

Sensitivity Enhancement  
of  
Liquid-State NMR  
and  
Improvement of the INPHARMA  
Method

Dissertation  
zur Erlangung des mathematisch-naturwissenschaftlichen  
Doktorgrades  
"Doctor rerum naturalium"  
der Georg-August-Universität Göttingen

vorgelegt von

Marcel Reese

aus Bielefeld

Göttingen, 2010

D 7

Referent: Prof. Christoph Schmidt

Koreferent: Prof. Christian Griesinger

Tag der mündlichen Prüfung: 2010-04-08

# Preface

This dissertation is a result of my doctorate studies at the University of Göttingen in supervision by Prof. Christoph Schmidt and at the Max Planck Institute for Biophysical Chemistry (MPIBPC) in supervision by Prof. Christian Griesinger.

I had the chance to work on two main topics of great interest:

The first topic relates to the INPHARMA method (interligand NOEs for pharmacophore mapping). This method was developed with my participation during my diploma thesis at the MPIBPC. The method is applicable in the context of structure-based drug design and is based on liquid-state NMR spectroscopy. It uses the two-dimensional NOESY NMR experiment under conditions of chemical exchange of ligands with their receptor.

The INPHARMA experiment allows for the determination of complex models of small, drug-like ligands binding to large macromolecular receptors. Until now, a given hypothetical, structural model could only be used as basis to simulate the expected outcome of an INPHARMA experiment. Then the comparison of the simulated and the experimental data was used to select models which were in best agreement with the experiment. A residual difference of the simulated and measured data was observed which made a narrow structural selection, based exclusively on INPHARMA, impossible.

Here, I show how experimental INPHARMA data can be used for directed structural refinement of models towards better agreement of experimental and calculated data. This is achieved by defining a INPHARMA-based penalty energy and by calculating its analytical gradient with respect to the coordinates of the involved atoms. Both, energy and its gradient have been implemented into the software XPLOR-NIH. Structural refinement using protocols with steepest decent minimization and simulated annealing molecular dynamics are presented. The coordinates of all receptor and ligand atoms influencing the INPHARMA experiment get optimized. The derived theoretical frame work can be directly applied to other NMR spectra, like transferred NOE (trNOE) or inter-ligand NOE (ILOE) spectra.

The second topic of my dissertation relates to sensitivity enhancement of liquid-state NMR spectroscopy in structural biology. Compared to other methods of molecular biology, NMR is very insensitive and requires about one thousand times more sample than many other methods. This is due to the small gyromagnetic ratio of nuclei and the resulting small net magnetic moment of NMR samples in high magnetic fields. Recently, techniques for sensitivity enhancement of NMR got increasing attention. The umbrella term for these techniques is dynamic nuclear polarization (DNP). Here, the intrinsically higher polarization of paramagnetic centers in a magnetic field is transferred to nuclei. Although signal enhancement factors of several thousand have been reported, no technique exists to allow for repeated, reproducible NMR experiments on one and the same macromolecular sample in liquid, aqueous solution. Repeated and reproducible NMR experiments are the prerequisite for many techniques of high-dimensional NMR spectroscopy, allowing for the determination of the molecular structure.

Here, I present the results of a design study for a liquid-state NMR spectrometer using DNP for the direct polarization of liquid, aqueous samples at ambient conditions. The samples are polarized at a low magnetic field and then are automatically transferred to a high magnetic field for high-resolution NMR experiments. Significant signal enhancement could be observed for small molecules dissolved in aqueous solution. Repeated and reproducible measurements on one and the same sample are possible. The applicability of the method to macromolecules is currently under investigation.

Marcel Reese  
Göttingen, 2010-03-17

minor editorial changes:  
Cambridge, USA, 2011-02-15

# Contents

<b>Preface</b>	<b>iii</b>
<b>Table of Contents</b>	<b>v</b>
<b>I INPHARMA</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Overview . . . . .	3
1.2 INPHARMA . . . . .	8
<b>2 Theory</b>	<b>15</b>
2.1 Simulation of INPHARMA Peak Volumes . . . . .	15
2.2 Derivation: INPHARMA Energy and Gradient . . . . .	18
2.2.1 Alternative Penalty Functions . . . . .	25
<b>3 XPLOr-NIH Implementation</b>	<b>29</b>
3.1 INPHARMA Gradient in a Nut Shell . . . . .	29
3.2 Numerical Approximations . . . . .	30
3.3 Minimum Example . . . . .	34
<b>4 Results</b>	<b>37</b>
4.1 6-Spin System . . . . .	37
4.1.1 One-dimensional Energy Landscape . . . . .	38
4.1.2 Numerical and Analytical Derivatives . . . . .	47
4.1.3 Minimization . . . . .	52

4.2	12-Spin System . . . . .	59
4.2.1	Minimization . . . . .	59
4.3	Tubulin - Epothilone A - Baccatin III . . . . .	63
4.3.1	Minimization . . . . .	64
4.3.2	Simulated Annealing . . . . .	86
<b>5</b>	<b>Conclusion and Outlook</b>	<b>107</b>
<b>II</b>	<b>DNP</b>	<b>111</b>
<b>6</b>	<b>Introduction</b>	<b>113</b>
6.1	Brief Theory . . . . .	116
<b>7</b>	<b>Summary</b>	<b>121</b>
7.1	Low-field DNP I, Peter Höfer et al., 2008 . . . . .	122
7.2	Low-field DNP II, Peter Höfer et al., 2008 . . . . .	124
7.3	Shuttle DNP I, Marcel Reese et al., 2008 . . . . .	126
7.4	Shuttle DNP II, Marcel Reese et al., 2009 . . . . .	128
7.5	Shuttle DNP III, Alexander Krahn et al., 2010 . . . . .	130
7.6	Low-field DNP III, Maria-Teresa Türke et al., 2010 . . . . .	132
<b>8</b>	<b>Conclusion and Outlook</b>	<b>137</b>
	<b>Appendix I, INPHARMA</b>	<b>141</b>
<b>9</b>	<b>Theory</b>	<b>143</b>
9.1	Simulation of INPHARMA Peak Volumes . . . . .	143
9.1.1	Setup of the Kinetic Matrix . . . . .	143
9.2	Derivation: INPHARMA Energy and Gradient . . . . .	144
9.2.1	Lemma 1 . . . . .	144
9.2.2	Lemma 2 . . . . .	146
9.3	Alternative Penalty Functions . . . . .	146

<b>10 Results</b>	<b>149</b>
10.1 6-Spin System . . . . .	149
10.1.1 One-dimensional Energy Landscape . . . . .	149
10.1.2 Minimization . . . . .	158
10.1.3 Numerical and Analytical Derivatives . . . . .	161
10.2 12-Spin System . . . . .	166
10.2.1 Minimization . . . . .	166
10.3 Tubulin - Epothilone A - Baccatin III . . . . .	171
10.3.1 Minimization . . . . .	171
10.3.2 Simulated Annealing . . . . .	181
<b>Appendix II, DNP</b>	<b>195</b>
<b>11 Contributions</b>	<b>197</b>
<b>12 Suppl. Figures</b>	<b>203</b>
<b>Bibliography</b>	<b>213</b>
<b>Acknowledgments</b>	<b>221</b>
<b>Curriculum Vitae</b>	<b>223</b>



Part I

INPHARMA



# Chapter 1

## Introduction

### 1.1 Overview

NMR is a powerful and established tool in structural biology. Recently it is gaining more and more importance in the field of structure based drug design.<sup>1-3</sup> Here, NMR can be helpful for primary screening, binding site localization, hit validation, structure determination and compound optimization.<sup>2</sup>

In structure based drug design molecular systems of ligand-receptor complexes are subjects of interest. The ligands will be small-molecular fragments, lead molecules, drug-like molecules, drug candidates or drugs having a weight of no more than a few hundred Dalton. The receptor or target, normally a protein or nucleic acid, will be of some physiological importance in the context of a disease and will be affected in function by binding of the ligand. Often the inhibition of the receptor's function is desired. For the ligand, being a drug, it is of utmost importance that it experiences a very specific and strong interaction with only the desired target, and does not accidentally interfere with other physiological processes. If the details of the atomic interaction between a specific physiological target molecule and a small drug-like molecule are understood this information can be used for improving the binding affinity of the ligand and its specificity to a specific target.

Several techniques have been developed to probe for the structural details of ligand receptor complexes. Many of these base on the physical effect of dipolar cross relaxation which is famously applied in the NOESY (nuclear Overhauser effect spectroscopy) experiment.<sup>4</sup> This two dimensional NMR experiment allows to determine the rotational correlation time of the molecule and the *distance dependent* cross-relaxation rates between protons.

These relaxation rates can then be used to measure the distances between the protons. These internuclear distances can then be used to determine the molecular structure by means of distance restrained optimization methods.

In a NOESY experiment of single molecules both, the magnetization transfer due to dipolar cross relaxation (nuclear Overhauser effect, NOE) and chemical exchange (moving spins) follow the same differential equation. The respective magnetization transfer rates add up linearly.

For macromolecules the NOE cross relaxation rates are virtually proportional to the rotational correlation time of the molecules, and thus to their molecular weight. The mass of macromolecular receptors will generally be many orders of magnitude higher than the mass of drug-like ligands. Thus, the NOE cross-relaxation rates inside of the receptor will be orders of magnitude higher than the ones in the free ligand.

The molecular size dependence and the distance dependence of NOE rates are used in the transferred nuclear Overhauser (trNOE) experiment. Here, in solution, a weakly binding ligand is added to a small amount of its receptor. In this situation the small ligand will quickly and repeatedly enter and leave the binding pocket of the large receptor. As mentioned, the NOE relaxation rates of molecules in solution depend linearly on the molecular weight. This is why the short, repeated binding events of the ligand and its resulting small, bound fraction (located in a bigger complex) will lead to an increase of the observed, average intra-ligand relaxation rates. The relaxation rate difference between complexed and free ligand can be so large that the observed, average relaxation rate of the ligand is completely dominated by the complexed population of the ligand.

These increased rates can be measured quantitatively by an NOESY experiment performed on such a mixture of receptor and ligand. If necessary, e.g. in the case of a low-molecular-weight receptor or a low ligand concentration (when the bound conformation is not completely dominating the relaxation rates), a reference spectrum without receptor can be used to correct for the influence of the free population of the ligand. In conclusion, transferred NOE (trNOE) measurements allows for the determination of the structure of the bound ligand, without interference of the free ligand's population.

Transferred NOE experiments are quite simple with respect to sample preparation as they do only require tiny amounts of *unlabeled* receptor; and isotope labeling of proteins or nucleotides is generally an tedious and expensive task and might even not be possible (as in the tubulin case explained later). As rule of thumb, the ratio of concentration of ligand and receptor is inversely proportional to the ratio of their molecular weights. The bigger the receptor, the less is required. The concentration of the ligand should be in the 10 micromolar to a low millimolar range.

Due to the low concentration of the receptor and due to its size its resonances are weak and broad. Normally this leads to the situation where the resonances become vanishingly small in the spectra and might only appear as a kind of baseline distortion. The detectable resonances only come from the ligand.

Several NMR experiments exist, using different kinds of samples, relying on chemical exchange of species during a NOESY experiment. For some of them the only theoretical differences are the details of the setup the NOE relaxation matrix, the kinetic matrix and the initial magnetization matrix (2.2, p. 15). Experiments like the transferred NOE (trNOE)<sup>5-7</sup> or inter ligand NOE (ILOE)<sup>8,9</sup> and INPHARMA<sup>10-14</sup> experiments follow the same differential equation. For these experiments the setup of the required matrices has been described.<sup>7,9,11,15,16</sup>

Other experiments, like the 1D-STD or 2D-STD (saturation transfer difference)<sup>17</sup> and WaterLOGSY (water-ligand observed via gradient spectroscopy)<sup>18</sup> experiments follow very similar differential equations and require very similar matrices. Here one species in solution is irradiated at or close to the Larmor-frequency of its spins. By this the spins of this species get magnetically saturated, the magnetization is reduced (mostly to zero). Spins of other species in solution, binding or interacting with the saturated species, are influenced by cross relaxation with the saturated spin either directly or via spin diffusion. This effect is detected and analyzed.

In contrast to the trNOE experiment, in INPHARMA experiments *two* ligands will be added to the receptor, competitively binding to the same binding pocket. The observed NOESY spectra of the ligands can be used to determine the orientation of binding of the ligands in respect to each other (relative orientation), or in respect to the receptor (absolute orientation),<sup>13</sup> and can be used to probe for the atomic structure of the binding pocket.<sup>11,12</sup> A detailed introduction into INPHARMA will be provided later (ch. 1.2). Inter ligand NOEs (ILOE) rely on a similar principle as INPHARMA. ILOE originate from two ligands binding simultaneously to different parts of the same binding pocket of a common receptor.<sup>8,9</sup>

The observed peak volumes of the trNOE experiment do not only depend on the distances of protons of the ligand. In the bound state many ligand protons will have some neighboring receptor protons affecting the evolution of magnetization transfer during a NOESY experiment.<sup>19,20</sup> The magnetization of the ligands starts to diffuse (spin diffusion) into the receptor and protons of the receptor can provide alternative, relayed pathways for magnetization transfer. These effects play a major role in NOESY experiments with longer mixing times (the time in a NOESY experiment where the spins are allowed to exchange magnetization with each other).

In short-mixing-time experiments the magnetization transfer will be dominated by the nearest neighbor of a proton. The NOE intensity between a pair of protons is directly proportional to the inverse of the sixth power of the distance between two protons.<sup>15</sup> This makes it easy to derive internuclear distances, and by this structures, from the measured cross relaxation rates using an isolated spin-pair model.<sup>21</sup> Just, small mixing time NOE cross-peaks are generally very weak and longer mixing times have to be used to achieve an satisfactory signal to noise ratio and to observe a decent number of cross peaks. Thus, in most situations, the application of the isolated spin-pair model is not satisfactory and a full relaxation matrix approach, considering all present protons and all relaxation pathways, is recommended for structure determination.<sup>22</sup>

The full relaxation matrix approach is well established for NMR based structure determination of single molecules in solution,<sup>22</sup> not experiencing chemical exchange. A corresponding optimization algorithm was implemented in DIANA<sup>23</sup> and X-PLOR.<sup>22,24</sup> X-PLOR is a program for NMR and crystallography based structure determination. First, the theoretical peak volumes are calculated from a given molecular conformation and compared to the experiment. Then a penalty energy is calculated being lower for models better fulfilling the experimental restraints. Furthermore the gradient of this penalty energy in respect to the coordinates of the involved protons is calculated. This gradient can be used to guide directed minimization procedures, like the method of steepest descent, to optimize a given conformation by convergence to a local minimum of the total energy of the system. From here any displacement of any atom would result in a higher allover energy of the system, also considering other, empirical molecular force field or experimental energy terms. XPLOR allows for further optimization by using simulated annealing protocols where the molecular dynamics of the system, generally in contact with a heat bath, will be simulated. Here, the calculated gradient of the penalty energy will be used to provide a force to guide the system into a new conformation better fulfilling the experimental NOE restraints. Simulated heating and slow cooling cycles (simulated annealing) allow to overcome local minima and, in principle, allow to find the global minimum of the total energy<sup>25,26</sup> and so to find the conformation in best agreement with the experimental data; while also guaranteeing for chemically meaningful conformations. It has been shown that protocols using these NOE gradients can lead to significant improvements of the accuracy of NMR structure determination.<sup>22</sup>

The calculation of the gradient of the penalty energy is the key to using NOESY data for accurate and efficient directed minimization or during simulated annealing molecular dynamics simulations. Regarding theory, this field has been pioneered by Yip and Case,<sup>27-29</sup> with a later contribution by

Nesterova and Chuprina.<sup>30</sup> Both concepts require the diagonalization of the relaxation matrix, an expensive  $\mathcal{O}(N^3)$  operation (with a large prefactor). The XPLOR implementation by Nilges et al. is based on the theoretical work of Yip and Case.<sup>28</sup>

Obviously, it is of great interest to find structural models perfectly explaining experimental NMR spectra based on the Overhauser effect and involving chemical exchange of ligands. For STD-NMR spectra iterative torsion angle refinement of the ligand has been reported and applied to structural refinement of complexes, only modifying the ligand, under chemical exchange conditions.<sup>31</sup> The complex structures, including the receptor atoms, have been considered for simulating the spectra. Not structural refinement, but parameter fitting has been performed for trNOE spectra using the Powell method for minimization,<sup>20</sup> without using analytical gradients. It is common to derive distance restraints from experimental trNOE spectra, neglecting the protons of the receptor and using a simple isolated spin model.<sup>32-34</sup> Then, the distance restraints are used for conventional NMR structure determination. Furthermore trNOE data was used for an analytical gradient based full relaxation matrix refinement using XPLOR.<sup>35</sup> However, here the refinement was neither considering the receptor protons nor chemical exchange. Instead, by using an effective rotational correlation time, the bound ligand was simulated to mimic a 'bigger free ligand' of a size between complex and free ligand to emulate the effect of binding.

Generally, it is important to consider the protons of the receptor in the simulation of trNOE data, also for the determination of distance restraints from experimental trNOE data.<sup>19,20</sup>

NOESY spectra of single species, without chemical exchange, can be nicely simulated and molecular models can be structurally refined using analytical gradient based methods. The methods include simulated annealing molecular dynamics and analytical gradient based minimization routines.

Until now, to my knowledge, no application or method has been reported for gradient based optimization and structural refinement using NMR spectra with chemical exchange. This is including both, purely numerical gradients and analytical gradients. Neither conjugate gradient nor steepest descent examples, neither from application nor from theory are reported.

INPHARMA, as already mentioned, it is based on NOESY spectra measured under conditions of chemical exchange between two ligands and a receptor. So far this method was only used to back-calculate INPHARMA spectra from structural models. Then the models were scored using experimental data as reference.<sup>11,14,35</sup> No directed refinement was employed.

## 1.2 INPHARMA

Transfer interligand NOEs between two ligands binding competitively to the same binding pocket of a common receptor have first been observed in our group.<sup>10-1</sup> These spectra, called INPHARMA spectra, can be used to determine the binding modes of two ligand-receptor complexes from a single measurement.<sup>11,13</sup> They are based on the trNOE effect (transfer nuclear Overhauser effect). In INPHARMA spectra, in contrast to trNOE spectra, not only one but *two* ligands will be in solution together with a common receptor. Normally, the receptor is big enough not to be seen in the NMR spectra, due to line broadening. The measurement of a normal NOESY spectrum on this sample will produce cross peaks between atoms of the two different ligands, indicating magnetization transfer. This transfer between the ligands is purely mediated by the receptor. The ligands don't bind to the same binding pocket at the same time, nor do they transfer magnetization when they are in the bulk solvent. Possibly occurring minor contributions to interligand magnetization transfer, like from unspecific binding of the ligands to the receptor, are neglected for now.

The application of INPHARMA spectra for structure determination, the determination of the binding epitopes, has first been done during my work on my diploma thesis.<sup>11,12</sup> There a model of epothilone A, an anti cancer agent, binding to tubulin, was determined. This model explained virtually all available SAR (structure activity relationship) data and mutagenesis data. In 2008 the method was successfully validated on a molecular reference system of complexes with structures known from X-ray crystallography.<sup>13</sup>

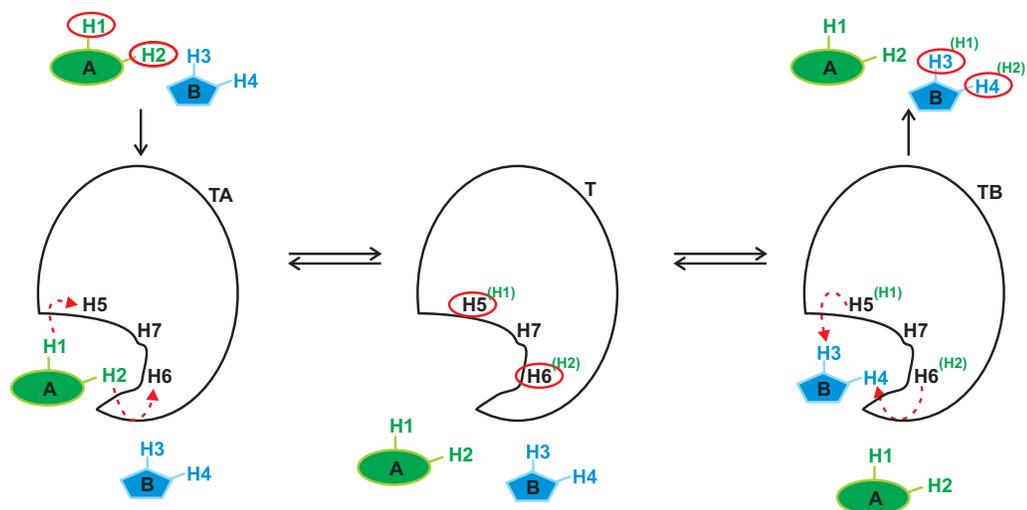
During the INPHARMA experiments the concentrations of both ligands, the receptor and the complexes are in chemical equilibrium. All species exist in a free and bound form. The principle of the method shall be illustrated in figure (fig. 1.1).

If the concentration of the ligands is large enough the free, uncomplexed state of the receptor can be neglected.<sup>14</sup> Experimentally, for the tubulin-epothilone A-baccatin III system, the concentration ratio receptor:ligandA:ligandB was approximately 1:50:50.<sup>10</sup> Thus the following (1.2) simplified model was used for the simulation the the INPHARMA spectra. It will be used during the remaining part of this document.

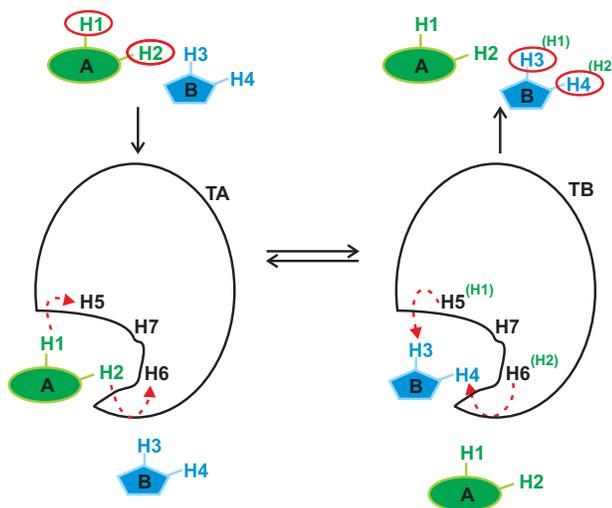
For two given complex structures TA and TB of the 2-step model the resulting INPHARMA spectrum can be simulated. As the free ligands, due to their small size, don't contribute significantly to the spectra their conformation is mostly irrelevant, though considered. In the ideal case only

---

<sup>1</sup>Similar spectra involving two ligands binding close to each other in the same binding pocket at the same time had first been reported by Dawei Li et al.<sup>8</sup>

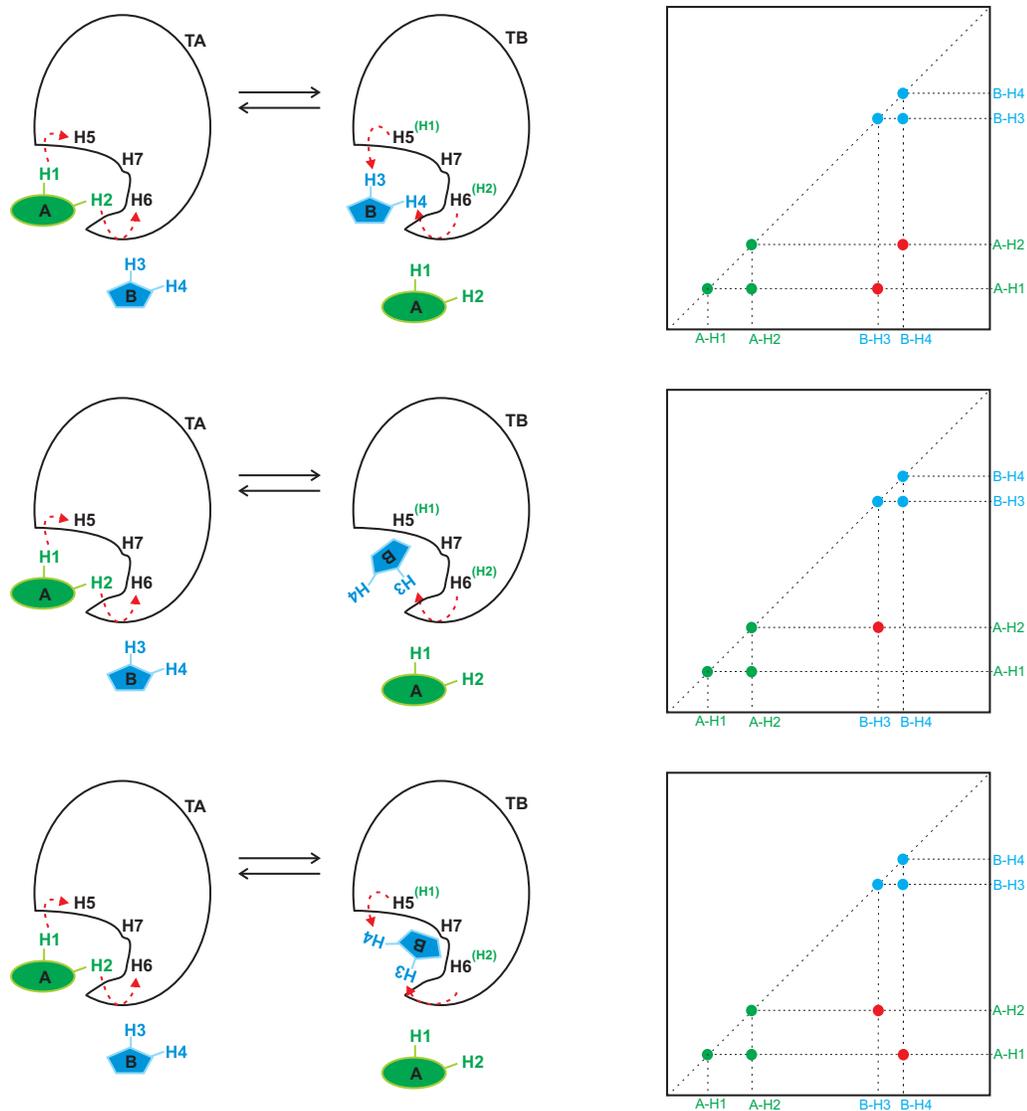


**Figure 1.1:** Sketch of the 3-step model of INPHARMA magnetization transfer. The  $z$ -magnetization (red) of the protons H1 and H2 of ligand A can be partially transferred to ligand B. Adapted from a figure by Dr. Víctor Sánchez-Pedregal.



