2.2 Structure calculations

The software SOLARIA was used to perform the automatic assignment of the spectra peak lists and to generate structures. The calculations were conducted on a SGI Origin 2100 cluster at the FMP, Berlin. The chemical shift tolerances were set to 0.4 ppm for carbon- and 0.5 ppm for nitrogen dimensions. The number of calculated structures was 20 for the first eight iterations and 100 for the final one. When using 100,000 cooling steps during the simulated annealing protocol, the calculation carried out by four processors in parallel took approximately 12 hours. The results of the calculations were evaluated by computing a pair-wise rmsd and an rmsd to the X-ray reference structure[56] of the 15 lowest-energy structures.