**Figure 1.2:** Sketch of the 2-step model of INPHARMA magnetization transfer. This model can be used if the ligands are in large excess in respect to the receptor.<sup>14</sup> Adapted from a figure by Dr. Víctor Sánchez-Pedregal.

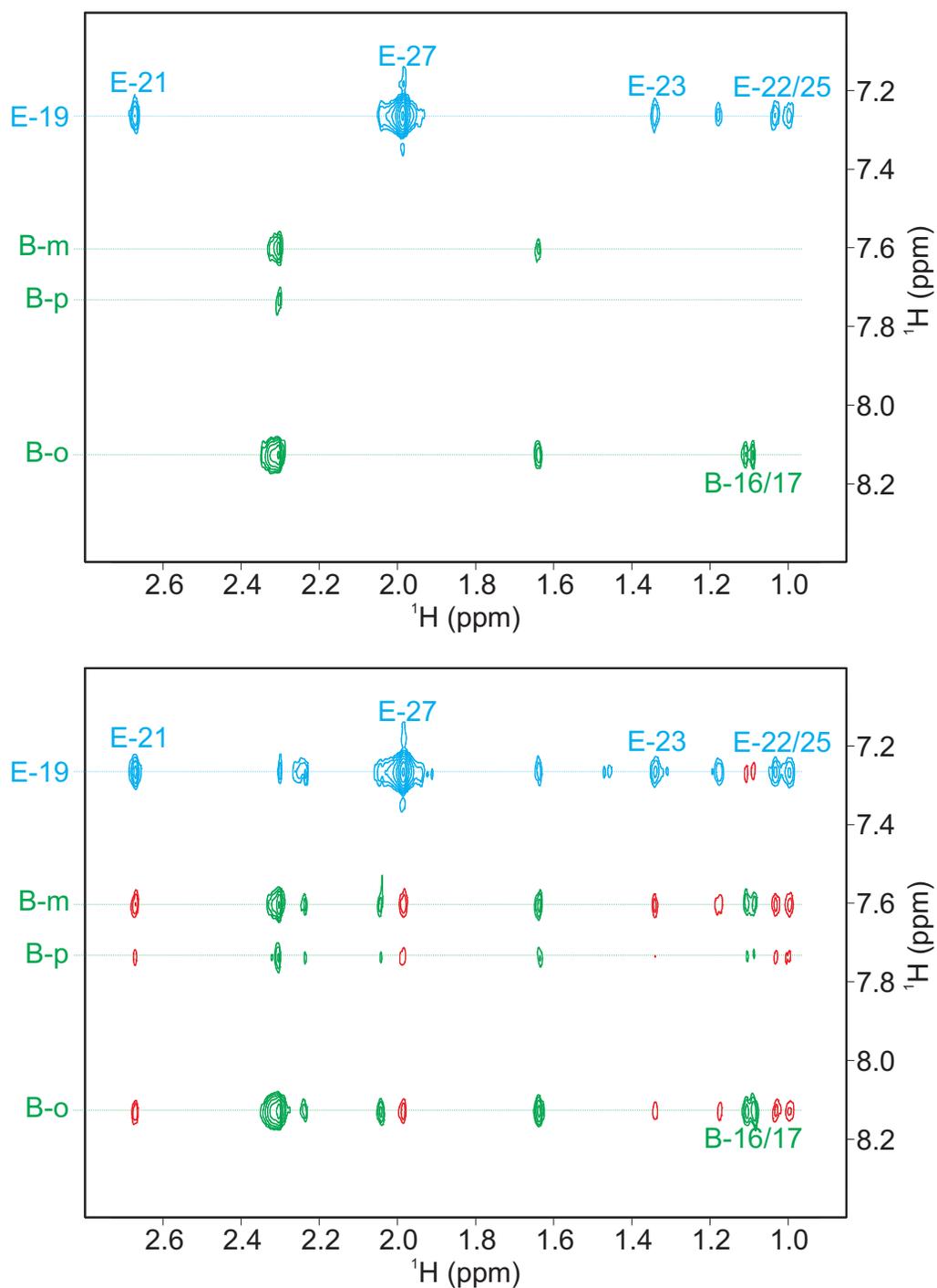
a single pair of complexes can be found to explain the experimental data. This would mean that the structure of both complexes would have been determined. The INPHARMA spectra will depend on the orientations of the ligands in the binding pocket, their conformation, and their precise relative position to neighboring atoms of the receptors, and so on the structure of the binding pocket. This dependence is illustrated in fig. 1.3.



**Figure 1.3:** Model of three different complex pairs and the resulting INPHARMA spectra. Purely intra-molecular peaks are highlighted in the color of the respective ligand. Inter-molecular INPHARMA peaks are highlighted in red. It is clearly visible how the different orientations of ligand B change the pattern of the INPHARMA peaks. At the same time the intra-molecular trNOE peaks stay hardly unchanged. In this figure only the orientation of ligand B is modified. The orientation of the other ligand shall be assumed to be known. If also ligand A is allowed to be moved, having an unknown orientation in respect to the receptor, there might be some different orientations leading to the same INPHARMA spectra. This case would occur if ligand A, in the lowest example, is turned 180°. The resulting spectrum would just look like the one on the top. Spin diffusion is neglected in this figure. If the mixing time, allowing for magnetization transfer, is chosen long enough all present protons will show cross peaks to each other. These peaks will be, in general, of different intensity, in contrast to the figure. The intensity of the peaks provides valuable gradual information on the magnetization exchange and can be used for structure determination. Adapted from a figure by Dr. Víctor Sánchez-Pedregal.

For the measurement of the INPHARMA spectrum a simple NOESY pulse sequence<sup>4</sup> is sufficient. In fig. 1.4 slices of such spectra are shown<sup>10</sup>

In the following chapter 'Theory' I will explain the theoretical basis of the INPHARMA method in greater detail. Then I will explain the theoretical and mathematical foundations for the analytical gradient based structural refinement of INPHARMA spectra. This theoretical approach can be directly applied to other systems, being on a general level mathematically equivalent to INPHARMA, like transferred NOE, interligand Overhauser effect spectra, or NOESY spectra of single molecules undergoing conformational exchange. In the chapter 'XPLOR-NIH Implementation' I will shortly explain the implementation details of the INPHARMA restraints into XPLOR-NIH.<sup>36,37</sup> Then, in the chapter 'Results', the application of INPHARMA restraints in the context of various test systems will be exemplified.



**Figure 1.4:** Slice of an experimental INPHARMA spectrum of the tubulin-epothilone A- baccatin III system. The upper measurement, using 20 ms mixing time, only shows intraligand trNOE peaks in green and blue. The lower measurements, using 70 ms mixing time, also shows INPHARMA peaks in red. Measurement by Dr. Víctor Sánchez-Pedregal, adapted from a figure by Dr. Víctor Sánchez-Pedregal.



# Chapter 2

## Theory

### 2.1 Simulation of INPHARMA Peak Volumes

The peakvolumes  $\mathbf{A}$  of INPHARMA spectra evolve according to the following differential equation as described previously:<sup>11,12</sup>

$$\frac{d\mathbf{A}}{dt} = -\mathbf{R}^* \cdot \mathbf{A} = -(\mathbf{R}^{\text{NOE}} + \mathbf{K})\mathbf{A}. \quad (2.1)$$

The NOE relaxation matrix  $\mathbf{R}^{\text{NOE}}$  and the chemical exchange matrix  $\mathbf{K}$  both contribute in physically different but mathematically equivalent ways to the magnetization exchange. Their sum will be called the effective relaxation matrix  $\mathbf{R}^*$ .

The NOE relaxation matrix  $\mathbf{R}^{\text{NOE}}$  can be seen as a matrix of submatrices  $\mathbf{R}^s$  and is of the general form

$$\mathbf{R}^{\text{NOE}} = \begin{pmatrix} \mathbf{R}^{\text{A}} & 0 & 0 & 0 \\ 0 & \mathbf{R}^{\text{B}} & 0 & 0 \\ 0 & 0 & \mathbf{R}^{\text{TA}} & 0 \\ 0 & 0 & 0 & \mathbf{R}^{\text{TB}} \end{pmatrix}. \quad (2.2)$$

The chemical species are denoted with the index  $s$ , here A,B,TA or TB. The non-diagonal elements of  $\mathbf{R}^{*s}$  are

$$\mathbf{R}_{ij}^s = \sigma_{ij}^s = \frac{1}{10} \gamma_i^2 \gamma_j^2 \left( \frac{\hbar \mu_0}{4\pi} \right)^2 \frac{1}{(r_{ij}^s)^6} [6J_{ij,2}^s - J_{ij,0}^s]. \quad (2.3)$$

Dipolar cross relaxation of a spin is induced by neighboring moving spins (due to Brownian motion) which create randomly fluctuating magnetic fields. These random fields induce random transitions causing relaxation.

For the simple model of a rigid molecule undergoing isotropic rotational diffusion the spectral densities  $J_{ij,n}^s$  are:

$$J_{ij,n}^s = \frac{\tau_{ij}^s}{1 + (n\omega_0\tau_{ij}^s)^2}. \quad (2.4)$$

The cross relaxation rate  $\sigma_{ij}^s$  between two individual protons  $i$  and  $j$  belonging to species  $s$  depends on their distance  $r_{ij}^s$ , on the isotropic rotational correlation time  $\tau_{ij}^s$  of the length of the internuclear vector, their Larmor-frequency  $\omega_o$  and so on the strength of the external magnetic field  $B_0$ . For the model of a rigid molecule, used here, the internuclear vector's rotational correlation time  $\tau_{ij}^s$  and the isotropic rotational correlation time of the molecular species  $\tau_c^s$  are identical. The gyromagnetic ratio of the proton  $\gamma_H$ , the reduced Planck-constant  $\hbar$  and the magnetic constant  $\mu_0$  are well known natural constants.

The diagonal elements of the matrix  $\mathbf{R}^s$  are

$$\mathbf{R}_{ii}^s = \sum_{j \neq i} \rho_{ij}^s \quad (2.5)$$

with

$$\rho_{ij}^s = \frac{1}{10} \gamma_i^2 \gamma_j^2 \left( \frac{\hbar \mu_0}{4\pi} \right)^2 \frac{1}{(r_{ij}^s)^6} [J_{ij,0}^s + 3J_{ij,1}^s + 6J_{ij,2}^s]. \quad (2.6)$$

The following kinetic model, visualized in (fig. 1.2), is used:



The chemical equilibrium state of the system is determined by the initial concentrations  $[\text{T}]_0$ ,  $[\text{A}]_0$ , and  $[\text{B}]_0$  and the kinetical rate constants  $k_{12}$  and  $k_{21}$ . These are input parameters of the simulation. All the equilibrium concentrations of all species  $[\text{A}]$ ,  $[\text{B}]$ ,  $[\text{TA}]$  and  $[\text{TB}]$  are calculated from them and the kinetic model (2.7). The complete derivation of the kinetic matrix  $\mathbf{K}$  was exemplified in [9] and has been published in detail in [11–13]. The calculation is repeated in the appendix of this thesis (ch. 9.1.1). This kinetic

matrix is used in the following:

$$\begin{aligned}
 \mathbf{K} &= \begin{pmatrix}
 \boxed{k_{21}[\text{TB}]\mathbf{I}} & 0 & -k_{12}[\text{B}]\mathbf{I} & 0 & 0 & 0 \\
 0 & \boxed{k_{12}[\text{TA}]\mathbf{I}} & 0 & 0 & -k_{21}[\text{A}]\mathbf{I} & 0 \\
 -k_{21}[\text{TB}]\mathbf{I} & 0 & \boxed{k_{12}[\text{B}]\mathbf{I}} & 0 & 0 & 0 \\
 0 & 0 & 0 & \boxed{k_{12}[\text{B}]\mathbf{I}} & 0 & -k_{21}[\text{A}]\mathbf{I} \\
 0 & -k_{12}[\text{TA}]\mathbf{I} & 0 & 0 & \boxed{k_{21}[\text{A}]\mathbf{I}} & 0 \\
 0 & 0 & 0 & -k_{12}[\text{B}]\mathbf{I} & 0 & \boxed{k_{21}[\text{A}]\mathbf{I}}
 \end{pmatrix} \\
 &\equiv \begin{pmatrix}
 \mathbf{K}^{\text{A}} & \mathbf{K}^{\text{B}} & \mathbf{K}_{\text{A}}^{\text{TA}} & \mathbf{K}_{\text{T}}^{\text{TA}} & \mathbf{K}_{\text{B}}^{\text{TB}} & \mathbf{K}_{\text{T}}^{\text{TB}}
 \end{pmatrix}
 \end{aligned} \tag{2.8}$$

Here, using the identity matrix  $\mathbf{I}$ , the initial peak volume matrix is given:

$$\mathbf{A}^0 = \begin{pmatrix}
 [\text{A}]\mathbf{I} & 0 & 0 & 0 \\
 0 & [\text{B}]\mathbf{I} & 0 & 0 \\
 0 & 0 & [\text{TA}]\mathbf{I} & 0 \\
 0 & 0 & 0 & [\text{TB}]\mathbf{I}
 \end{pmatrix}. \tag{2.9}$$

The efficient and fast calculation of eigenvalues and eigenvectors of a quadratic matrix requires symmetric matrices as input. Though the NOE contribution to the effective relaxation matrix is symmetric the kinetic matrix contribution is not. The diagonal symmetrization matrix  $\mathbf{C}$  with  $C_{ii} = 1/\sqrt{K_{ii}}$  is used for symmetrization.

Not all atoms of the molecule will have the same chemical shift, like protons in methyl groups. All cross peak volumes of them will be actually the observation of the sum of many individual-proton cross peaks. This is considered in the grouping matrix  $\mathbf{G}$ , having one column for every observable resonance, a group. The entries of the column are "1" for atoms belonging to the group and "0" for all other atoms.

The solution for the time evolution of the grouped peak volumes<sup>11,12</sup> is:

$$\begin{aligned}
\mathbf{A}_m^g(\tau_m) &= \mathbf{G}\mathbf{A}_m(\tau_m)\mathbf{G}^{\text{tr}} \\
&= \mathbf{G} \exp(-\mathbf{R}^* \tau_m) \mathbf{A}^0 \mathbf{G}^{\text{tr}} \\
&= \mathbf{G}\mathbf{C} \exp(-\mathbf{R}^{*s} \tau_m) \mathbf{C}^{-1} \mathbf{A}^0 \mathbf{G}^{\text{tr}} \\
&= \mathbf{G}\mathbf{C}\boldsymbol{\chi}^{*s} \exp(-\boldsymbol{\lambda}^{*s} \tau_m) (\boldsymbol{\chi}^{*s})^{-1} \mathbf{C}^{-1} \mathbf{A}^0 \mathbf{G}^{\text{tr}} \\
&= \mathbf{G}\mathbf{C}\boldsymbol{\chi}^{*s} \exp(-\boldsymbol{\lambda}^{*s} \tau_m) (\boldsymbol{\chi}^{*s})^{\text{tr}} \mathbf{C}^{-1} \mathbf{A}^0 \mathbf{G}^{\text{tr}}.
\end{aligned} \tag{2.10}$$

The following symbols are used:

- $\mathbf{A}_m^g$  Grouped peak volume matrix for the mixing time with index  $m$
- $\mathbf{A}_m$  Individual-atom peak volume matrix for the mixing time with index  $m$
- $\mathbf{A}^0$  Individual-atom initial peak volume matrix
- $\mathbf{G}$  Grouping matrix
- $\boldsymbol{\lambda}$  Diagonal matrix of the eigenvalues of the effective, symmetrized relaxation matrix  $\mathbf{R}^{*s}$
- $\boldsymbol{\chi}^{*s}$  Orthonormal matrix of eigenvectors of the effective, symmetrized relaxation matrix  $\mathbf{R}^{*s}$  with  $\boldsymbol{\chi}^{*str} = \boldsymbol{\chi}^{*s-1}$
- $\mathbf{C}$  Diagonal symmetrization matrix
- $\tau_m$  Mixing time of the NOESY experiment with index  $m$

For clarity I define  $\mathbf{L} \equiv \boldsymbol{\chi}^{*s}$  and  $\boldsymbol{\Lambda} \equiv \boldsymbol{\lambda}^{*s}$ . Now the equation becomes

$$\begin{aligned}
\mathbf{A}_m^g &= \mathbf{G}\mathbf{A}_m\mathbf{G}^{\text{tr}} \\
&= \mathbf{G}\mathbf{C}\mathbf{L} \exp(-\boldsymbol{\Lambda}\tau_m) \mathbf{L}^{\text{tr}} \mathbf{C}^{-1} \mathbf{A}^0 \mathbf{G}^{\text{tr}}
\end{aligned} \tag{2.11}$$

## 2.2 Derivation: INPHARMA Energy and Gradient

For using directed-minimization-based or molecular-dynamics-based optimization methods for structure determination it is in general necessary, or at least beneficial, to calculate the derivative of a penalty energy with respect to the structural parameters of a molecule. This derivative or gradient is either used to find the "best way down" to the minimum of the energy landscape or to provide a force  $\mathbf{F} = -\nabla_{x,y,z} E$  during molecular dynamics simulations, biasing the atom movements towards a lower energy conformation.

The INPHARMA penalty energy  $E_{\text{INPH}}$  is a function of the experimental peak volumes  $\mathbf{A}^{\text{exp}}$ , their experimental errors  $\mathbf{A}^{\text{err}}$  and the calculated (grouped) peak volumes  $\mathbf{A}^g$ .

The INPHARMA energy (the penalty energy) is defined to be an R-factor (2.15) scaled by an allover energy constant  $k_{\text{INPH}}$

$$E_{\text{INPH}} = k_{\text{INPH}} R. \quad (2.12)$$

Now the energy with respect to all coordinates of all involved atoms has to be calculated. The INPHARMA peak volumes depend on the precise conformations of the free ligands as well as on both precise conformations of the complexes. As a results, this also applies to the penalty energy and to the gradient. Each atom has direct influence on all forces on all other atoms.

The following derivation of the analytical matrix derivative of the INPHARMA penalty energy is based on the example of Nesterova and Chuprina, 1993.<sup>30</sup> There, only NOESY spectra of a single molecule were considered using the following solution for the calculation of the individual-atom peak volumes

$$\mathbf{A}_m^{\text{calc}} = k \cdot \exp(-\mathbf{R}^{\text{NOE}} \tau_m). \quad (2.13)$$

This solution is much simpler than the respective solution for the INPHARMA peak volumes (2.10, third line). It assumes the same initial magnetization on all individual spins, does not account for chemical exchange nor chemical shift overlap. In the INPHARMA case the initial peak volume matrix  $\mathbf{A}^0$ , the kinetic matrix  $\mathbf{K}$ , the symmetrization matrix  $\mathbf{C}$ , and the grouping matrix  $\mathbf{G}$  have to be considered additionally.

I have chosen two possible penalty functions to calculate the INPHARMA energy. The following first one is based on the example used in [30] and similarly in [22]. The first penalty function *well* is used during the following derivation. The second penalty function *pwell* is introduced at the end of this chapter. The calculation of the derivative, in general, is not limited to any specific penalty function and can be easily adapted.

The function *well* is here defined as

$$well(cal, exp, err, f_{\text{small}}) \equiv \begin{cases} f_{\text{small}}((exp - err) - cal) & \text{if } cal \leq exp - err, \\ cal - (exp + err) & \text{if } cal \geq exp + err, \\ 0 & \text{else.} \end{cases} \quad (2.14)$$

A measure of fit is the R-factor

$$R = \frac{\sum_{m=1}^M \sum_{\alpha\beta} w_{m,\alpha,\beta} \cdot well(\mathbf{A}_{m,\alpha,\beta}^{\text{cal}}, k_m^s \mathbf{A}_{m,\alpha,\beta}^{\text{exp}}, \mathbf{A}_{m,\alpha,\beta}^{\text{err}})}{\sum_{m=1}^M \sum_{\alpha\beta} w_{m,\alpha,\beta} \cdot k_m^s \mathbf{A}_{m,\alpha,\beta}^{\text{exp}}}, \quad (2.15)$$

$$= \frac{\sum_{m=1}^M \sum_{\alpha\beta} w_{m,\alpha,\beta} \cdot well(\mathbf{A}_{m,\alpha,\beta}^{\text{cal}}, k_m^s \mathbf{A}_{m,\alpha,\beta}^{\text{exp}}, \mathbf{A}_{m,\alpha,\beta}^{\text{err}})}{R_D}$$

based on [22, 30]. The index  $m$  is denoting the spectrum (or mixing time), the indexes  $\alpha$  and  $\beta$  are denoting peak groups (or columns) in  $\mathbf{G}$  (nuclei with the same or similar chemical shift, peaks). The experimental and calculated peak volumes are accordingly indexed by the spectrum index  $m$  and the index pair  $\alpha, \beta$  of the cross-peak. For simplicity, the denominator will be called  $R_D$ .

To account for a unknown proportionality factor between the NMR-spectrum-derived peak volumes  $A_{m,\alpha,\beta}^{\text{exp}}$  and the calculated peak volumes  $A_{m,\alpha,\beta}^{\text{cal}}$  a calibration factor for each spectrum (and thus for each mixing time) is introduced. The calibration factor is defined as

$$k_m^s = \frac{\sum_{\alpha,\beta} \mathbf{A}_{m,\alpha,\beta}^{\text{cal}}}{\sum_{\alpha,\beta} \mathbf{A}_{m,\alpha,\beta}^{\text{exp}}} \quad (2.16)$$

where the sum runs over all well determined peaks  $A_{m,\alpha,\beta}^{\text{exp}}$  which are the ones at least three times as big as their error estimate. When the different spectra of a series of experiments should have a common calibration factor (for experimental reasons) this factor can be computed by summing over all well determined peaks of all spectra.

The individual weights are

$$w_{m,\alpha,\beta} = \frac{1}{\mathbf{A}_{m,\alpha,\beta}^{\text{exp}}}. \quad (2.17)$$

They are used to increase the the influence of small peaks on the R-factor.<sup>22</sup>

In the following, I am using the single atom position  $\mathbf{r}_i$  and the distance  $r_{ij}$  between two atoms  $i$  and  $j$ .

As for now, it is assumed that only a single mixing time, a single spectrum, is used (the index  $m$  is not denoting spectra anymore).

The derivate of  $E_{\text{INPH}}$  with respect to the Cartesian coordinates of the protons can be calculated as follows, following the chain rule of differentiation:

$$\frac{\partial E}{\partial \mathbf{r}_i} = \sum_{\alpha\beta klmn j} \frac{\partial E}{\partial \mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \frac{\partial \mathbf{A}_{kl}}{\partial \mathbf{R}_{mn}^{*s}} \frac{\partial \mathbf{R}_{mn}^{*s}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \mathbf{r}_i}. \quad (2.18)$$

$\partial E / \partial \mathbf{r}_i$  denotes the derivative of the energy with respect to all Cartesian components of the position vector  $\mathbf{r}_i$ .

The derivative with respect to the coordinates will be calculated in steps.

The first step is to calculate the derivative of the energy with respect to the

individual-atom cross peaks:

$$\begin{aligned}
\mathbf{F}_{kl} &= \sum_{\alpha\beta} \frac{\partial E}{\partial \mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \\
&= \sum_{\alpha\beta} \frac{k_{\text{INPH}} w_{\alpha\beta}}{R_D} \left. \frac{\partial \text{well}}{\partial \text{cal}} \right|_{\text{cal}=\mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \\
&= \sum_{\alpha\beta} \frac{k_{\text{INPH}} w_{\alpha\beta}}{R_D} \begin{cases} -f_{\text{small}} & \text{if } \text{cal} \leq \text{exp} - \text{err}, \\ +1 & \text{if } \text{cal} \geq \text{exp} + \text{err}, \\ 0 & \text{else.} \end{cases} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}}.
\end{aligned} \tag{2.19}$$

The derivative of a grouped peak volume with respect to the non-grouped peak volumes is derived easily

$$\begin{aligned}
\mathbf{A}_{\alpha\beta}^g &= \sum_{kl} \mathbf{G}_{\alpha k} \mathbf{A}_{kl} \mathbf{G}_{l\beta}^{\text{tr}} \\
\frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} &= \sum_{nm} \mathbf{G}_{\alpha n} \frac{\partial \mathbf{A}_{nm}}{\partial \mathbf{A}_{kl}} \mathbf{G}_{m\beta}^{\text{tr}} \\
&= \sum_{nm} \mathbf{G}_{\alpha n} \delta_{nk} \delta_{ml} \mathbf{G}_{m\beta}^{\text{tr}} \\
&= \mathbf{G}_{\alpha k} \mathbf{G}_{l\beta}^{\text{tr}} \\
&= \mathbf{G}_{\alpha k} \mathbf{G}_{\beta l} \\
&= \begin{cases} 1 & k \in \text{peakgroup}_{\alpha} \text{ and } l \in \text{peakgroup}_{\beta}, \\ 0 & \text{else.} \end{cases}
\end{aligned} \tag{2.20}$$

$\text{peakgroup}_{\alpha}$  and  $\text{peakgroup}_{\beta}$  are the sets of individual atoms of the respective peaks  $\alpha$  and  $\beta$ . Using this result I gain

$$\begin{aligned}
\mathbf{F}_{kl} &= \sum_{\alpha\beta} \frac{\partial E}{\partial \mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \\
&= \sum_{\alpha\beta} k_r w_{\alpha\beta} \begin{cases} -f_{\text{small}} & \text{if } \text{cal} \leq \text{exp} - \text{err}, \\ +1 & \text{if } \text{cal} \geq \text{exp} + \text{err}, \\ 0 & \text{else.} \end{cases} \begin{cases} 1 & k \in \text{Peakgroup}_{\alpha} \text{ and } l \in \text{Peakgroup}_{\beta}, \\ 0 & \text{else.} \end{cases}
\end{aligned} \tag{2.21}$$

To calculate  $\partial E / \partial \mathbf{R}_{mn}^{\text{*s}}$  I define, for readability,  $\mathbf{R} \equiv \mathbf{R}^{\text{*s}} = \mathbf{L}\mathbf{A}\mathbf{L}^{\text{tr}}$ , with

$$\mathbf{R}^{\text{tr}} = \mathbf{R} \text{ and } (\mathbf{R}^N)^{\text{tr}} = \mathbf{R}^N.$$

$$\begin{aligned} \frac{\partial E}{\mathbf{R}_{mn}} &= \sum_{kl} \mathbf{F}_{kl} \frac{\partial \mathbf{A}_{kl}}{\partial \mathbf{R}_{mn}} \\ &= \sum_{kl} \mathbf{F}_{kl} \frac{\partial (\mathbf{C} \exp(-\tau \mathbf{R}) \mathbf{C}^{-1} \mathbf{A}^0)_{kl}}{\partial \mathbf{R}_{mn}} \\ &= \sum_{kl} \mathbf{C}_k \mathbf{F}_{kl} \left[ \frac{\partial}{\partial \mathbf{R}_{mn}} \left( 1 - \tau \mathbf{R} + \frac{1}{2!} \tau^2 \mathbf{R}^2 - \frac{1}{3!} \tau^3 \mathbf{R}^3 + \dots \right)_{kl} \right] \mathbf{C}_l^{-1} \mathbf{A}_l^0 \end{aligned} \quad (2.22)$$

In the next step I use the following lemma, derived in the appendix (9.4, 9.5, p. 144):

$$\frac{\partial (\mathbf{R}^N)_{kl}}{\partial \mathbf{R}_{mn}} = \delta_{mk} (\mathbf{R}^{N-1})_{nl} + \sum_{f=1}^{N-2} (\mathbf{R}^f)_{km} (\mathbf{R}^{N-1-f})_{nl} + \delta_{nl} (\mathbf{R}^{N-1})_{km} \quad (2.23)$$

Now I can gain the result of the derivative of a single summand of the power series (2.22). The factor  $(-\tau)^N/N!$  is skipped here:

$$\sum_{kl} \mathbf{C}_k \mathbf{F}_{kl} \frac{\partial (\mathbf{R}^N)_{kl}}{\partial \mathbf{R}_{mn}} \mathbf{C}_l^{-1} \mathbf{A}_l^0 = \left( \sum_{f=1}^N \mathbf{R}^{f-1} \mathbf{F}^A \mathbf{R}^{N-f} \right)_{mn} \quad (2.24)$$

Here the definition  $\mathbf{F}^A \equiv \mathbf{C} \mathbf{F} \mathbf{A}^0 \mathbf{C}^{-1}$  is used. The details for this derivation can also be found in the appendix (9.6, p. 146).

Now, including the previously skipped factor  $(-\tau)^N/N!$ , the derivative of the energy  $E$  with respect to the symmetrized relaxation matrix  $\mathbf{R} = \mathbf{R}^{*s}$  gets

$$\frac{\partial E}{\partial \mathbf{R}} = \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} \mathbf{R}^{f-1} \mathbf{F}^A \mathbf{R}^{N-f}. \quad (2.25)$$

This can be further simplified using  $\mathbf{R} = \mathbf{R}^{*s} = \mathbf{L} \mathbf{A} \mathbf{L}^{\text{tr}}$  with  $\mathbf{L}^{\text{tr}} \mathbf{L} = \mathbf{I}$  and

so  $\mathbf{R}^N = \mathbf{L}\mathbf{\Lambda}^N\mathbf{L}^{\text{tr}}$ . I gain:

$$\begin{aligned}
\frac{\partial E}{\partial \mathbf{R}} &= \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} (\mathbf{L}\mathbf{\Lambda}\mathbf{L}^{\text{tr}})^{f-1} \mathbf{F}^{\mathbf{A}} (\mathbf{L}\mathbf{\Lambda}\mathbf{L}^{\text{tr}})^{N-f} \\
&= \mathbf{L} \left( \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} \mathbf{\Lambda}^{f-1} \cdot \mathbf{L}^{\text{tr}} \mathbf{F}^{\mathbf{A}} \mathbf{L} \cdot \mathbf{\Lambda}^{N-f} \right) \mathbf{L}^{\text{tr}} \\
&= \mathbf{L} \left( \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} \mathbf{\Lambda}^{f-1} \cdot \mathbf{L}^{\text{tr}} \mathbf{C} \mathbf{F}^{\mathbf{A}} \mathbf{C}^{-1} \mathbf{L} \cdot \mathbf{\Lambda}^{N-f} \right) \mathbf{L}^{\text{tr}} \\
&= -\tau \mathbf{L} \left( \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^{N-1}}{N!} \mathbf{\Lambda}^{f-1} \cdot \mathbf{M}^{\mathbf{A}} \cdot \mathbf{\Lambda}^{N-f} \right) \mathbf{L}^{\text{tr}}.
\end{aligned} \tag{2.26}$$

In the last step of this derivation the definition

$$\begin{aligned}
\mathbf{M}^{\mathbf{A}} &\equiv \mathbf{L}^{\text{tr}} \mathbf{C} \mathbf{F}^{\mathbf{A}} \mathbf{C}^{-1} \mathbf{L} \\
&= \mathbf{L}^{\text{tr}} \mathbf{F}^{\mathbf{A}} \mathbf{L}
\end{aligned} \tag{2.27}$$

is introduced.

Using the Yip and Case series  $\mathbf{J}$  [28, p. 643, formula 10], I get

$$\frac{\partial E}{\partial \mathbf{R}} = -\tau \mathbf{L} \mathbf{J} (\mathbf{M}^{\mathbf{A}}, -\tau \mathbf{\Lambda}) \mathbf{L}^{\text{tr}}, \tag{2.28}$$

where  $\mathbf{J}$  is defined as

$$\mathbf{J}(\mathbf{A}, \mathbf{B}) = \begin{cases} \frac{\mathbf{A}_{ij}}{\mathbf{B}_{ii} - \mathbf{B}_{jj}} [\exp(\mathbf{B}_{ii}) - \exp(\mathbf{B}_{jj})] & \text{if } \mathbf{B}_{ii} \neq \mathbf{B}_{jj} \\ \mathbf{A}_{ij} \exp(\mathbf{B}_{ii}) & \text{if } \mathbf{B}_{ii} = \mathbf{B}_{jj}. \end{cases} \tag{2.29}$$

The next step is to calculate the partial derivative with respect to the distances  $r_{ij}$ :

$$\begin{aligned}
\frac{\partial E}{\partial r_{ij}} &= \sum_{mn} \frac{\partial E}{\partial \mathbf{R}_{mn}} \frac{\partial \mathbf{R}_{mn}}{\partial r_{ij}} \\
&= \sum_{mn} \frac{\partial E}{\partial \mathbf{R}_{mn}} \frac{\partial (\mathbf{C}^{-1} \mathbf{R}^* \mathbf{C})_{mn}}{\partial r_{ij}} \\
&= \sum_{mn} \left( \mathbf{C}^{-1} \frac{\partial E}{\partial \mathbf{R}} \mathbf{C} \right)_{mn} \frac{\partial \mathbf{R}_{mn}^*}{\partial r_{ij}}.
\end{aligned} \tag{2.30}$$

Defining the matrix

$$\mathbf{Q} \equiv \mathbf{C}^{-1} \frac{\partial E}{\partial \mathbf{R}} \mathbf{C} = -\tau \mathbf{C}^{-1} \mathbf{L} \mathbf{J} (\mathbf{M}^{\mathbf{A}}, -\tau \mathbf{\Lambda}) \mathbf{L}^{\text{tr}} \mathbf{C} \tag{2.31}$$

and replacing  $\mathbf{R}^*$  with  $\mathbf{R}^{\text{NOE}} + \mathbf{K}$  the following equation is derived:

$$\begin{aligned}
\frac{\partial E}{\partial r_{ij}} &= \sum_{mn} \mathbf{Q}_{mn} \frac{\partial \mathbf{R}_{mn}^*}{\partial r_{ij}} \\
&= \sum_{mn} \mathbf{Q}_{mn} \frac{\partial (\mathbf{R}^{\text{NOE}} + \mathbf{K})_{mn}}{\partial r_{ij}} \\
&= \sum_{mn} \mathbf{Q}_{mn} \frac{\partial \mathbf{R}_{mn}^{\text{NOE}}}{\partial r_{ij}}.
\end{aligned} \tag{2.32}$$

The derivative  $\partial \mathbf{R}_{mn}^{\text{NOE}} / \partial r_{ij}$  is calculated by considering the formulas for the cross (2.6) and diagonal terms (2.5) of the relaxation matrix  $\mathbf{R}^{\text{NOE}}$ . I get, analogously to formula [15] of Nesterova and Chuprina,<sup>30</sup>  $\partial E / \partial r_{ij}$ :

$$\mathbf{H}_{ij} \equiv \frac{\partial E}{\partial r_{ij}} = -\frac{6}{r_{ij}} [(\mathbf{Q}_{ii} + \mathbf{Q}_{jj}) S_{ij}^{\text{R}} d_{ij} + (\mathbf{Q}_{ij} + \mathbf{Q}_{ji}) \mathbf{R}^{\text{NOE}}] \tag{2.33}$$

with  $S_{ij}^{\text{R}} = \frac{1}{10} \gamma_i^2 \gamma_j^2 \left(\frac{\hbar \mu_0}{4\pi}\right)^2 r_{ij}^{-6} [J_{ij,0} + 3J_{ij,1} + 6J_{ij,2}]$ .

Now the derivative with respect to the Cartesian coordinates of all atoms can be calculated:

$$\frac{\partial E}{\partial x_i} = \sum_j \frac{\partial E}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial x_i} = \sum_j \mathbf{H}_{ij} \frac{x_i - x_j}{r_{ij}} \tag{2.34a}$$

$$\frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial y_i} = \sum_j \mathbf{H}_{ij} \frac{y_i - y_j}{r_{ij}} \tag{2.34b}$$

$$\frac{\partial E}{\partial z_i} = \sum_j \frac{\partial E}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial z_i} = \sum_j \mathbf{H}_{ij} \frac{z_i - z_j}{r_{ij}} \tag{2.34c}$$

Until now, in the derivation of the derivative, it has been assumed that we only have a single mixing time and a single experimental spectrum. To consider several mixing times the grouped peak volumes for the new mixing time,  $\mathbf{A}^g$ , the  $\mathbf{F}$  matrix and, based on this, the  $\mathbf{Q}$  matrix have to be recalculated. Before calculating the forces the  $\mathbf{Q}$  matrices for the various mixing times have to be summed. Then  $\mathbf{H}$  and the forces should be calculated. The eigenvalues  $\mathbf{\Lambda}$ , eigenvectors  $\mathbf{L}$  and the relaxation matrices  $\mathbf{R}^{\text{NOE}}$  and  $\mathbf{R}^{*s} = \mathbf{R}$  can be reused. The most expensive, time consuming step, the calculation of eigenvectors and eigenvalues, can be avoided.

### 2.2.1 Alternative Penalty Functions

The penalty function used until now was based on the *well* function (2.14):<sup>22,30</sup>

$$well(cal, exp, err, f_{small}) \equiv \begin{cases} f_{small} ((exp - err) - cal) & \text{if } cal \leq exp - err, \\ cal - (exp + err) & \text{if } cal \geq exp + err, \\ 0 & \text{else} \end{cases}$$

Its derivative with respect to the calculated peak volume is

$$\frac{\partial well(cal, exp, err, f_{small})}{\partial cal} = \begin{cases} -f_{small} & \text{if } cal \leq exp - err, \\ +1 & \text{if } cal \geq exp + err, \\ 0 & \text{else.} \end{cases} \quad (2.35)$$

The function *well* has the property of a discontinuous derivative at the point where the difference between calculated and experimental value leave the error bound. As shown later (ch. 4.1.3) this is problematic in the context of gradient-based Powell minimization using XPLOR-NIH (Powell minimization in XPLOR-NIH is a generic routine and, in contrast to Powell's method,<sup>38</sup> also uses full gradients).

To circumvent these problems I defined a new, slightly modified function "power well" or *power well*. It uses a function of the form  $a \cdot cal^b$  to create a transition area which gurantees a smooth transition from a non-penalty to a penalty situation. Substituting the *well* against the *power well* function in the calculation of the allover R-Factor leads to a new penalty energy. Now the energy as well as its derivative are continuous. The size of the transition area is specified in multiples of the experimental error bound as exemplified in figure 2.1 and in the following figures. Special attention should be given to the fig. 2.2 where a parabolic profile in the transition area is attained as  $k_i = k_o$ . The non-penalty area is reduced to a single point, later shown to support good structural convergence (ch. 4.1.3).

Here, the mathematical description of the function "power well" or *power well*:

$$\begin{aligned}
p_{well}(cal, exp, err, k_i, k_o, f_{small}) \equiv & \\
\left\{ \begin{array}{ll}
a \cdot (cal - (exp + (1 - k_i) \cdot err))^b & \text{if } (cal \leq exp + err \cdot (1 + k_o) \\
& \text{and} \\
& cal \geq exp + err \cdot (1 - k_i)), \\
f_{small} \cdot a \cdot (-cal - (-exp + (1 - k_i) \cdot err))^b & \text{if } (cal \geq exp - err \cdot (1 + k_o) \\
& \text{and } cal \leq exp - err \cdot (1 - k_i)), \\
well(cal, exp, err, f_{small}) & \text{else.}
\end{array} \right. \tag{2.36}
\end{aligned}$$

with

$$x_0 = err(k_i + k_o) \tag{2.37a}$$

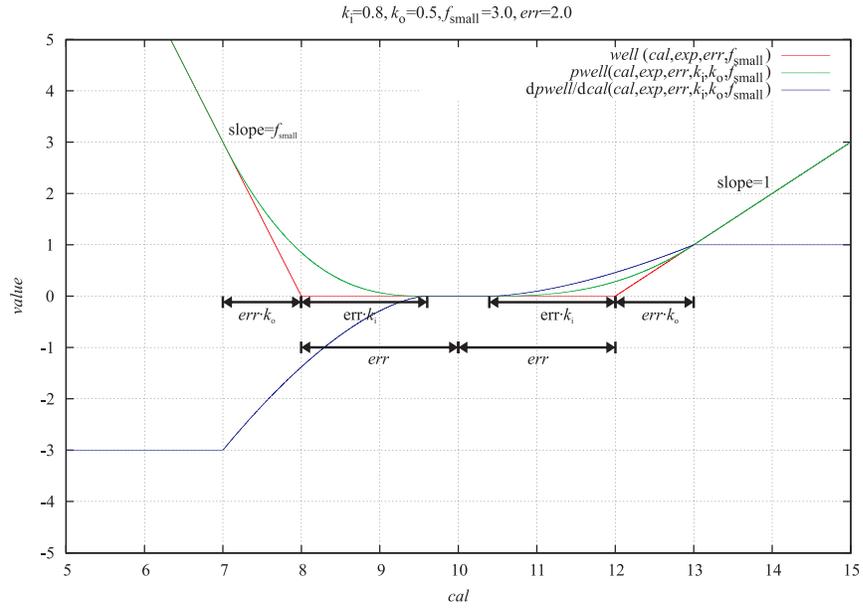
$$y_0 = err \cdot k_o \tag{2.37b}$$

$$b = \frac{x_0}{y_0} \tag{2.37c}$$

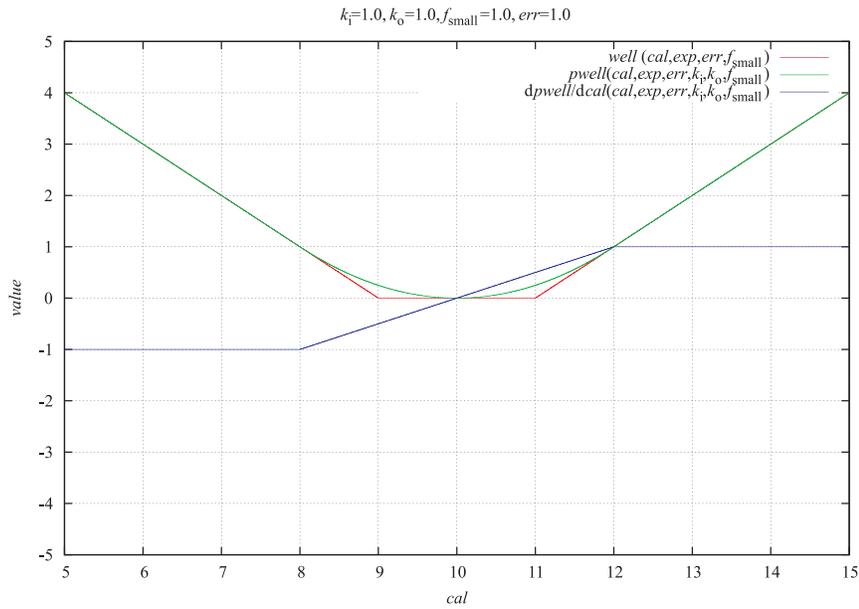
$$a = \frac{y_0}{x_0^b} \tag{2.37d}$$

and its derivative

$$\begin{aligned}
\frac{\partial p_{well}(cal, exp, err, k_i, k_o, f_{small})}{\partial cal} = & \\
\left\{ \begin{array}{ll}
a \cdot b \cdot (cal - (exp + (1 - k_i) \cdot err))^{b-1} & \text{if } (cal \leq exp + err \cdot (1 + k_o) \\
& \text{and} \\
& cal \geq exp + err \cdot (1 - k_i)), \\
-f_{small} \cdot a \cdot b \cdot (-cal - (-exp + (1 - k_i) \cdot err))^{b-1} & \text{if } (cal \geq exp - err \cdot (1 + k_o) \\
& \text{and } cal \leq exp - err \cdot (1 - k_i)), \\
\partial well(cal, exp, err, f_{small}) / \partial cal & \text{else.}
\end{array} \right. \tag{2.38}
\end{aligned}$$



**Figure 2.1:** Here examples of a *well* and a corresponding *pwell* function and its derivative are given. The smooth transition area around the error bound is specified in multiples of the error bound. The slope towards the negative *cal*-axis can be scaled using the factor  $f_{small}$ . The parameters are given on the top of the figure. The experimental value *exp* is 10.



**Figure 2.2:** In this example with  $k_i = 1$  only a single point without any penalty for *pwell* is existing. A parabolic profile is attained as  $k_i = k_o$  (2.37a).

Further examples of *pwell* functions with different parameters can be found in the appendix (ch. 9.3, p. 146).

# Chapter 3

## XPLOR-NIH Implementation

### 3.1 INPHARMA Gradient in a Nut Shell

The sequence of operations for calculating the gradient of the INPHARMA energy, following the results of the previous chapter 'Theory', is the following:

1. Setting up the effective relaxation matrix  $\mathbf{R} \equiv \mathbf{R}^{*s}$  and initial peak volume matrix  $\mathbf{A}^0$ , chemical exchange matrix  $\mathbf{K}$ , symmetrization matrix  $\mathbf{C}$ , and the grouping matrix  $\mathbf{G}$  (2.1-2.9).
2. Calculating the eigenvectors  $\mathbf{L} \equiv \boldsymbol{\chi}^{*s}$  and eigenvalues  $\boldsymbol{\Lambda} \equiv \boldsymbol{\lambda}^{*s}$ .
3. Calculating the matrix of grouped peak volumes (2.11) for a given mixing time  $\tau_m$

$$\mathbf{A}_m^g = \mathbf{GCL} \exp(-\boldsymbol{\Lambda}\tau_m)\mathbf{L}^{\text{tr}}\mathbf{C}^{-1}\mathbf{A}^0\mathbf{G}^{\text{tr}}. \quad (3.1)$$

4. Calculating the  $\mathbf{F}$  matrix (2.21) (for the given  $\tau_m$ )

$$\mathbf{F}_{kl} = \sum_{\alpha\beta} k_r w_{\alpha\beta} \begin{cases} -f_{small} & \text{if } cal \leq exp - err, \\ +1 & \text{if } cal \geq exp + err, \\ 0 & \text{else.} \end{cases} \cdot \begin{cases} 1 & k \in \text{Peakgroup}_\alpha \text{ and } l \in \text{Peakgroup}_\beta, \\ 0 & \text{else.} \end{cases} \quad (3.2)$$

5. Calculating the matrix  $\mathbf{M}^A$  (2.27) (for the given  $\tau_m$ )

$$\mathbf{M}^A = \mathbf{L}^{\text{tr}}\mathbf{CFA}^0\mathbf{C}^{-1}\mathbf{L}. \quad (3.3)$$

6. Calculating the matrix  $\mathbf{J}(\mathbf{M}^A, -\tau\mathbf{\Lambda})$  (2.29) (for the given  $\tau_m$ )

$$\mathbf{J}(\mathbf{A}, \mathbf{B}) = \begin{cases} \frac{\mathbf{A}_{ij}}{\mathbf{B}_{ii} - \mathbf{B}_{jj}} [\exp(\mathbf{B}_{ii}) - \exp(\mathbf{B}_{jj})] & \text{if } \mathbf{B}_{ii} \neq \mathbf{B}_{jj} \\ \mathbf{A}_{ij} \exp(\mathbf{B}_{ii}) & \text{if } \mathbf{B}_{ii} = \mathbf{B}_{jj}. \end{cases} \quad (3.4)$$

7. Calculating the matrix  $\mathbf{Q}$  (2.31) (for the given  $\tau_m$ )

$$\mathbf{Q} = -\tau\mathbf{C}^{-1}\mathbf{L}\mathbf{J}(\mathbf{M}^A, -\tau\mathbf{\Lambda})\mathbf{L}^{\text{tr}}\mathbf{C}. \quad (3.5)$$

8. Calculating further matrices  $\mathbf{Q}$  for the remaining mixing times, starting at (3.1). Summing them up to gain the final matrix  $\mathbf{Q}$ .
9. Calculating the matrix  $\mathbf{H}$  (2.33)

$$\mathbf{H}_{ij} = -\frac{6}{r_{ij}} [(\mathbf{Q}_{ii} + \mathbf{Q}_{jj}) S_{ij}^s + (\mathbf{Q}_{ij} + \mathbf{Q}_{ji}) \mathbf{R}^{\text{NOE}}] \quad (3.6)$$

$$\text{with } S_{ij}^s = \frac{1}{10} \gamma_i^2 \gamma_j^2 \left(\frac{\hbar\mu_0}{4\pi}\right)^2 r_{ij}^{-6} [J_{ij,0}^s + 3J_{ij,1}^s + 6J_{ij,2}^s].$$

10. Calculating the forces, like here for  $x$  (2.34)

$$\frac{\partial E}{\partial x_i} = \sum_j \mathbf{H}_{ij} \frac{x_i - x_j}{r_{ij}}. \quad (3.7)$$

This procedure is valid if several spectra have to be calculated only differing in mixing time. If any of the parameters influencing the relaxation matrices have changed - like the field, concentrations or kinetic constants - the effective relaxation matrix has to be set-up again and diagonalization is necessary. Then, the resulting  $\mathbf{H}$  matrices for different spectra should be summed before the calculation of the forces.

## 3.2 Numerical Approximations

For a fast computation of the matrix derivative it is important to be able to use the same speed-up techniques as for calculating the peak volumes,<sup>11,12</sup> namely only using and determining the  $M \leq N$  *smallest* eigenvalues and corresponding eigenvectors of the symmetrized effective relaxation matrix  $\mathbf{R}^{*s}$  of size  $N \times N$ .  $N$  is the total number of all considered individual atoms in both complexes and both free ligands. The number of eigenvalues  $M$  required for reliable simulation of INPHARMA spectra and INPHARMA gradients can be significantly lower ( $\approx 10$  fold with a 8.5 Å cutoff in the

tubulin system than the total number of eigenvalues  $N$ . The required computation time is roughly the fraction  $M/N$  of the time required for all eigenvalues.

The technique is justified by the fact that, according to (2.11), the matrix exponential of  $\mathbf{\Lambda}$  (the diagonal matrix of the eigenvalues of the symmetrized effective relaxation matrix) multiplied by  $-\tau_m$  has to be calculated:

$$\mathbf{E} \equiv \exp(-\tau_m \mathbf{\Lambda}) \quad (3.8)$$

The matrix exponential  $\mathbf{E}$  is diagonal, again. Its elements for larger eigenvalues quickly become vanishingly small. In further multiplications, these small elements, and vectors (columns and rows of matrices) multiplied with them, can be neglected without affecting the all-over result.

To calculate the number of necessary eigenvectors, first the trace of  $\mathbf{E}$ ,  $E_{\text{sum}}$ , is calculated for the shortest of all mixing times  $\tau_{\text{min}}$ , requiring the most eigenvectors. For this, only once, all the  $N$  eigenvalues, not the eigenvectors, for the system will be calculated. They will be sorted in a way that  $\mathbf{\Lambda}_{i-1} \leq \mathbf{\Lambda}_i$ , with  $\mathbf{\Lambda}_i \equiv \mathbf{\Lambda}_{ii}$ .

$$E_{\text{sum}} \equiv \text{trace}(\mathbf{E}) = \sum_i^N \exp(-\mathbf{\Lambda}_i \tau_{\text{min}}) \quad (3.9)$$

Now, the number of eigenvalues and eigenvectors  $I$  to be used as input for the diagonalization routine can be determined from the smallest  $I$  fulfilling the following condition with an arbitrary  $0 \leq k_{\text{esum}} \leq 1$ :

$$E_{\text{sum}}^{I \leq N} = \sum_{i=1}^I \exp(-\mathbf{\Lambda}_i \tau_{\text{min}}) \geq k_{\text{esum}} \cdot E_{\text{sum}}. \quad (3.10)$$

$k_{\text{esum}}$  is the fraction of the trace of  $\mathbf{E}$  ( $E_{\text{sum}}$ ) which has to be explained by the number of required eigenvalues  $I$ . As this approximation should only be done once, requiring the calculation of all eigenvalues which needs some time, an additional security factor is applied to gain the number of actually calculated eigenvalues  $M = \min(f \cdot I, N)$ .  $f$  is normally chosen to be 1.1. For the small 6-spin test system it had to be 1.5 (to use 3 instead of 2 eigenvalues).

In all following diagonalizations, using modified structural models, the number of  $M$  eigenvectors and eigenvalues will be determined.

For longer mixing time spectra even less eigenvalues can be used for the matrix multiplications as less eigenvalues, multiplied by  $-\tau_m$ , still give rise to significant values of the exponential  $\exp(-\Lambda_i \tau_m)$ . To consider this effect the condition (3.9) has been applied analogously, using the trace of the  $M$  size matrix exponential of the actually calculated eigenvalues,  $E_{\text{sum}}^M$ , as reference and choosing the smallest  $p_m$  fulfilling the condition

$$E_{\text{sum}}^{p_m} = \sum_{i=1}^{p_m} \exp(-\Lambda_i \tau_m) \geq k_{\text{esum2}} \cdot E_{\text{sum}}^M, \quad (3.11)$$

with  $k_{\text{esum2}} = 0.99999$ .

This means that for every mixing time a different number  $p_m$  of  $M$  available eigenvectors and eigenvalues will be used for all the following matrix multiplications.

The eigenvalues will always be sorted to fulfill the condition  $\Lambda_{i-1} \leq \Lambda_i$ . The smallest ones should be "on the left hand side" of the diagonal matrix  $\Lambda$ . The eigenvectors are sorted accordingly.

Although all vanishing elements of the  $N \times N$ -sized matrices  $\mathbf{A}$  and  $\mathbf{L}$  could just be set to zero, further speed-up is achieved by considering the actually relevant matrix sizes, for all or just a few of the considered matrices. For the matrices  $\mathbf{A}$  and  $\mathbf{L}$  this sizes would be  $M \times M$  and  $N \times M$ , respectively. This approach is facilitated by the chosen general form of the NOE relaxation matrix  $\mathbf{R}^{\text{NOE}}$  and the corresponding single-atom peak volume matrix  $\mathbf{A}$ . Only the free ligand concentrations are used for grouping and thus only the  $L \times L$  ( $L$  is the number of protons in both ligands) single-atom cross-peaks in the upper left corner of  $\mathbf{A}$  have to be computed instead of  $N \times N$ .

Explaining all the details for all matrix multiplications would lead to far in the context of this thesis. The by far largest fraction of CPU time is used for the matrix diagonalization. In general, the following matrix multiplications only contribute a few percent to the allover CPU time. Considering the sizes of the matrices explicitly can reduce this fraction of needed CPU time to  $< 1\%$ . The used matrix multiplication algorithm `dgemm` from the BLAS library (included in Intel's MKL library) is equipped with respective arguments to facilitate the multiplication of parts of matrices. In the implementation of INPHARMA restraints into XPLOR-NIH all the relevant matrix sizes were considered and the sequence of matrix chain multiplications was optimized.



### 3.3 Minimum Example

Normally XPLOR-NIH will be executed by piping a script into it. Examples of protocols for using INPHARMA with XPLOR-NIH are given in the appendix (8, p. 143), referring to the applications in the following chapters. Here, the result of executing the XPLOR-NIH implemented HELP command of INPHarma is given. If the script INPH HELP END STOP is the content of the file `script.inp` it can be executed with the command `xplor < script.inp > script.out`. It gives a complete list of INPHARMA subcommands and the respective options:

XPLOR-NIH version 2.20-custom

C.D. Schwieters, J.J. Kuszewski,	Progr. NMR Spectr. 48, 47-62 (2006).
N. Tjandra, and G.M. Clore	J. Magn. Res., 160, 66-74 (2003).
<a href="http://nmr.cit.nih.gov/xplor-nih">http://nmr.cit.nih.gov/xplor-nih</a>	based on X-PLOR 3.851 by A.T. Brunger

```

User: mare          on: linux-gues(x86/Linux      )  at: 12-Feb-10 22:30:34
X-PLOR>INPH HELP END STOP
INPHarma {<inph-statement>} END
<INPH-statement>:== {
  TAUF <real>
  ! free ligands' correlation time /s
  ! {default = 100e-12}
  TAUB <real>
  ! complexes' correlation time /s
  ! {default = 1000e-9}
  K12 <real>
  ! kinetik constant for L1T+L2 -> L2T+L1
  ! usually in (micromolar*second)^-1
  ! {default = 1e5}
  K21 <real>
  ! kinetik constant fro L1T+L2 <- L2T+L1
  ! usually in (micromolar*second)^-1
  ! {default = 1e5}
  CL10 <real>
  ! initial ligand 1 concentration /micromolar
  ! {default = 100}
  CL20 <real>
  ! initial ligand 2 concentration /micromolar
  ! {default = 100}
  CTO <real>
  ! initial target concentration /micromolar
  ! {default = 1}
  NUE <real>

```

```

! B0 proton frequency / Hz
! {default = 900e6}
CUTOff <real>
! cutoff /Angstrom
! {default = 10}
FORCeconstant <real>
! force constants: <kforce>
! {default = 100}
EXPS <real>
! fraction of exp(-lambda tau_min) trace for eigenvector
! calculation speed-up ( >1.0 ==> Use all eigenvectors )
! ( <=1.0 ==> Use some eigenvectors)
! {default = 1.1}
FUNC <funcarg> <funcarg>
! select penalty function and set parameters
! <funcarg>=WELL for well or <funcarg>=PWEL for power well, {default = WELL}
! WELL: <funcarg>:==:<real>
! slope fsmall for too small peaks, {default = 1.0}
! PWELL: <funcarg>:==<real> <real> <real>
! fsmall, k_i, k_o
! slope fsmall for too small peaks, {default = 1.0}
! factor ki for defining inner transition area, {default = 1.0}
! factor ki for defining outer transition area, {default = 1.0}
REWEight
! switch REWEight ON: Recalculate scales and weight for next E(INPH) evaluation.
! Peakvolumes, forces and correlation are outdated until then.
WRITE
! Write the correlation data, experimental and calculated peakvolumes
! as well as regression results.
MIXIngtime ADD <real> [ADD <real> ...] END
! Define a sequence of mixingtimes /s
! {default = 0.1}
ACTIve <selection> <selection> <selection> <selection> <selection> <selection>
! 6 selections of NMR-active target protons:
! <ligfree1> <ligfree2> <tar1> <lig1> <tar2> <lig2>
GROUp <name> <sel>
! [<name> <sel> ...]
END
! definition of protons with identical
! chemicalshift for grouping matrix
VOLUmes
! <groupname> <groupname> <value> <errorvalue>
! [<groupname> <groupname> <value> <errorvalue> ...]
END
! definition of to be calculated peaks

```

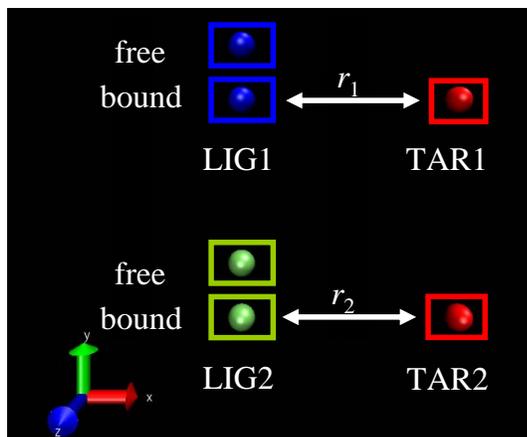
```
! their volumes, and errors
! use <value>=<errorvalue>=-999 for calculating
! the volume, but
! not using it for regression or R-values
}
HEAP:  maximum use= 6015799 current use= 5998137
X-PLOR: total CPU time= 0.0000 s
X-PLOR: entry time at 22:30:34 12-Feb-10
X-PLOR: exit time at 22:30:34 12-Feb-10
```

# Chapter 4

## Results

### 4.1 6-Spin System

To verify the validity of the gradient calculation and to visualize and test the XPLOR-NIH implementation of INPHARMA restraints a most simple "molecular" test system was created (fig. 4.1).



**Figure 4.1:** This is one conformation of the simplest complete INPHARMA spin-system composed of only 6 spins. The ligands (LIG) as well as on the receptors (TAR, target) consist of only a single atom. This system was used to compare the analytical gradients with the numerical ones (ch. 4.1.2, p. 47), minimization and convergence tests and for plotting the energy landscape. The latter was easily possible as only the two internuclear distances  $r_1$  and  $r_2$  define the distance dependent entries of the relaxation matrix.

Here the ligands and the protein are assembled from a single atom. Considering the two free and bound ligands as well as the two protons of the

complexes there are six atoms in total. The relaxation matrix has  $6 \times 6$  entries. There are only two distances (between two protons in the two complexes) which define the matrix, the peak volumes and the R-factor. This makes it easy to sample a relevant part of conformational space of the system and to plot the INPHARMA energy as a function of these two distances (figs. 4.13, 4.14).

The used pdb file to describe the reference coordinates of the system was the following:

ATOM	1	HL1	LIG	1	0.000	8.000	0.000
ATOM	2	HL1	LIG	2	0.000	2.000	0.000
ATOM	3	HR1	REC	3	2.000	6.000	0.000
ATOM	4	HL1	LIG	4	0.000	6.000	0.000
ATOM	5	HR1	REC	5	2.000	0.000	0.000
ATOM	6	HL1	LIG	6	0.000	0.000	0.000

### 4.1.1 One-dimensional Energy Landscape

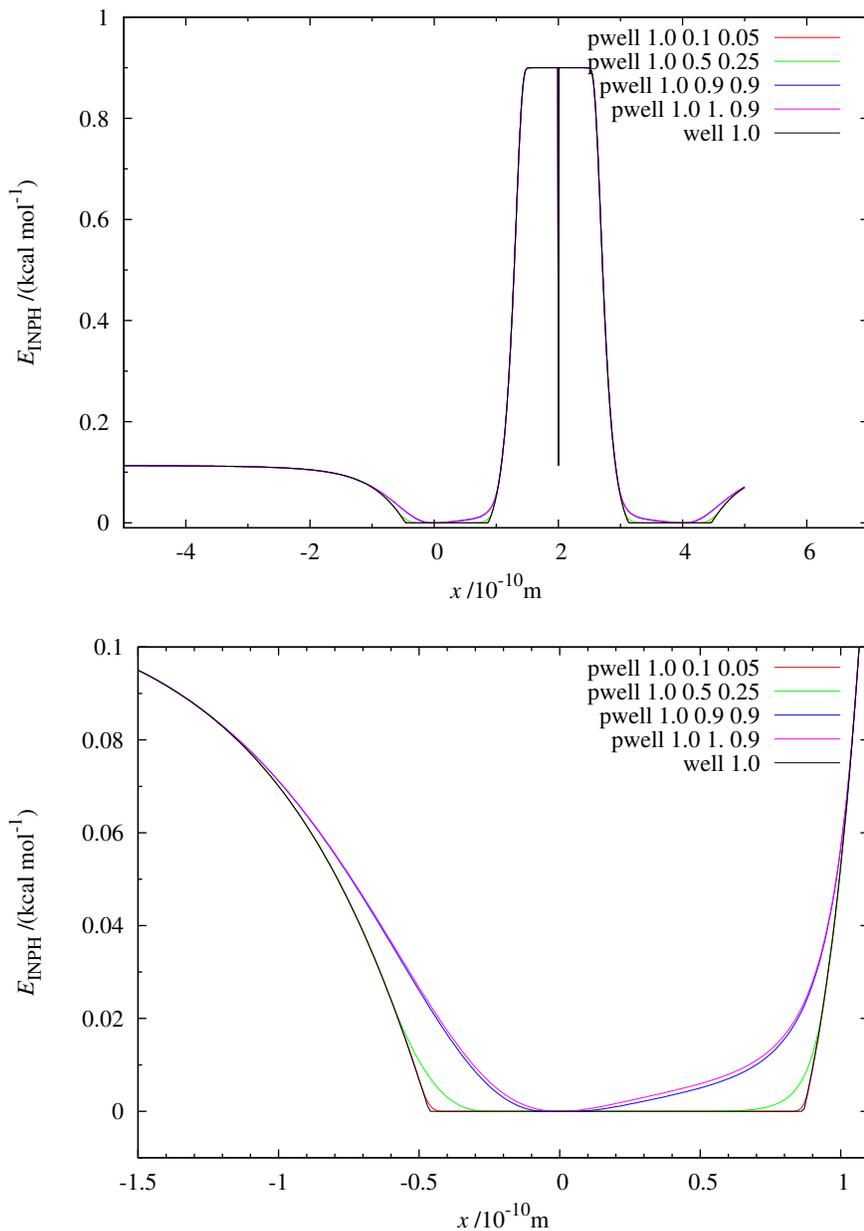
To create a simple, single dimensional plot of the energy landscape first a single reference peak was calculated from the reference conformation. This peak was used as pseudo-experimental input. Moving any atom from its place would potentially change the calculated intensity and so alter the INPHARMA energy  $E_{\text{INPH}}$ . The diagonal peak of atom 1 (ligand 1, free) was selected as the pseudo-experimental peak volume. Atom 4 (ligand 1, bound) was picked to be moved along the  $x$ -axis in the range from  $-5$  to  $5$  Å. At the position  $x = 0$  Å and  $x = 4$  Å its reference fit to its neighbor is achieved and  $E_{\text{INPH}}$  is zero for all penalty functions. At  $x = 2$  Å the distance to the neighbor becomes zero (numerically: *almost* zero) and the value of the energy should be discarded here (2.6, 2.5).

The following parameters for the penalty functions have been chosen to compare the profile of the resulting INPHARMA energies  $E_{\text{INPH}}$  and the derivatives (the parameters are  $f_{\text{small}}$  or  $f_{\text{small}}, k_i, k_o$ ):

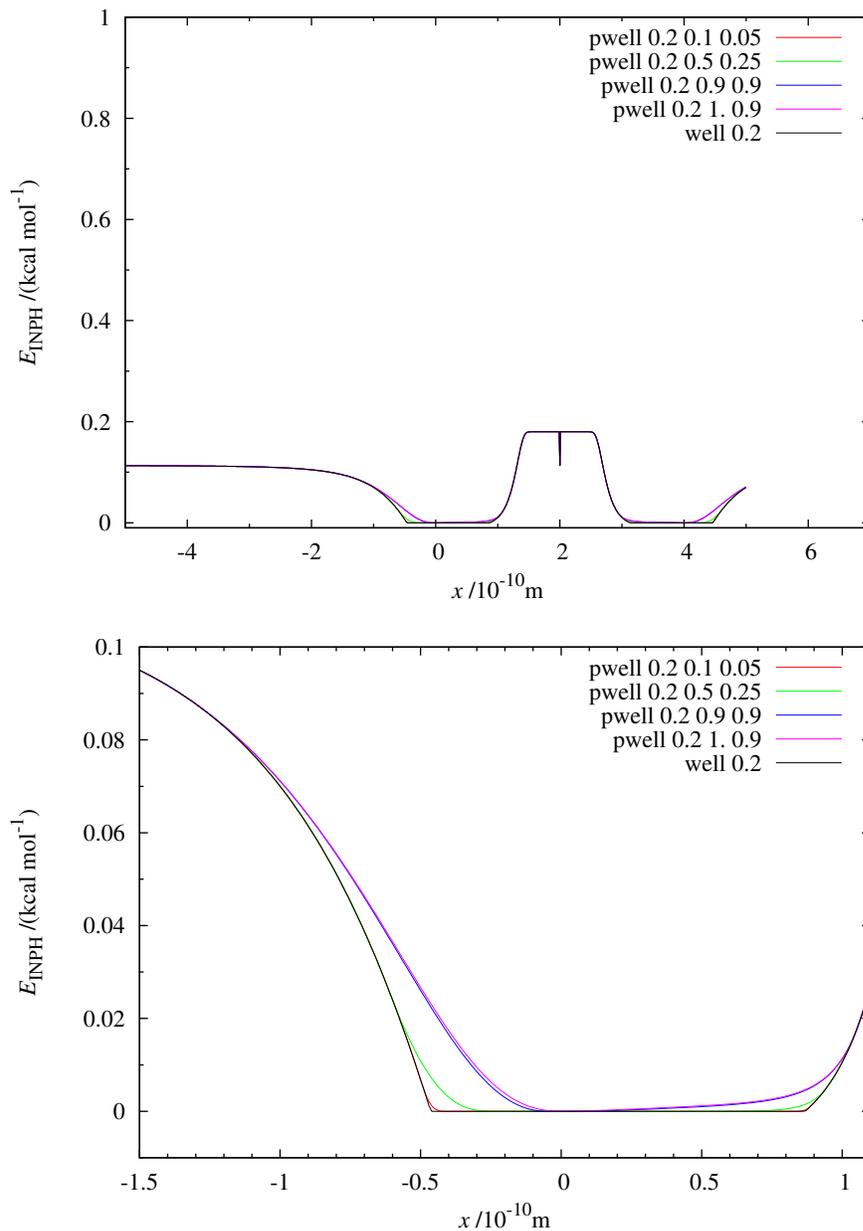
```
FUNC WELL 1.0
FUNC PWEL 1.0 0.1 0.05
FUNC PWEL 1.0 0.5 0.25
FUNC PWEL 1.0 0.9 0.9
FUNC PWEL 1.0 1. 0.9
```

```
FUNC WELL 0.2
FUNC PWEL 0.2 0.1 0.05
FUNC PWEL 0.2 0.5 0.25
FUNC PWEL 0.2 0.9 0.9
FUNC PWEL 0.2 1. 0.9
```

The XPLOR-NIH script to calculate the plots of  $E_{\text{INPH}}$  as a function of the  $x$  coordinate of atom 4 is given in the appendix (ch. 10.1.1). It can be seen that the hardness of the transition between the penalty and non-penalty situation as well as the extent of the transition area and the non-penalty area can be finely tuned using the *pwel* function (fig. 4.2). This also applies when the  $f_{\text{small}}$  factor is used to rescale the penalty of peaks being to small (fig. 4.3). It can be seen that the smooth *pwel* function avoids edges and allows to specify the extent of the non-penalty area precisely. The *well* function shows a distinct kink when the error bound gets violated. The kink is a little bit smeared out in the close-up figures due to the finite sampling interval  $\Delta x = 0.01 \text{ \AA}$ .



**Figure 4.2:** Top: The INPHARMA energies, for the system conformations as described in the text and in fig. 4.1, under a linear move of one atom, is plotted considering only a single diagonal peak. The parameters of the functions *well* and *pwell* are given in the key in the sequence  $f_{\text{small}}, k_i, k_o$  (see fig. 2.1). Bottom: Close-up of the low-penalty area around  $x = 0 \text{ \AA}$ .

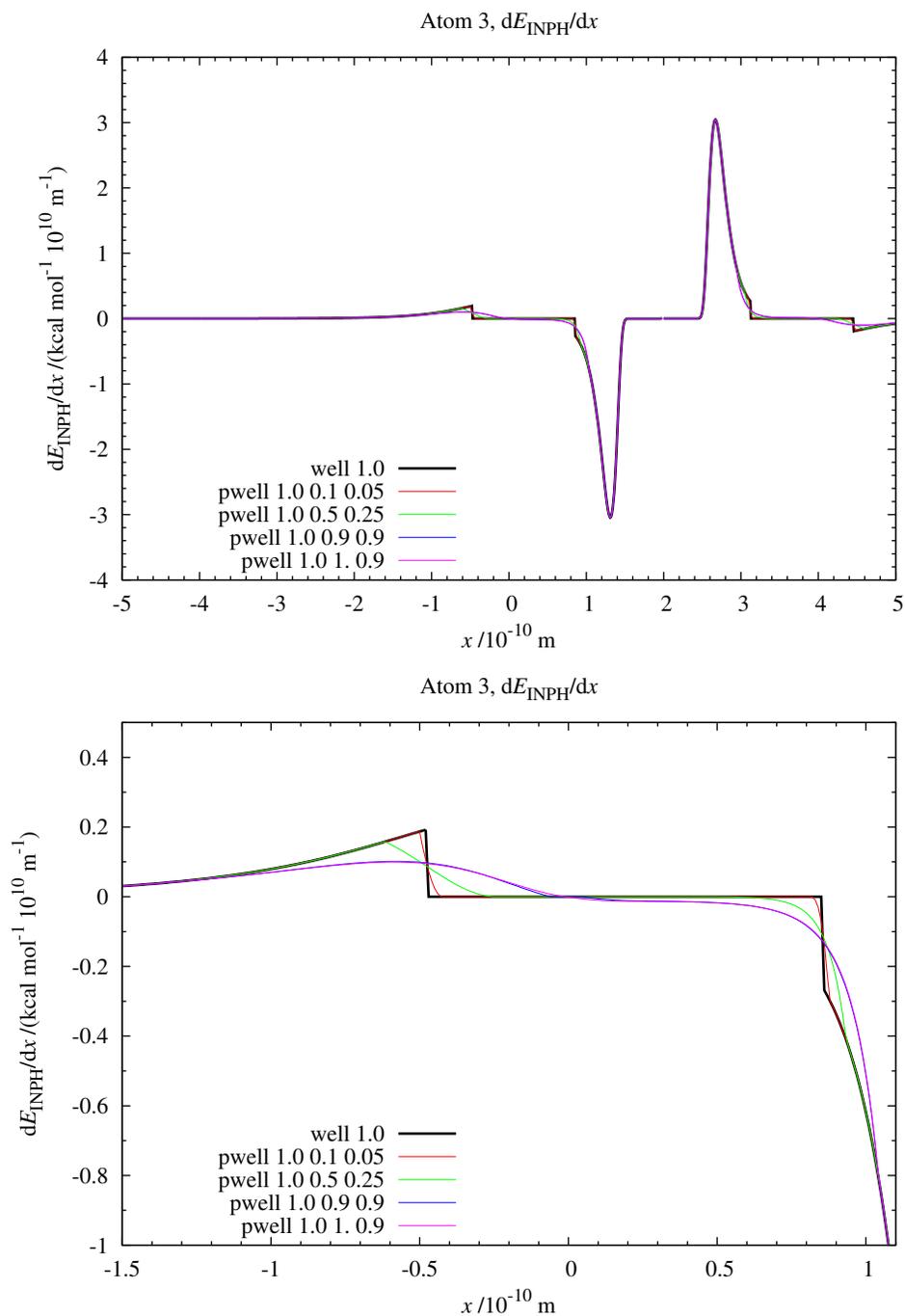


**Figure 4.3:** A figure corresponding to fig. 4.2 is given, using  $f_{\text{small}} = 0.2$  instead of  $f_{\text{small}} = 1.0$ . The penalty is reduced for the coordinates with calculated peak volumes smaller than the reference peak. As a diagonal peak was chosen here, this is the case when the two protons in the complex 1 are closer to each other than in the reference conformation. The shorter distance increases the relaxation rates leading to magnetization loss. The penalty function is not affected for conformations where the calculated peak is bigger than the reference peak ( $x < 0 \text{ \AA}$  or  $x > 4 \text{ \AA}$ ). Bottom: A close-up of the low-penalty area around  $x = 0 \text{ \AA}$  is shown.

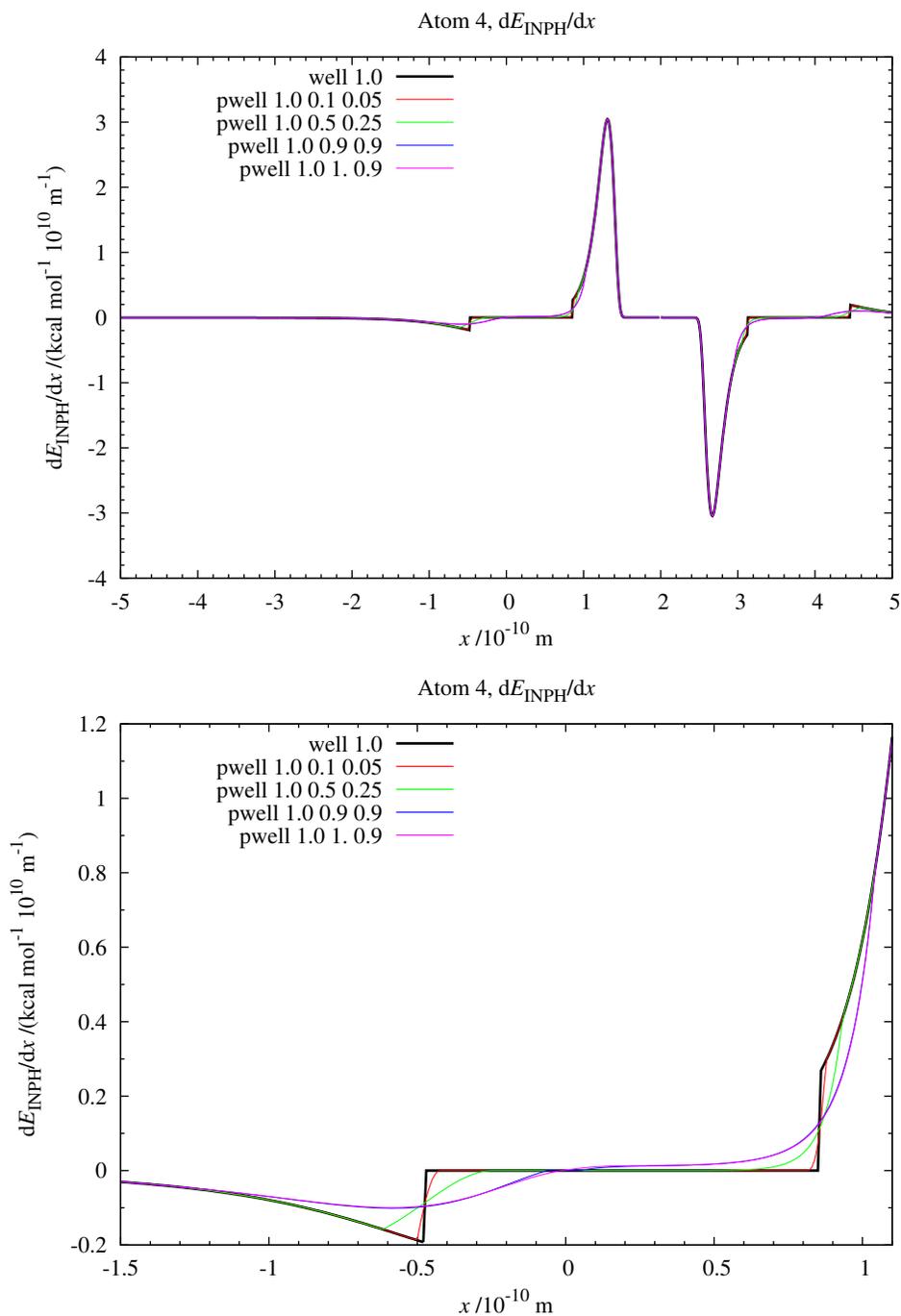
The figures fig. 4.4-4.7 show all the non-zero derivatives occurring during the translation of atom 4. In the main text only the figures for  $f_{\text{small}} = 1$  are given. The corresponding figures for  $f_{\text{small}} = 0.2$  can be found in the appendix (fig. 10.1-10.4, p. 154).

Only the atoms 3, 4, 5 and 6, all forming the two complexes, experience a force which only acts along the  $x$ -axis. The latter can be anticipated as any translation of the atoms of the complexes along the  $y$  or  $z$ -axis does not differentially change the length of the internuclear vector, lying along the  $x$ -axis. The free atoms will not experience any force as they are alone in their species and don not have any neighboring atom, and so do not have any defined distances. This is why their position is irrelevant for the INPHARMA energy.

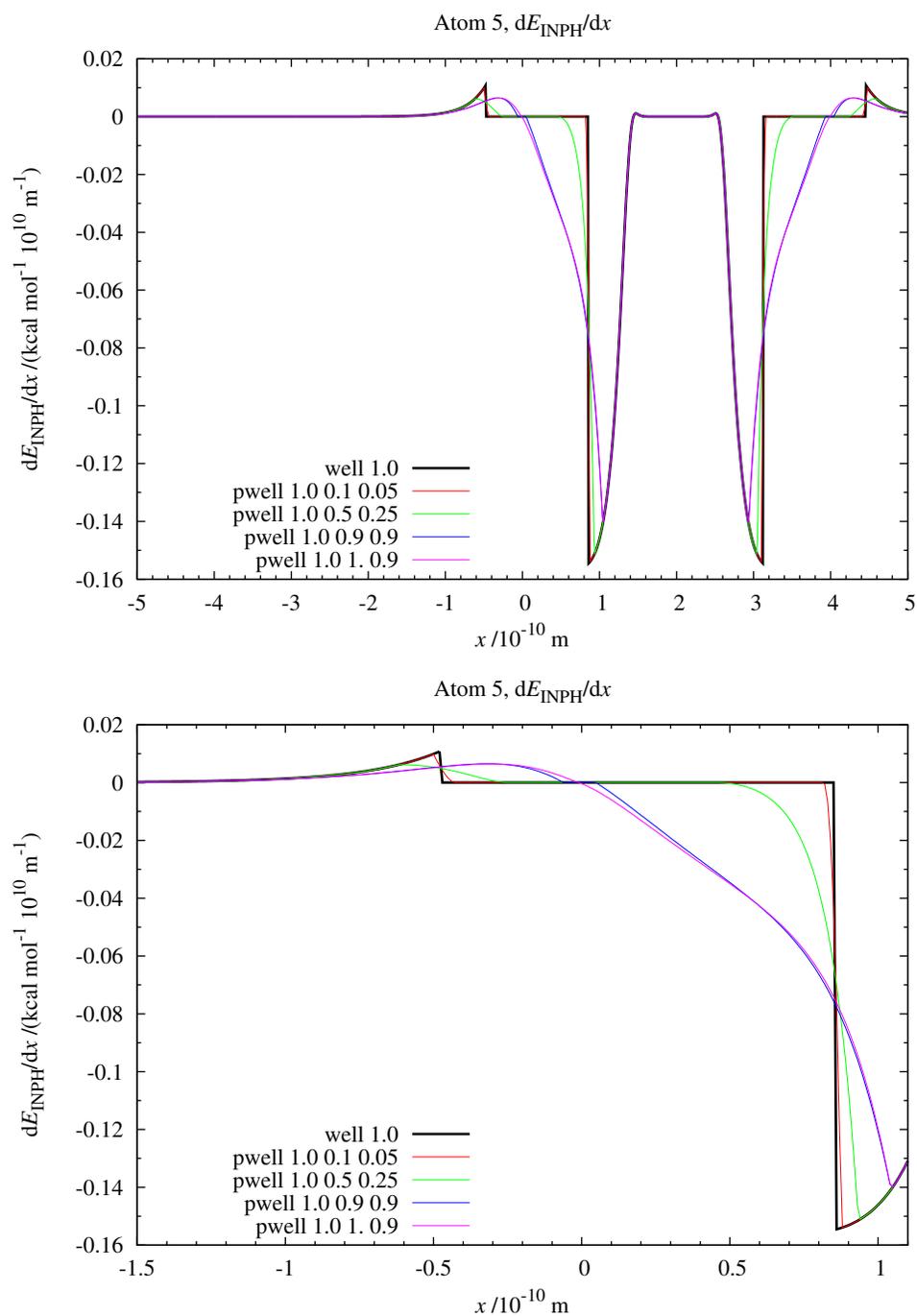
While the *well* function in all cases experiences discrete jumps in the force, these jumps are avoided using any of the *pwell* functions. Depending on the extend of  $k_i$  the extend of the area without any force within the experimental error bound can be tuned. For  $k_i = 1$  there are only two discrete positions without any remaining force.



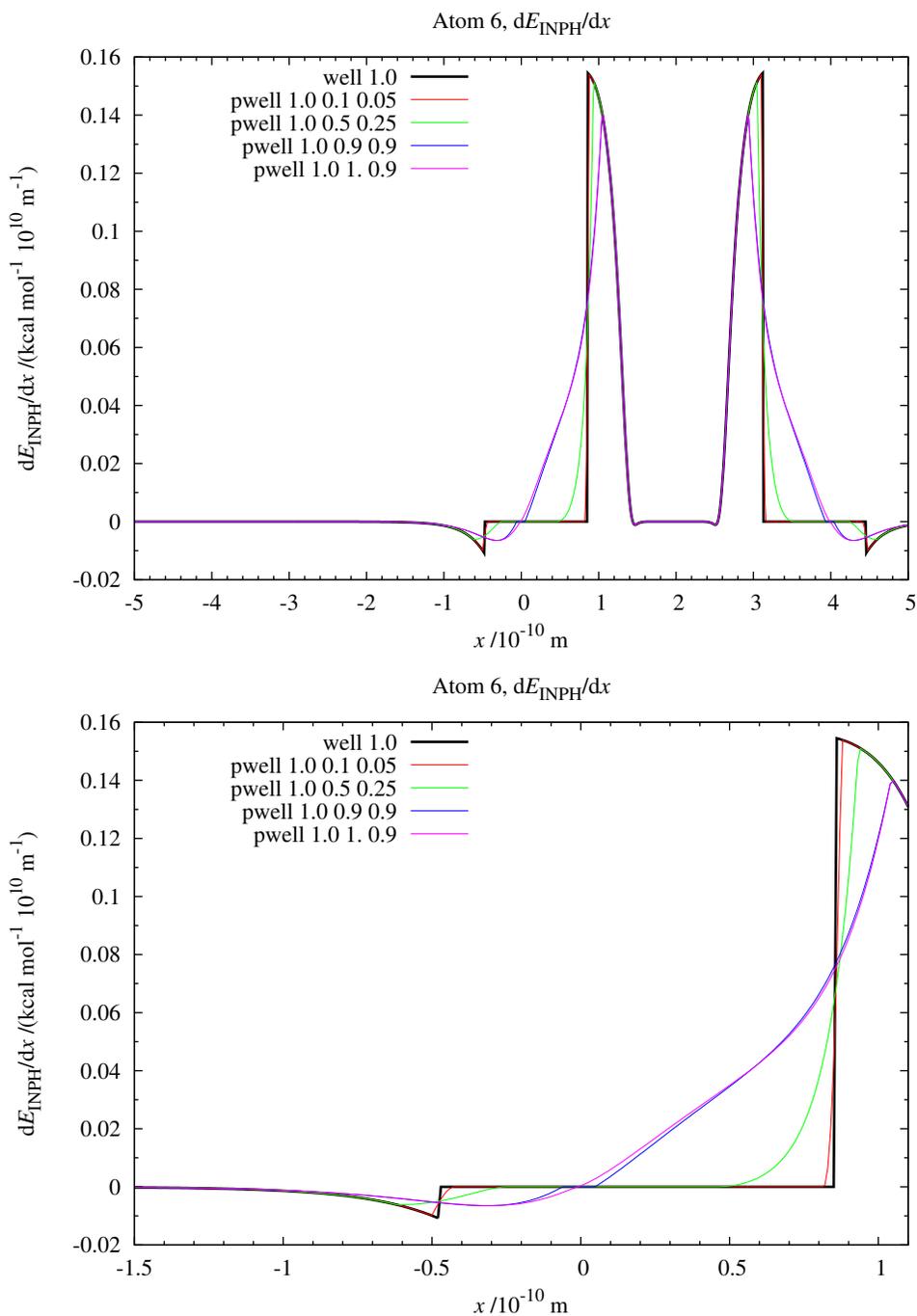
**Figure 4.4:** Top: The derivative of the INPHARMA energies for atom 3 (target 1, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of atom 4 (ligand 1, bound) is plotted considering only a single experimental diagonal peak. The functions *well* and *pwell* have been used with  $f_{\text{small}} = 1.0$ . All their parameters are given in the key in the sequence  $f_{\text{small}}, k_i, k_o$  (see fig. 2.1). Bottom: A close-up of the low-penalty area around  $x = 0 \text{ \AA}$  is shown.



**Figure 4.5:** Top: The derivative of the INPHARMA energies for atom 4 (ligand 1, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of the same atom (atom 4; ligand 1, bound) is plotted considering only a single experimental diagonal peak. Bottom: Close-up of the low-penalty area around  $x = 0$ .



**Figure 4.6:** Top: The derivative of the INPHARMA energies for atom 5 (target 2, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of atom 4 (ligand 1, bound) is plotted considering only a single experimental diagonal peak. Bottom: Close-up of the low-penalty area around  $x = 0$  Å.



**Figure 4.7:** Top: The derivative of the INPHARMA energies for atom 6 (ligand 2, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of atom 4 (ligand 1, bound) is plotted considering only a single experimental diagonal peak. Bottom: Close-up of the low-penalty area around  $x = 0$  Å.

### 4.1.2 Numerical and Analytical Derivatives

I also calculated and sampled the complete derivative vector for all atoms in the relevant part of conformational space. The results of the analytical derivative was compared with the result of an alternatively implemented, but very slow, numerical derivative. The match between analytical and numerical derivative is fulfilled in all situations. All the previous plots of analytical derivatives (figs. 4.4-4.7, 10.1-10.4) can be reproduced using the numerical derivative.

Here, a simple example for the agreement between numerical and analytical derivative is given. In contrast to the plots of the previous section, using a single peak and a single mixing time, here two mixing times and three cross peaks for each mixing time are used.

The numerical gradient was calculated using symmetric finite differences following this equation for all atoms  $i$  along the Cartesian axes  $x$ ,  $y$  and  $z$ :

$$\left. \frac{\partial E_{\text{INPH}}}{\partial x_i} \right|_{x_i=x_i^0} = \frac{E_{\text{INPH}}(x_i^0 + h) - E_{\text{INPH}}(x_i^0 - h)}{2 \cdot h} \quad (4.1)$$

with  $h = 0.005 \text{ \AA}$ . As any finite difference derivative will evaluate the derivative more or less in the middle of an interval the *symmetric* finite difference delivers a rather accurate estimation for the derivative at  $x_i^0$ .

The calculation of this numerical gradient requires the evaluation of two INPHARMA energies for a single entry of the gradient and 6 evaluations for every atom. Each time the effective relaxation matrix needs to be diagonalized. For a small 6-atom test system this time is negligible. For larger systems having matrix sizes above  $1000 \times 1000$ , more than 6000 diagonalizations are required and the numerical evaluation of the gradient becomes almost prohibitively slow or at least impractical. Considering a typical computation time of 0.6s for a single energy evaluation, using a cutoff of 8.5  $\text{\AA}$  and a matrix size of 1000 the calculation of the numerical derivative would require 60 minutes. In contrast, the calculation of the corresponding analytical derivative would only require 0.6 s, being 6000 times faster.

The finite difference gradient only depends on the *values* of the energy function  $E_{\text{INPH}}$ . It does not require any knowledge about the energy's functional dependence on the coordinates, in contrast to the calculation of the analytical gradient. Thus the numerical gradient can be used to check if the analytical gradient is correct for a particular set of coordinates and parameters.

To compare the numerical and analytical gradients, atom 4 (ligand 1, bound)

of the 6-spin system was picked and moved along the  $x$ -axis in the range from  $-5$  to  $5$  Å. At the position  $x = 0$  Å it has its reference position and at  $x = 2$  Å its neighbor atom is placed.

These peak volumes for two mixing times were used:

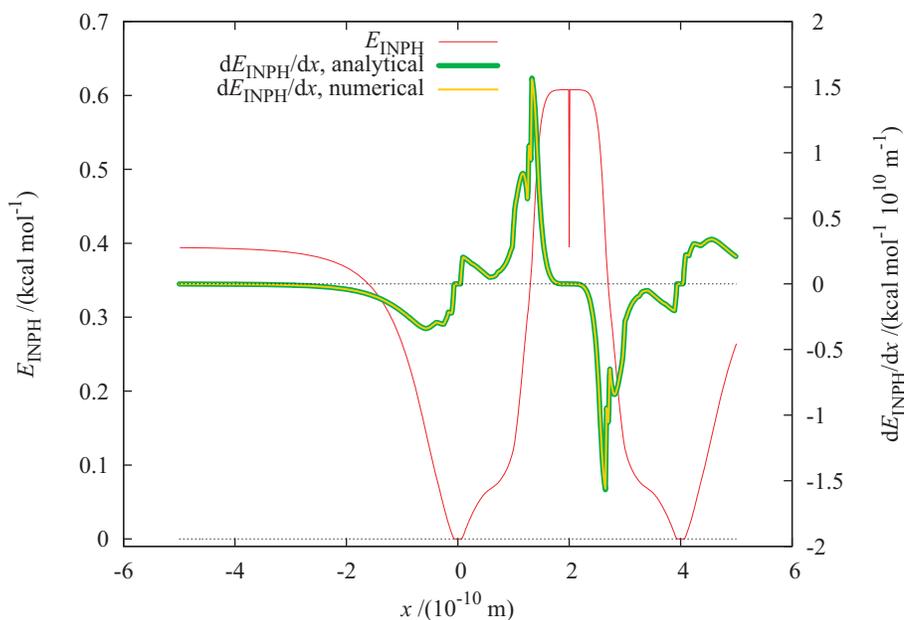
VOLUmes

```
ADD "A1" "A1" 0.57002E+01 0.57002E+00 0.48891E+01 0.48891E+00
ADD "A1" "B1" 0.30432E+00 0.30432E-01 0.12202E+01 0.12202E+00
ADD "B1" "B1" 0.37764E+01 0.37764E+00 0.26114E+01 0.26114E+00
```

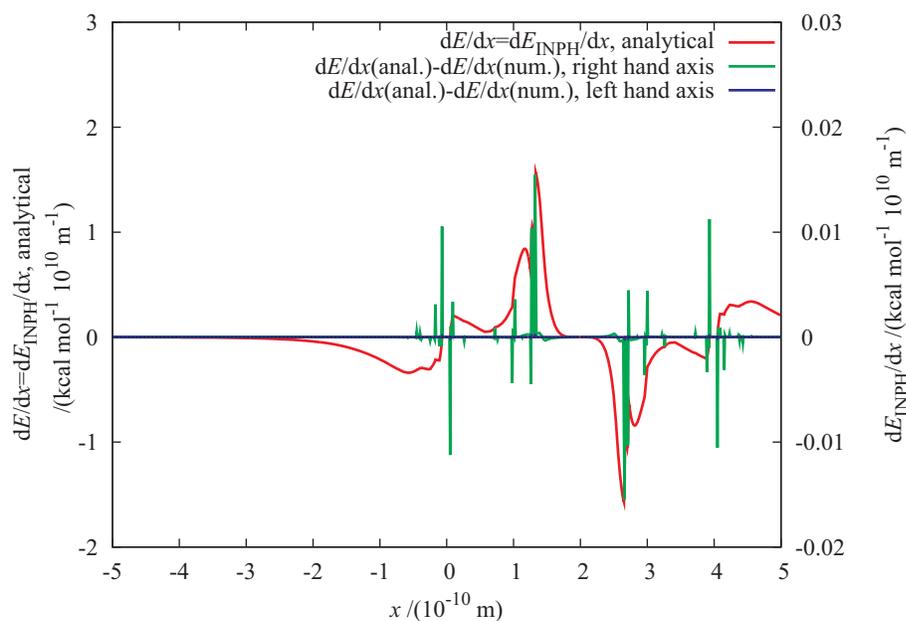
END

The used protocol can be found in the appendix (ch. 10.1.3, p. 161).

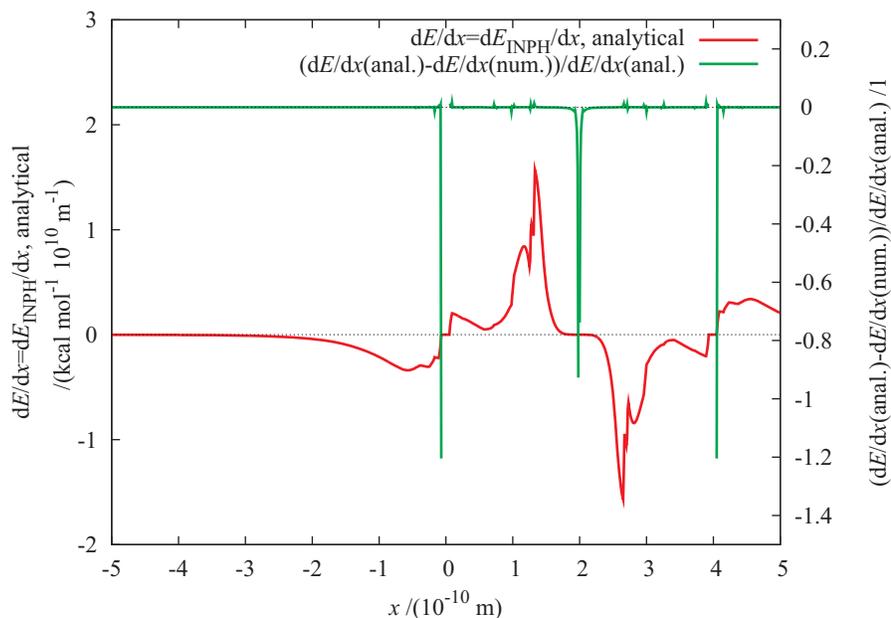
In the following figures (fig. 4.8-4.12) the INPHARMA energy  $E_{\text{INPH}}$  and gradients or functions of them along the  $x$ -coordinate of the moved atom are given.



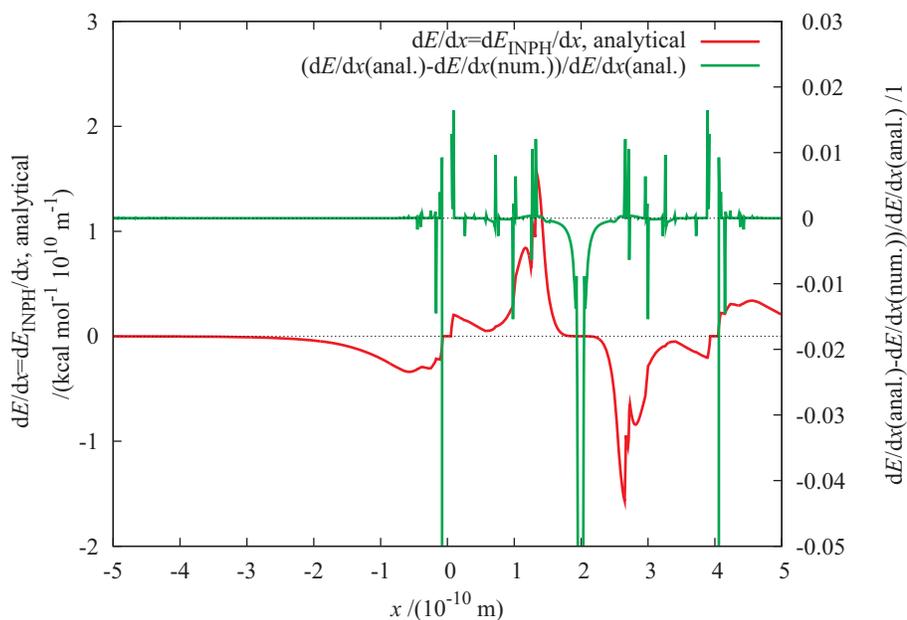
**Figure 4.8:** Plot of the INPHARMA energy  $E_{\text{INPH}}$  (red) and the numerical (yellow) and analytical (green) derivatives along the  $x$ -coordinate of the moved atom (see text). There are no apparent differences between the analytical and numerical deviates and both lines lie just on top of each other. Two mixing times and three peaks for each mixing time were used. It can be seen that the energy as a function of atom coordinate is smooth, having a continuous derivative. Nevertheless, the derivative has kinks where the second derivative of the energy suddenly changes.



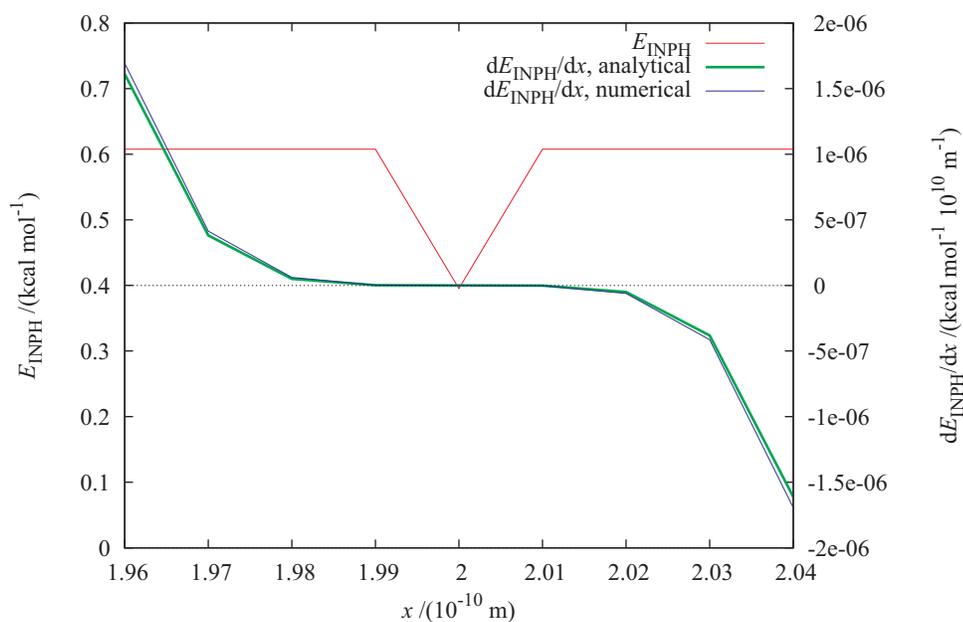
**Figure 4.9:** The analytical derivative, as in fig.4.8, and the absolute difference of analytical and numerical derivative are plotted. The blue line is the difference in the same scale as the derivative itself. Here the differences seem small compared to the range of values of the derivative itself (red line). The green line plots the absolute differences on a smaller scale making them more visible. The maximum absolute difference is about  $0.017 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  while the maximum derivative itself has a 100 times bigger value of about  $1.7 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$



**Figure 4.10:** Here the numerical derivative (green) and its relative difference to the analytical derivative (red) are given. The maximum relative difference is  $1.2 = 120\%$ . Again, the difference is biggest when the value of the derivate is very small and mainly in places where a kink in the derivative is present.



**Figure 4.11:** Here, the numerical derivative (green) and the relative difference of the analytical derivative (red) are given just as in figure 4.10. Here, the scale of the difference axis has been reduced. It can be seen that the relative difference of the numerical to the analytical derivative is normally well below  $0.02 = 2\%$ .



**Figure 4.12:** The INPHARMA energy (red), the numerical (blue) and analytical (green) derivatives are plotted in an area around  $x = 2 \text{ \AA}$ . Here a large relative difference of the derivatives of about 100% is visible. These differences, on an absolute scale, are still virtually invisible. The largest relative difference is actually due to the following absolute values  $grad_{\text{anal.}} = 0.157 \cdot 10^{-08}$  and  $grad_{\text{num.}} = 0.302 \cdot 10^{-08}$ .

It can be seen that the absolute difference between numerical and analytical gradient is small (fig. 4.8, 4.9) . The relative difference of the gradients  $(grad_{\text{analytical}} - grad_{\text{numerical}}) / grad_{\text{analytical}}$  can reach up to a factor of  $1.2 = 120\%$  (fig. 4.10, 4.11). Nevertheless, this only happens where the absolute value of the gradient is almost zero, and so are the absolute differences, and preferably in positions where the derivative is not differentiable and so has a kink (fig. 4.12). Here the numerical derivative does not give reliable estimates for the actual derivative.

The agreement of the numerical and analytical derivative is very good. The remaining differences probably come from the technical drawbacks of the numerical derivative or numerical inaccuracies. No severe differences between numerical or analytical derivative have been found, neither in the presented test case nor in other tested cases (data not shown).

From this, it is concluded that the theoretical derivation of the analytical gradient of the INPHARMA energy as well as its implementation into XPLOR-NIH are correct and can be applied to structural refinement.

### 4.1.3 Minimization

The 6-spin system was used for sampling a large, relevant part of the conformational space to visualize the energy landscape. Furthermore the minimization procedure was tested for this simple system. In the course of the minimization test each of the grid-sampled conformations, described by  $r_1$  and  $r_2$  (the two distances between the protons in the receptors) were taken as initial coordinates for minimization. During the course of minimization the distances should converge back close to their initial, reference positions as these have been used to determine the reference or pseudo-experimental peak volume.

The sampling of the energy landscape and initial minimization positions was done by independently moving atom 4 and atom 6, the bound ligands, to modify their internuclear distances in the range of  $0.4 \text{ \AA} \leq r \leq 10.3 \text{ \AA}$  for both distances  $r_1$  and  $r_2$ . The step size was  $0.01 \text{ \AA}$ . The range of distances  $r < 0.4 \text{ \AA}$  was skipped as in a few cases the minimization procedure failed and atoms got lost, meaning that their coordinate values left the available range of XPLOR, causing XPLOR to stop. This can happen if the initial gradient is small but the penalty energy is rather high. Then the Powell minimization routine of XPLOR-NIH automatically tends to pick too large step sizes, sometimes leading to divergence. This kind of divergence will not be observed in bigger, molecular systems as the distance of closest approach of non-bonded protons is inherently limited by Pauli repulsion, considered in the Van der Waals potential of XPLOR-NIH, to approximately  $r \geq 1.6 \text{ \AA}$ .<sup>39</sup>

If the protons are covalently bonded parts of a molecule distance divergence will be prohibited by the molecular cohesion.

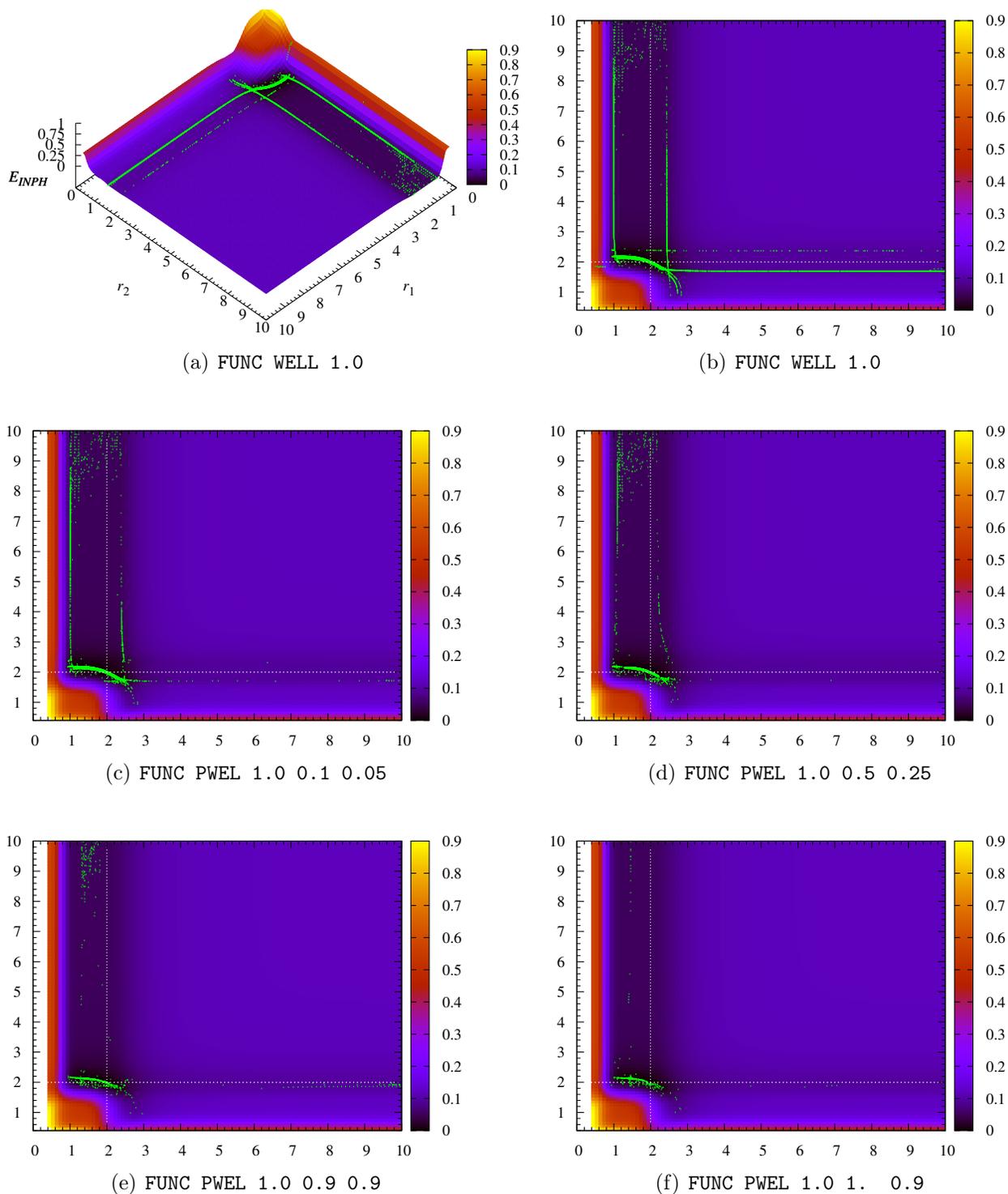
The peak volumes used for the minimizations are both diagonal peaks and the inter-ligand cross peak for a single mixing time, calculated at the reference position with  $r_1 = r_2 = 2 \text{ \AA}$ . The error bound was chosen to be 10% of each reference peak:

```
VOLUmes
  ADD "A1" "A1"          0.57002E+01  0.57002E+00
  ADD "A1" "B1"          0.30432E+00  0.30432E-01
  ADD "B1" "B1"          0.37764E+01  0.37764E+00
END
```

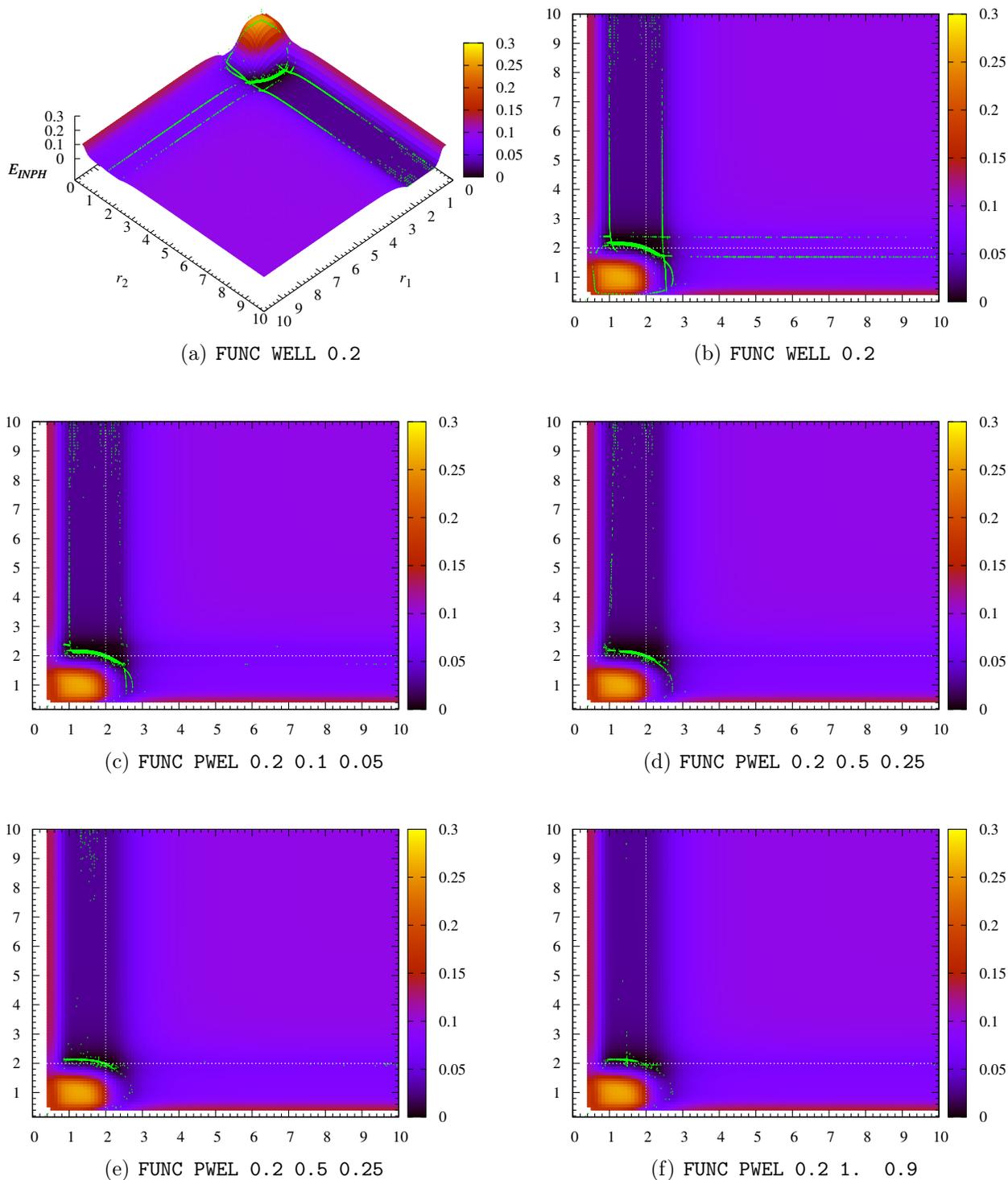
One example protocol used for minimization (with the parameters FUNC PWEL 1.0 1. 0.9) is given in the appendix (ch. 10.1.2, p. 158).

In figures fig. 4.13 and fig. 4.14 the energy landscapes for *pell* and *pwel* functions, using a range of parameters, are given. In fig. 4.13 the parameter  $f_{\text{small}}$  was 1.0 and in fig. 4.14 it was 0.2. The green dots in all figures depict the post-minimization distances inside the two complexes. The white dotted lines cross each other in the reference state of the system where each distance was 2 Å.

It can be seen that there is a tendency of the minimization to get stuck around edges of the energy landscape. This is most evident in fig. 4.13(a) and fig. 4.14(a) using the *well* function. The figures are sorted left-right, top-down getting smoother and having a reduced-size non-penalty area. Finally in figs. 4.13(f) and 4.14(f) the non-penalty area shrunk to a single point. The tendency to get stuck in edges is getting reduced by introducing a transition area with defined curvature via the *pwel* function. Lower curvatures or larger transition areas seem favorable. As long as there is a non-penalty area the post-minimization state of the system does not always converge back to the initial distances, just as expected. Then, an extended area of the conformational space is covered by the post-minimization states. When the non-penalty area is reduced to a single point most models converge closely to the initial distances.



**Figure 4.13:** The energy landscapes of the 6-spin system under the sampling of the bound ligand atoms'  $x$ -positions are shown. The parameters of the respective penalty functions are given. The green dots indicate post-minimization distances to the neighboring atoms and their post-minimization INPHARMA energy. The reference and target position was at  $r_1 = r_2 = 2 \text{ \AA}$  and is indicated by the intersection the dotted white lines.  $f_{\text{small}}$  was set to 1.0. The color bar corresponds to  $E_{\text{INPH}}$ .



**Figure 4.14:** The energy landscapes of the 6-spin system under the sampling of the bound ligand atoms'  $x$ -positions are shown. The parameters of the respective penalty functions are given. The green dots indicate post-minimization distances to the neighboring atoms and their post-minimization INPHARMA energy. The reference and target position was at  $r_1 = r_2 = 2 \text{ \AA}$  and is indicated by the intersection the dotted white lines.  $f_{\text{small}}$  was set to 0.2. The color bar corresponds to  $E_{\text{INPH}}$ .

The degree of convergence is not easily visualized using the previously described figures fig. 4.13 and 4.14 as the density of the points is visually generally underestimated. This comes from the fact that actually many points lie just on top of each other and only appear to be a single point.

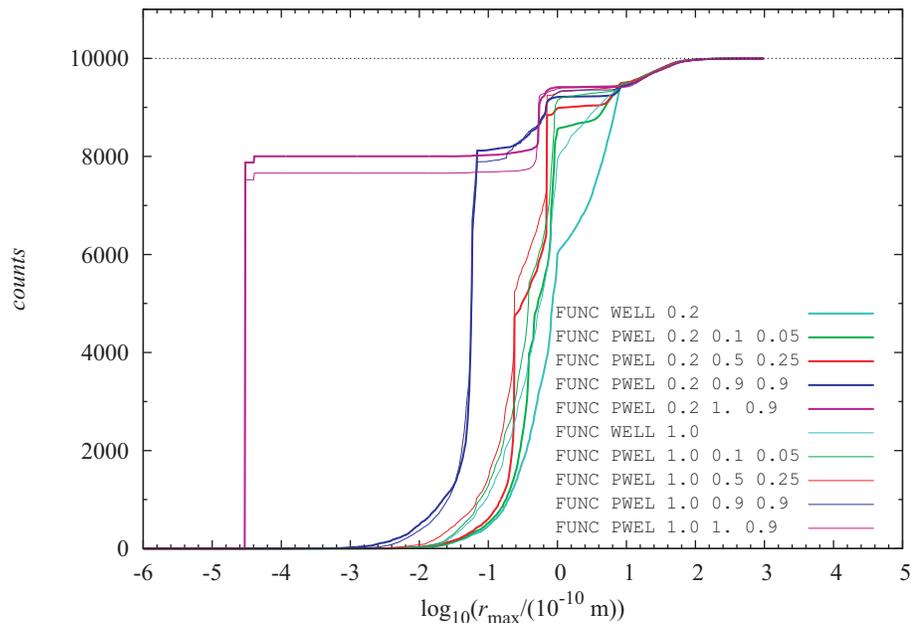
The degree of coordinate and energy convergence was quantified using the cumulative histograms in fig. 4.15 and 4.16. Various parameters for the *well* and *pwell* function were tried.  $f_{\text{small}}$  was the scaling factor for too small peaks, and  $k_i$  and  $k_o$  defined the extent of the inner and outer transition area of the *pwell* function (in multiples of the error bound) (fig. 2.1, p. 27).

Clearly, the best distance convergence is achieved using the *pwell* function with  $k_i = 1.0$  and  $k_o = 0.9$ . The difference in using  $f_{\text{small}} = 1.0$  or  $f_{\text{small}} = 0.2$  is small in this test case. For both about 80% of all models the two atom-atom distances in the complexes get minimized closer than  $10^{-4.5}$  Å to the initial distances. The smaller  $k_i$  is chosen the worse the convergence tends to be for a given upper difference limit  $r_{\text{max}}$ . The choice of  $k_i = 1.0$  is obviously the best as for all non-optimal conformations a residual force drives the system's minimization.

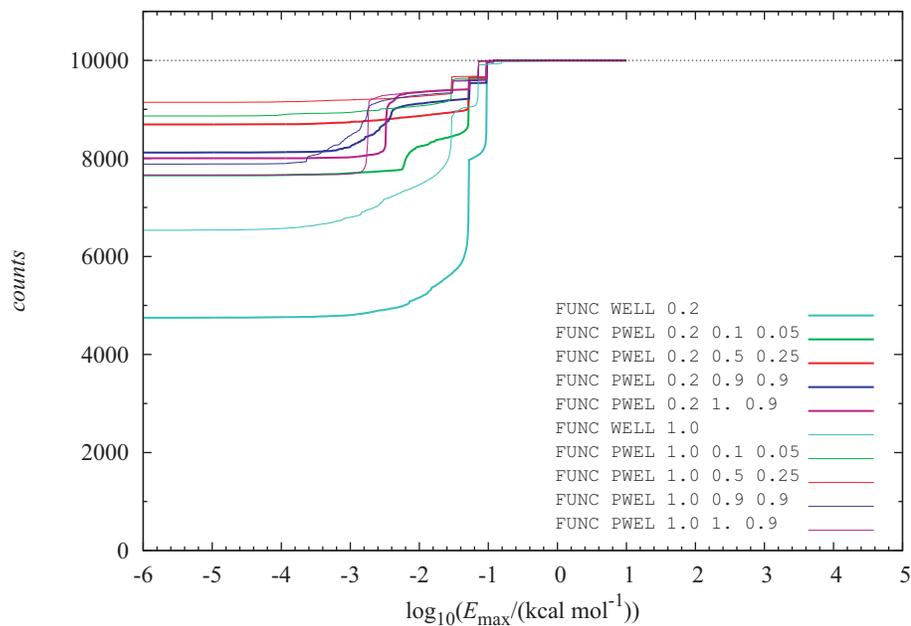
Energy convergence histograms (fig. 4.16) give a mixed impression on how well the minimization converges depending on the restraining function. The *well* functions show rather bad convergence (cyan lines in fig. 4.16) with no more than 65% of the models converging below the limit  $10^{-6}$  kcal mol<sup>-1</sup>. Actually, the energies of conformations with less than  $10^{-6}$  kcal mol<sup>-1</sup> were generally zero. One of the best functions is the *pwell* function with the parameters FUNC PWEL 1.0 0.5 0.25, having more than 90% of all models converging better than  $10^{-6}$  kcal mol<sup>-1</sup>. The function *pwell* with FUNC PWEL 1.0 1. 0.9 only performs in the average range with  $\approx 75\%$  of the models converging better than  $10^{-6}$  kcal mol<sup>-1</sup>. Nevertheless, the post-minimization INPHARMA energy histograms are probably not the best choice to judge the potential of a particular restraining function to find the correct *coordinates*. Obviously the energy convergence should get better having a larger non-penalty area because there are just more non-penalty positions to choose from. This does not mean that the resulting distances are very close to the reference distances.

The  $f_{\text{small}}$  parameter, scaling the penalty for peaks being too small, does not seem to be too relevant in this simple 6-spin example. It might become useful in the exclusive use of cross-peaks, as typical for the use of experimental INPHARMA and trNOE data. Here too small peaks would correlate with too large spin-spin distances (or unfavorable chemical exchange conditions). As the slope of the INPHARMA energy gets almost zero for too large distances some extra force, using  $f_{\text{small}} \gg 1.0$ , might help to improve convergence.

In general, for INPHARMA energy driven energy minimization, a smoother energy function, having a larger transition area and especially avoiding non-penalty areas is favorable for good distance convergence to the reference distances.



**Figure 4.15:** Plot of cumulative histograms for post-minimization distance differences  $|\Delta r_{1,2}| = |r_{1,2} - r_0|$  to the common original distance  $r_0 = 2 \text{ \AA}$ . *counts* is the number of models fulfilling the condition  $\max(|\Delta r_1|, |\Delta r_2|) \leq r_{\max}$ .



**Figure 4.16:** Plot of cumulative histograms for post-minimization INPHARMA energies  $E_{\text{INPH}}$ . For the to-be-found reference position the value of  $E_{\text{INPH}}$  is zero. *counts* is the number of models fulfilling the condition  $E_{\text{INPH}} \leq E_{\max}$ .

## 4.2 12-Spin System

To test INPHARMA minimization in a more complicated but still simple system I designed a 12-spin system where each ligand and each receptor is consisting of two atoms (fig. 4.17). The atoms are not bonded to each other and each species is put on a linear chain with a 2 Å spacing.

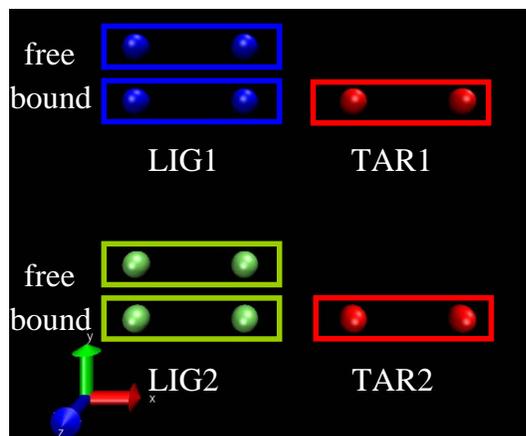


Figure 4.17: 12-spin-system

### 4.2.1 Minimization

This minimization test was just used to check if the minimization works in principle and is not thoroughly analyzed.

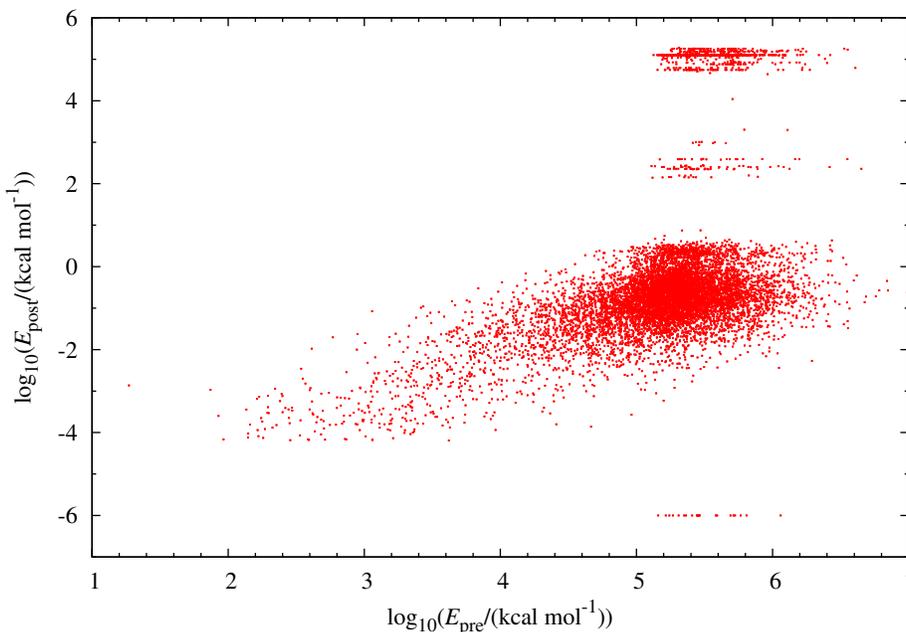
Visualizing the INPHARMA energy landscape and testing the minimization in the 12-spin system is generally not as simple as in the 6-spin system as now there is no simple possibility to map or even display the high dimensional INPHARMA energy landscape. Altogether, there are 16 relevant distances and the systematic sampling of a large set of possible conformations is hardy possible. More importantly the conformational space will possess a large number of local INPHARMA energy minima, distributed over the whole conformational space. This makes it difficult to define a meaningful metric to compare the very different pre- and post-minimization conformations.

A protocol was used to change the reference coordinates randomly and to minimize the system afterward, using the reference peak volumes of the reference conformation as pseudo experimental input. The noise level was increased in steps of 0.01 Å to a maximum of 1 Å. 100 models were created for each noise level, and so 10,000 models altogether. All the 10 possi-

ble cross and diagonal peaks of the reference conformation have been chosen as pseudo experimental input using two mixing times. 10% of each peak volume was chosen as the pseudo experimental error. The force constant was  $k_{\text{INPH}} = 1,000,000$  and the penalty function's parameter were `FUNC PWEL 0.2 1 1`.

The full protocol (ch. 10.2.1, p. 166) and the initial coordinates (p. 170) can be found in the appendix.

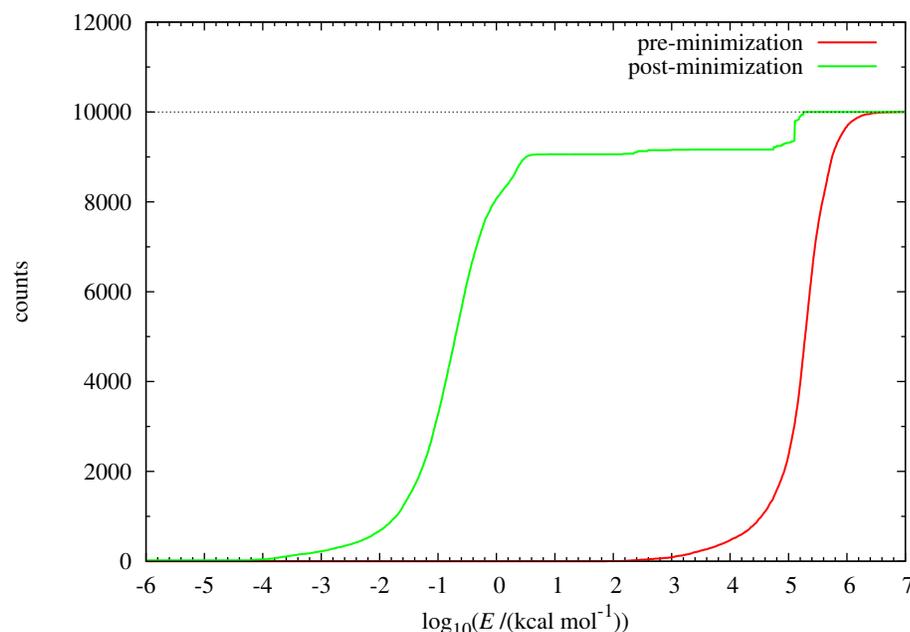
No systematic test for structural similarity of the post-minimization coordinates to the reference coordinates was performed. Visual inspection of a few post-minimization coordinates shows that the coordinate shift can be rather high, many Å, for higher noise levels. For lower noise levels, much below 1 Å, the structural difference can be very small.



**Figure 4.18:** Correlation plot of the pre- and post-minimization energies of all 10,000 minimizations using the 12-spin system with incremented structural noise.

In fig. 4.18 the post-minimization energy is plotted over the pre-minimization energy. There are several clusters of energy pairs but in general the post-minimization energy drops significantly, between 1 to 12 orders of magnitude, below the pre-minimization energy. Although there is a dominating cluster showing correlation between pre- and post-minimization energy

there seem to be clusters where the minimization gets stuck. This is likely due to XPLOR's Powell minimization method stopping the minimization in unfavorable cases (see ch. 4.1.3). This behavior has not been analyzed in detail and I turned my attention to more realistic molecular systems, actually showing fine convergence. Simulated annealing protocols will generally not get stuck at edges of the energy landscape. Using them before final minimization, reaching a point close to the minimum, should allow the minimization to converge in most cases, as less transitions from a penalty to a non-penalty situations will remain.



**Figure 4.19:** Cumulative histograms of the pre- and post-minimization energies of all 10,000 minimization using the 12-spin system with incremented structural noise.

After all, the cumulative histogram of pre- and post-minimization energies in fig. 4.19 clearly shows the energy improvement due to minimization. While the 90th percentile of the pre-minimization energies is about  $10^5 \text{ kcal mol}^{-1}$ , the post-minimization energies have a 90th percentile of about  $10^{0.5} = 3.2 \text{ kcal mol}^{-1}$ .

It has been demonstrated that the INPHARMA minimization in XPLOR-NIH works reliably using systems of several atoms per species, several mixing times and peak volumes. It can minimize structural deviations along all three Cartesian axes.

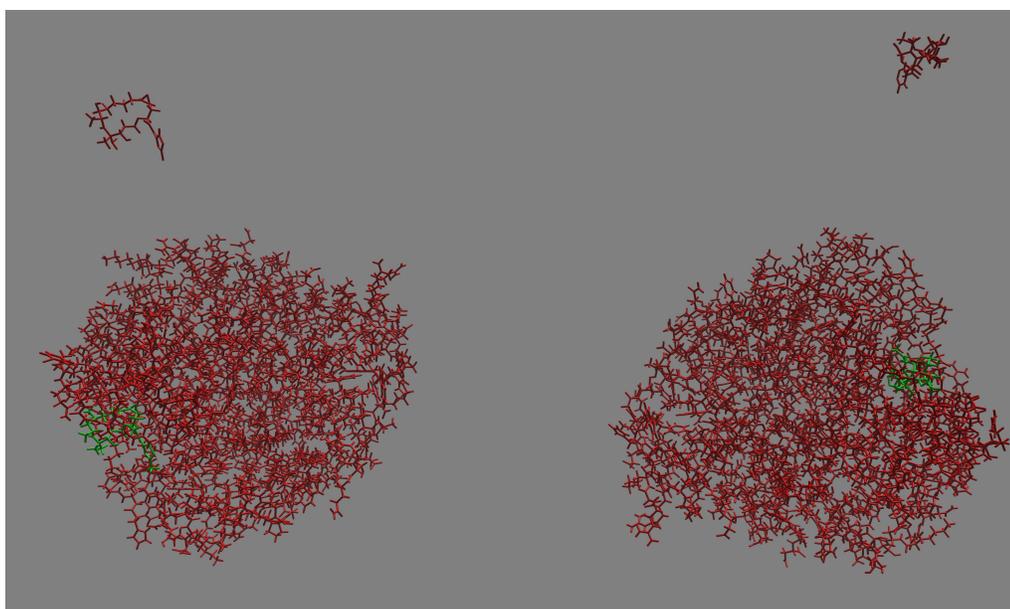
The XPLOR-NIH Powell minimization has problems with edges in the energy landscape and the use of the power well function *pwel* can avoid them.

It is important to avoid non-penalty areas by setting  $k_i = 1$  and the transition area should be rather smooth, achieved e.g. by setting  $k_o \approx 1.0$ .

The average CPU time for running one out of two times 36911 simulations was 15 CPU min. 4 cores of the type Xeon 5160 (Woodcrest, Dual Core, 3 GHz, 4 MB L2 Cache,  $\approx 12$  GFLOPS/core) have been used on a cluster of dual CPU (quad core) shared memory machines. The average CPU load was about 80%, which is likely close to the scaling factor of parallelization. XPLOR-NIH was compiled with the Intel compiler version 11.0. Highest optimization and auto-parallelization had been activated to create a parallelized binary for the non-INPHARMA parts of XPLOR-NIH. The INPHARMA part is mainly parallelized via the used Intel-MKL library 10.2.1.017.

### 4.3 Tubulin - Epothilone A - Baccatin III

The actual application of INPHARMA driven minimization would be large macromolecular complexes with ligands of biological or pharmaceutical interest. Here INPHARMA is tested in the context of the tubulin-epothilone A-baccatin III system. It was the first experimental system to demonstrate the experimental and computational aspects of INPHARMA.<sup>10,12</sup> The previous application of INPHARMA led to a model of the complex of tubulin with the anti cancer agent epothilone A in agreement with the very most of the known SAR (structure activity relationship) and mutational data.



**Figure 4.20:** Overview over the two complexes (red: receptor; green: ligands) and the two free ligands (red, on the top). The species keep separated from each other during minimization and simulated annealing. Anyway, the species only interacted with each other by means of INPHARMA forces as all other inter-species interactions had been switched off explicitly.

In a first step it is demonstrated that INPHARMA driven minimization can be used to improve macromolecular models to better fulfill the experimental data (ch. 4.3.1). Minimization will normally not stop in the global minimum of the energy landscape but just in the next local minimum.

In a second step a simple simulated annealing molecular dynamics protocol is used to achieve further structural improvements, beyond the next local minimum. With simulated annealing protocols it is principally possible to find the global minimum of the energy landscape<sup>25,26</sup> and so to determine

the model which can best explain the experimental data.

The tubulin - epothilone A - baccatin III system was chosen for initial tests of the applicability of INPHARMA restraint driven minimization and simulated annealing protocols to the refinement of macromolecular complexes. Fig. 4.20 gives an impression of the molecular system. One complex pair as used for INPHARMA minimization and simulated annealing is shown.

In principle about 27 million different complex pairs could be created from the existing docking models. It was still practicable to calculate all the INPHARMA scores for all 27 million combinations.<sup>11,12</sup> For one pair a simulation took 0.7 CPUs. Minimization or even simulated annealing of all pairs, each single one taking 15 CPUmin (p. 4.2.1) or 3 CPUh (p. 4.3.2) respectively, is not an option. If the previously best INPHARMA-scored models are somehow valid docking and INPHARMA models it should be possible to refine the better ones of them, either by minimization or simulated annealing, to get better models without missing any good candidates.

In contrast to the calibration factors (2.16, p. 20) and weights (2.17) defined in the chapter 'Theory', here a uniform weight of "1" and a common calibration factor for all spectra is chosen.

The calibration factor used here is

$$k^s = \frac{\sum_{m,\alpha,\beta} \mathbf{A}_{m,\alpha,\beta}^{\text{cal}}}{\sum_{m,\alpha,\beta} \mathbf{A}_{m,\alpha,\beta}^{\text{exp}}} \quad (4.2)$$

where the sum runs over all spectra  $m$  and all well defined (p. 20) cross-peaks. This common calibration factor is justified by the fact that the used spectra were calibrated manually for comparability of their absolute peak volumes.

### 4.3.1 Minimization

The initial models for minimization were taken from the HADDOCK<sup>40</sup> docking models, generated to find the binding modes of epothilone A and baccatin III to tubulin, using INPHARMA.<sup>11,12</sup> These models base on an electron crystallography model (PDB id: 1jff<sup>41</sup>) and had been created, during my diploma thesis,<sup>11</sup> in a protocol allowing increasing flexibility of the complex where finally explicit water refinement had been performed. As the bound conformations of the ligands had been measured or validated using trNOE experiments<sup>42</sup> the bound conformations had been restrained using artificial dihedral angle restraints during docking. In initial steps of the model generation NOE restraints had been used to guide the ligand into the binding pocket. These restraints got virtually ineffective when the lig-

and had approached any atom of the defined binding pocket closer than 4 Å. Experimental NOE restraints had also been included for epothilone. As the conformation of the thiazole ring of epothilone A had only been ambiguously defined through CCR (cross-correlated relaxation) measurements. This is why the thiazole ring had not been fixed but was allowed to switch between the two conformations.

All the described restraints, used to generate the HADDOCK docking models, have been included in the here presented minimization protocol. Explicit water was not included. Only atoms of residues having at least one atom closer than 14 Å to any of the ligands have been allowed to move. All other atoms were fixed. The non-bonded interactions were only active inside the chemical species. That means that the free ligands could neither feel each other nor the complexes, nor the other way around. All species could in principle occupy the same space without interacting, except through INPHARMA. As the experimental sample was kept in deuterated water all exchangeable protons have been excluded of the list of active protons. They will not participate in the magnetization transfer process during the INPHARMA simulation. The simulations were carried out using the penalty function parameters `FUNC PWEL 0.2 1 1` (see. ch 2.2.1, p. 25, fig. 2.1). An extended non-penalty area was avoided. To make use of the fast diagonalization procedure the parameter `EXPS=0.99999` ( $k_{esum} = 0.99999$ , p. 31) was used. The force constant was either 100, 1000 or 1,000,000. Five mixing times with 51 peak volumes each have been considered. The RMSD between the pre- and post-minimization model was calculated on the basis of all non-fixed atoms.

For generating the HADDOCK docking models the program CNS has been used. Though XPLOR-NIH is very similar to CNS there are remaining differences in the force-field parameter files. Due to them and the lack of explicit water in the following minimization runs it is expected that the starting models will not lie in a local minimum of all the non-INPHARMA energy terms.

The minimization with force constants  $k_{\text{INPH}} = 100$  and  $k_{\text{INPH}} = 1000$  have been performed for the best  $\approx 37,000$  of all docking model pairs, judged based on the initial regression coefficient  $R$  of the models.  $R$  is corresponding to a fit of a straight line  $f(x) = a + bx$  to the correlation plot of experimental and calculated data.<sup>11</sup>

The protocol for  $k_{\text{INPH}} = 1,000,000$  can be found in the appendix (ch. 10.3.1, p. 171). Only the value of  $k_{\text{INPH}}$  needs to be adapted to obtain the other protocols. The used pair of complex models corresponds to the one published previously, explaining virtually all SAR and mutational data.<sup>12</sup>

To get an impression on how the energies evolve during minimization the

figures 4.21, 4.22; figs. 4.23, 4.24; and figs. 4.25, 4.26 were prepared corresponding to minimizations with  $k_{\text{INPH}} = 100$ ,  $k_{\text{INPH}} = 1000$ , and  $k_{\text{INPH}} = 1,000,000$ , respectively.

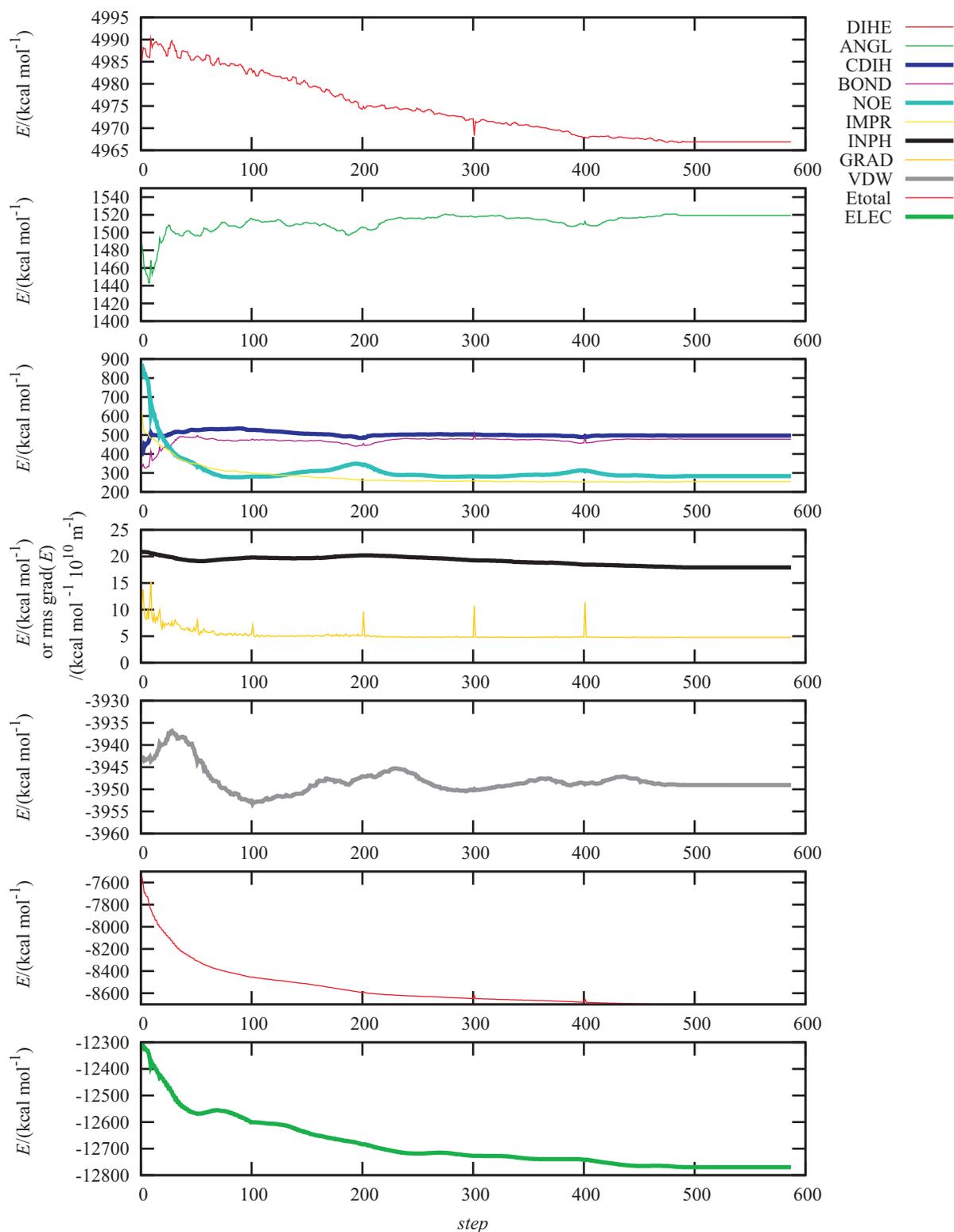
The corresponding correlation plots of experimental and calculated data are given in figures 4.27-4.30, p. 74.

The figures 4.21 and 4.22 show the evolution of all the energies during minimization for  $k_{\text{INPH}} = 100$ . Fig. 4.21 has a broken energy axis to show the changes of individual energy terms in greater detail. Fig. 4.22 shows the same data mapped on a unbroken energy axis to put the different energies in scale to each other. It is evident that here the minimization is mainly driven by non-INPHARMA energies. The INPHARMA line (bold black, in the middle of both plots) does only drop from 20.9 to 17.9 by 3 kcal mol<sup>-1</sup>. The main energy contributions to the total energy drop of  $\Delta E_{\text{total}} = 1180$  kcal mol<sup>-1</sup> come from the NOE ( $\Delta E_{\text{NOE}} = 586$  kcal mol<sup>-1</sup>) and the electrostatics ( $\Delta E_{\text{ELEC}} = 464$  kcal mol<sup>-1</sup>) terms. As expected the total energy drops during minimization and so does the total rms gradient of all forces. The spikes in some energy functions are caused by restarting the energy minimization after every 100 steps. Then the step size is not yet adapted and a trial with total energy increase can be observed. Though these attempts to search for better energy position are plotted they will be rejected by the Powell minimization procedure of XPLOR-NIH. The structural change for the given example was quantified to an RMSD of 0.117 Å on all atoms allowed to move. Dividing the post-minimization INPHARMA energy by  $k_{\text{INPH}} = 100$  yields the scaled INPHARMA energy of  $E_{\text{INPH,scaled}} = 0.179$  kcal mol<sup>-1</sup>, which is almost, discarding the unit, the value of the post-minimization R-factor. The correlation plot can be found in fig. 4.28.

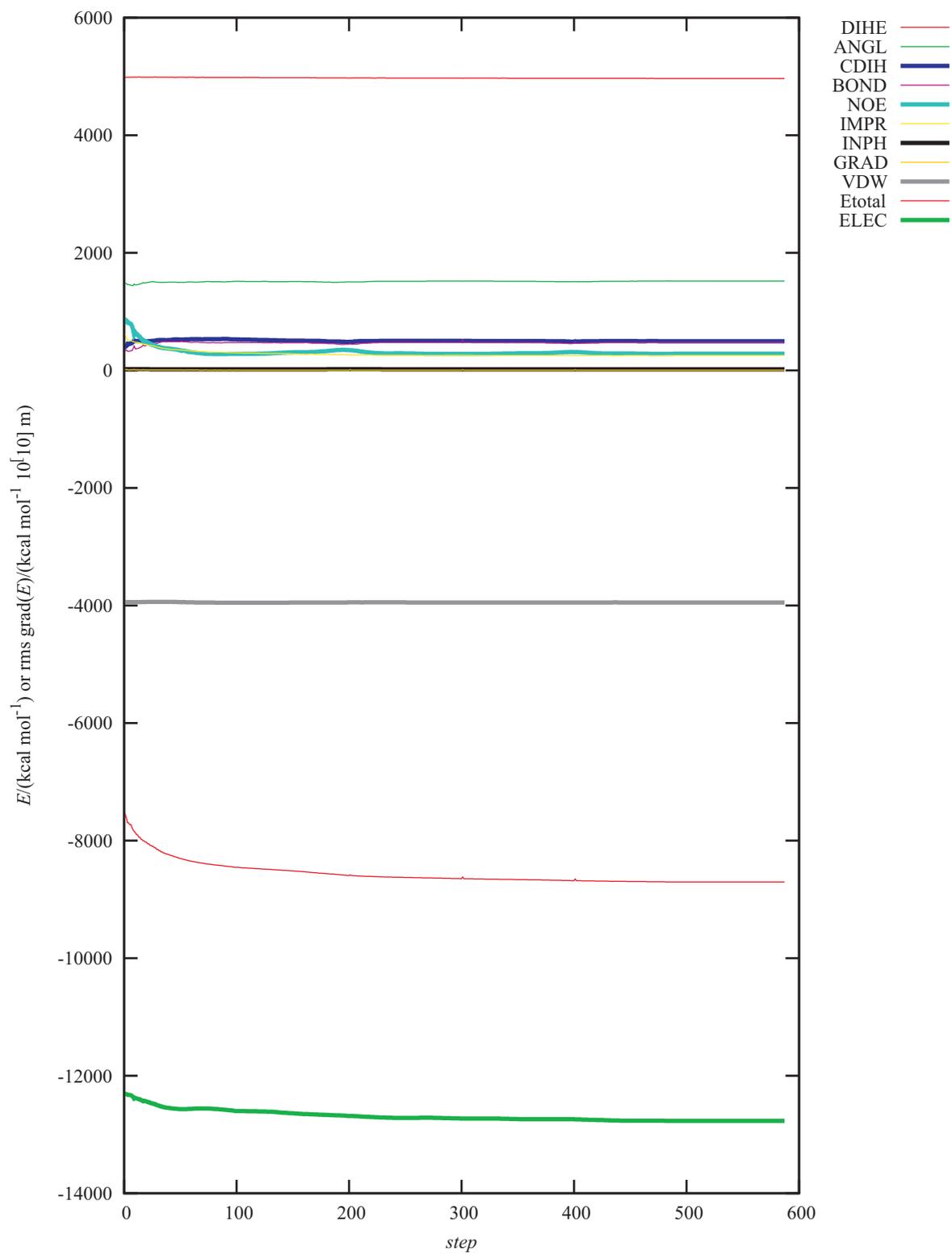
The figures 4.23, 4.24 show, just as the previously explained figures, the evolution of all the energies during minimization, but for  $k_{\text{INPH}} = 1000$ . Also here it is evident that the minimization is mainly driven by non-INPHARMA energies. The INPHARMA energy, again in the middle of both plots, does only drop from 182.0 to 88.4 by 93.6 kcal mol<sup>-1</sup>. The main energy contributions to the total energy drop of  $\Delta E_{\text{total}} = 1283$  kcal mol<sup>-1</sup> come from the NOE ( $\Delta E_{\text{NOE}} = 585$  kcal mol<sup>-1</sup>) and the electrostatics ( $\Delta E_{\text{ELEC}} = 460$  kcal mol<sup>-1</sup>) terms. The structural change for the given example was quantified to an RMSD of 0.147 Å on all atoms allowed to move. The scaled post-minimization INPHARMA energy, using  $k_{\text{INPH}} = 1000$ , is  $E_{\text{INPH,scaled}} = 0.0936$  kcal mol<sup>-1</sup>. Compared the the minimization with  $k_{\text{INPH}} = 100$ , with  $E_{\text{INPH,scaled}} = 0.179$  kcal mol<sup>-1</sup>, the INPHARMA restraints get better explained (fig. 4.29), as expected from their higher weight.

The figures 4.25 and 4.26 show, just as the previously explained figures, the evolution of all the energies during minimization, but here for  $k_{\text{INPH}} = 1,000,000$ . Here, the minimization is mainly driven by the INPHARMA energy. The INPHARMA energy, now on top of both plots, does drop from 208656 to 22036 by 186620 kcal mol<sup>-1</sup>. This is the main energy contribution to the total energy drop of  $\Delta E_{\text{total}} = 178914$  kcal mol<sup>-1</sup>. The NOE energy drop is  $\Delta E_{\text{NOE}} = 553$  kcal mol<sup>-1</sup> and the electrostatics one  $\Delta E_{\text{ELEC}} = 207$  kcal mol<sup>-1</sup>. The structural change for the given example was quantified to an RMSD of 0.157 Å on all atoms allowed to move. The scaled post-minimization INPHARMA energy, using  $k_{\text{INPH}} = 1,000,000$ , is  $E_{\text{INPH,scaled}} = 0.022$  kcal mol<sup>-1</sup>. Compared to the minimization with  $k_{\text{INPH}} = 100$  or  $k_{\text{INPH}} = 1000$  the INPHARMA restraints get even better fulfilled (fig. 4.30).

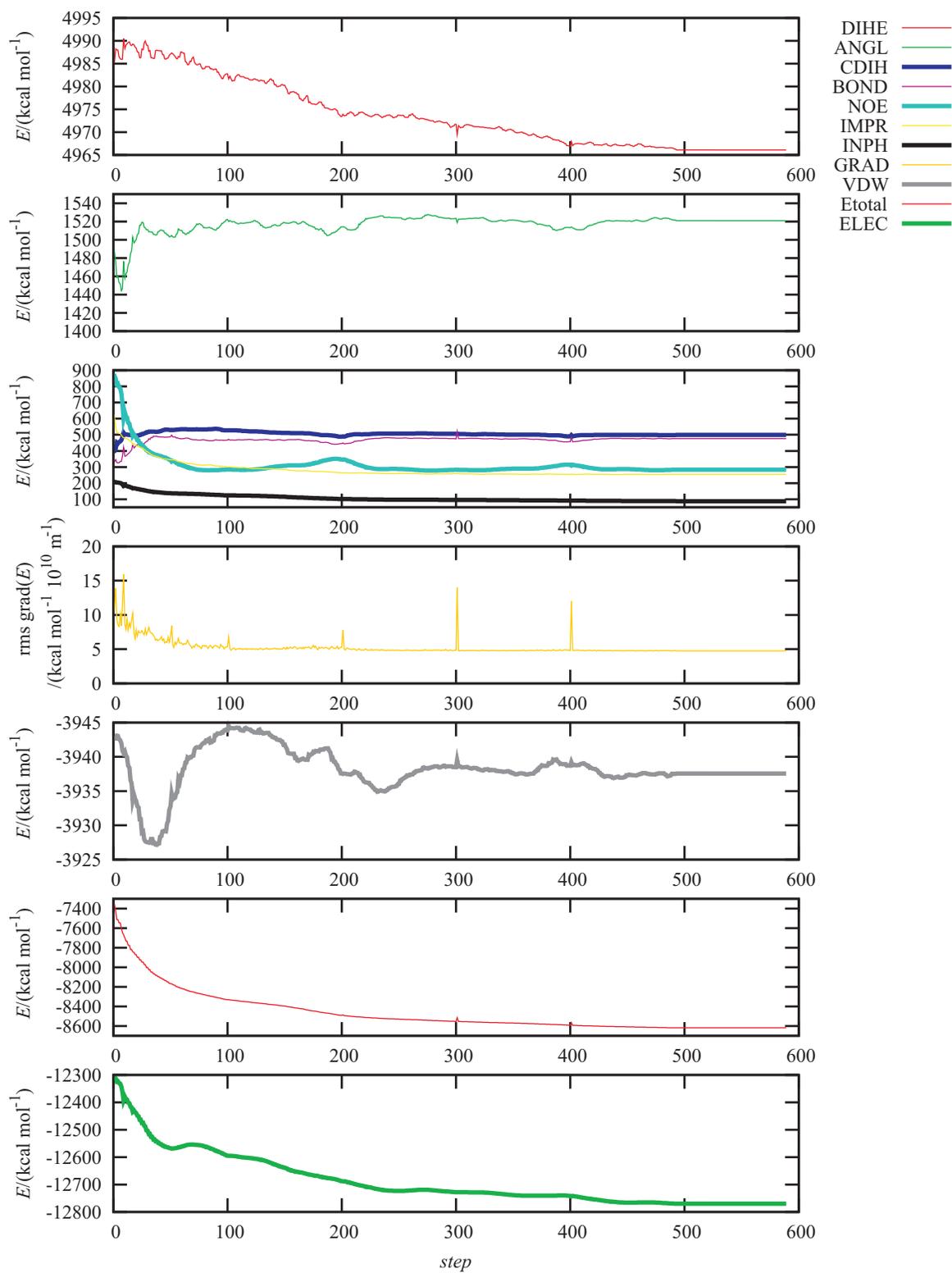
Allover it can be seen that in all the described situations with  $k_{\text{INPH}} = 100$ ,  $k_{\text{INPH}} = 1000$  or  $k_{\text{INPH}} = 1,000,000$  the total energy is minimized effectively. Although the energy contributions of non-INPHARMA terms is dominating the minimization for the first two cases the INPHARMA restraints get minimized as well. In the last case with  $k_{\text{INPH}} = 1,000,000$  the INPHARMA energy is clearly dominating. Although it is minimized heavily the other energy contributions don't seem to suffer too much. They don't get by far as much increased as the INPHARMA energy is dropping. Looking at the RMSD changes, of 0.117, 0.147 and 0.157 Å for the three weights, it becomes evident that only very small structural changes are responsible for the large changes in fit quality:  $E_{\text{INPH,scaled}} = 0.179$ , 0.0936, or 0.022 kcal mol<sup>-1</sup>, respectively. Though, the average RMSD of the selected, movable atoms is not necessarily a good measure for the structural change close to the ligands, as the closest atoms are most contributing to the INPHARMA energy. An impression on the maximum structural changes can be gained from the figures 4.31 and 4.32.



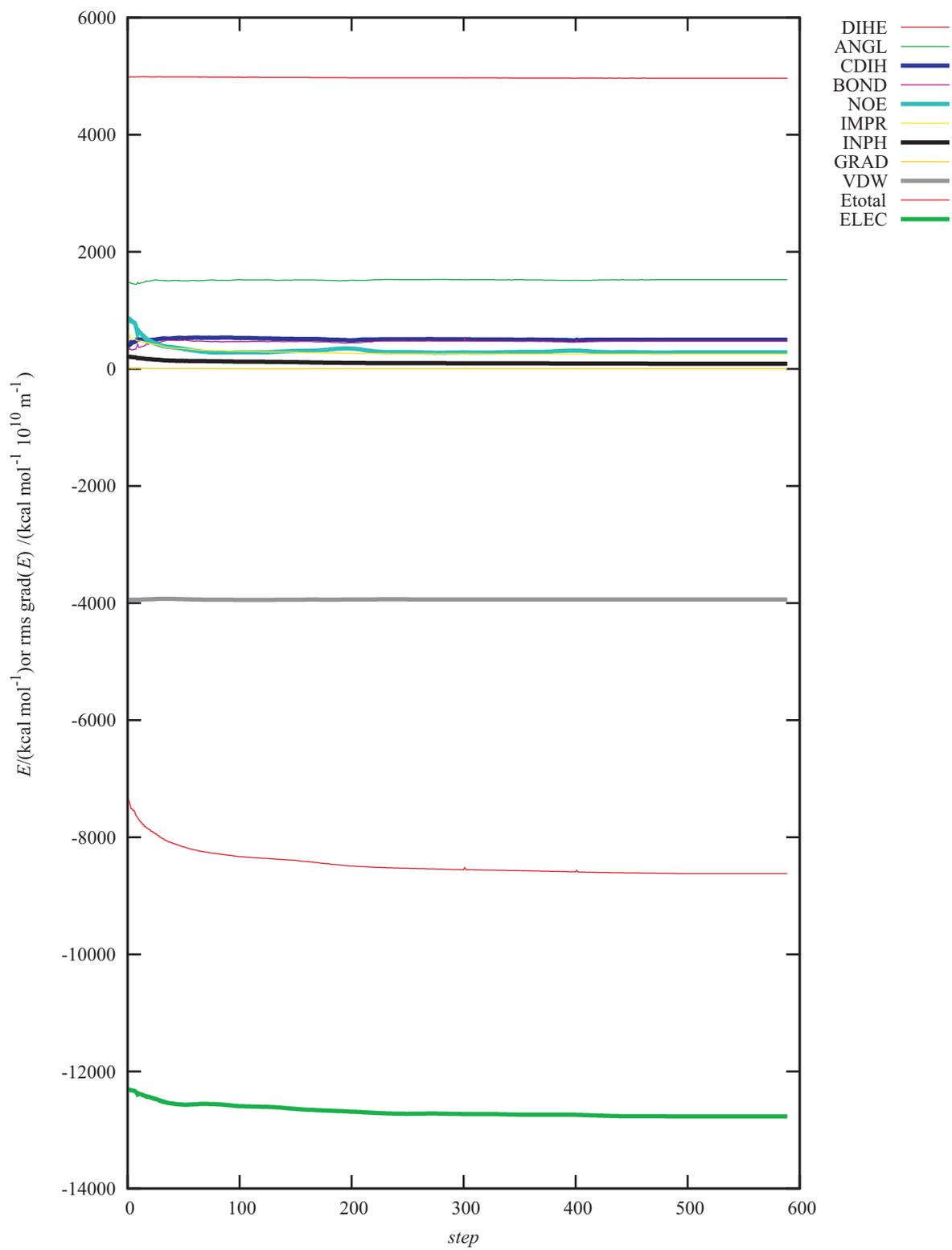
**Figure 4.21:** Evolution of all active energy terms during INPHARMA minimization with  $k_{\text{INPH}} = 100$ . The  $E$ -axis is interrupted for a better visibility of small changes. The shortcuts used in the key are: DIHE, dihedral energy; ANGL, angular e.; CDIH, dihedral restraint e.; BOND, bond stretching e.; NOE, distance constraint e.; IMPR, improper e. (planarity and chirality); INPH, INPHARMA e.; GRAD, root mean square gradient; VDW, Van der Waals e.; Etotal, total e.; ELEC, electrostatic e.



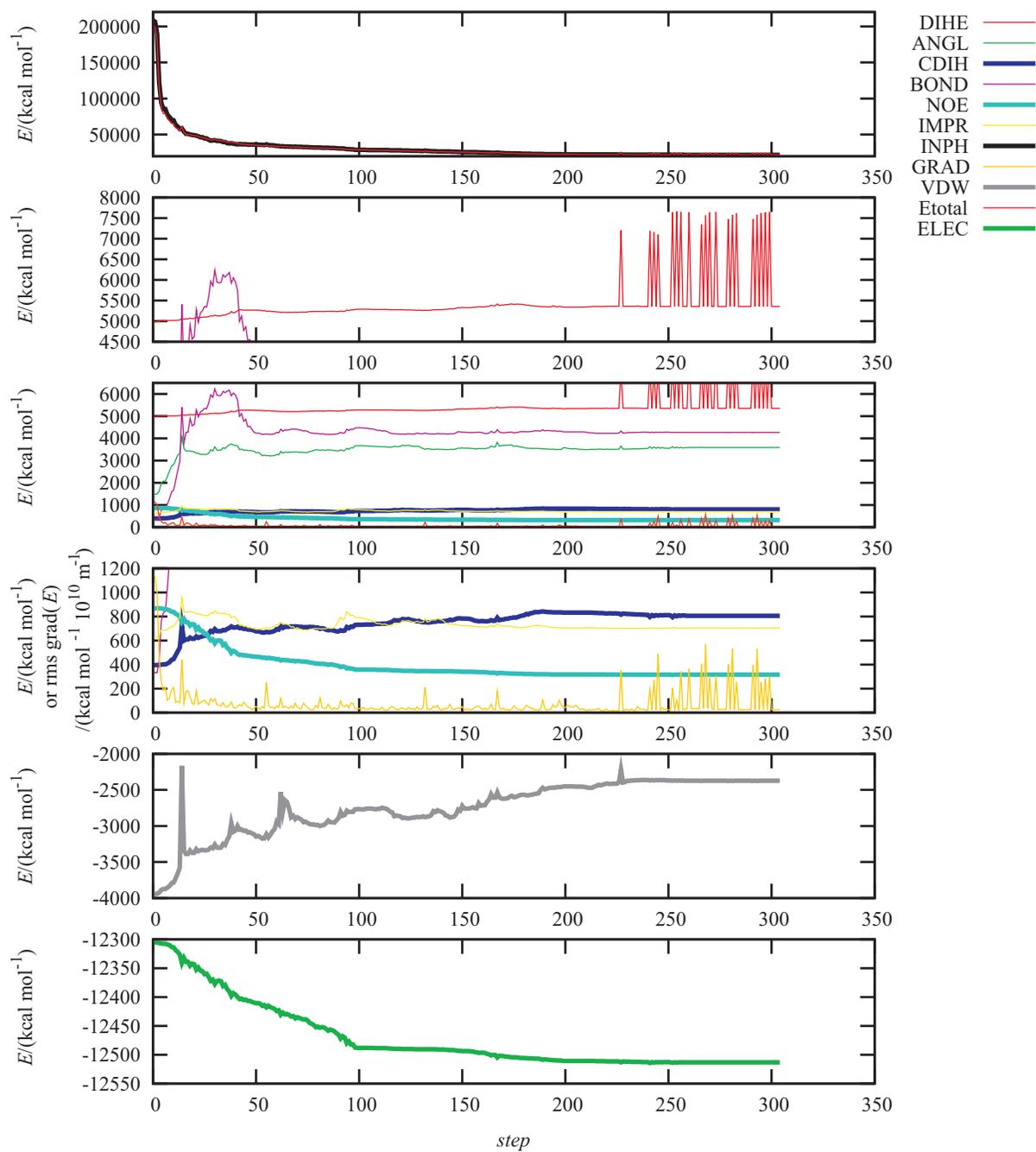
**Figure 4.22:** Evolution of all active energy terms during INPHARMA minimization with  $k_{\text{INPH}} = 100$ .



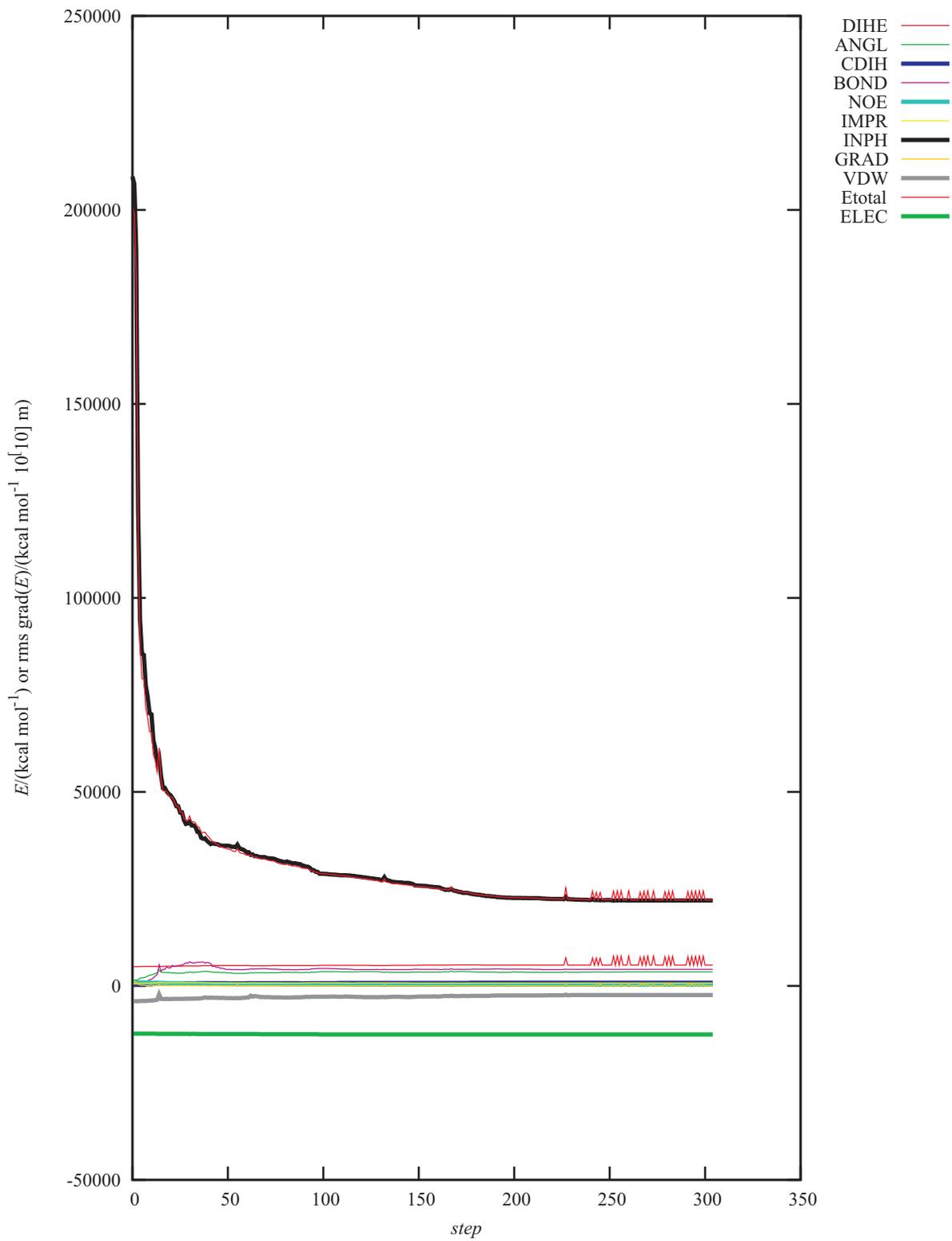
**Figure 4.23:** Evolution of all active energy terms during INPHARMA minimization with  $k_{\text{INPH}} = 1000$ . The  $E$ -axis is interrupted for a better visibility of small changes.



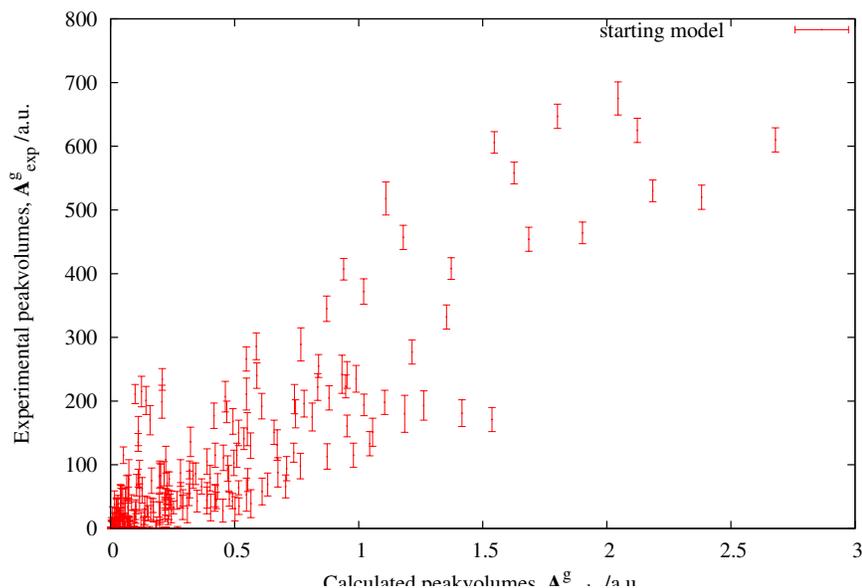
**Figure 4.24:** Evolution of all active energy terms during INPHARMA minimization with  $k_{\text{INPH}} = 1000$ .



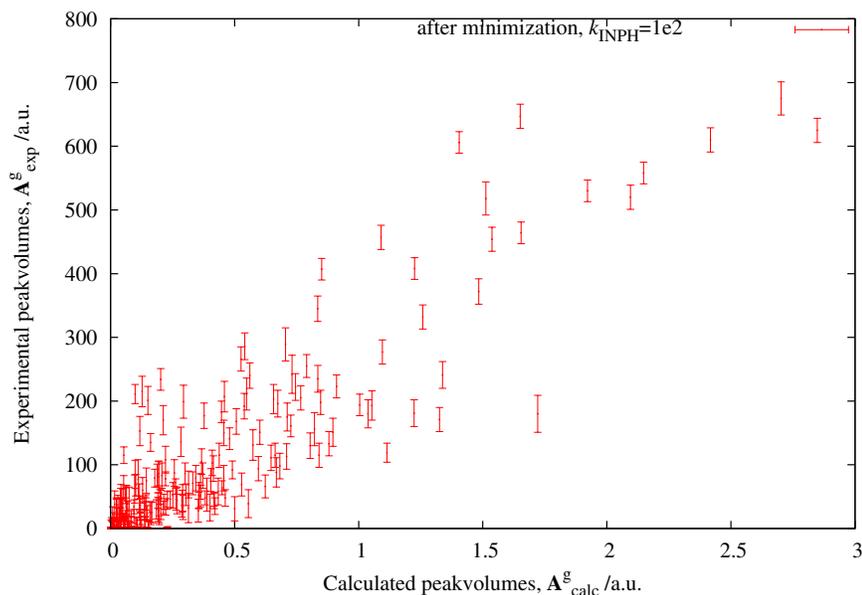
**Figure 4.25:** Evolution of all active energy terms during INPHARMA minimization with  $k_{\text{INPH}} = 1,000,000$ . The  $E$ -axis is interrupted for a better visibility of small changes.



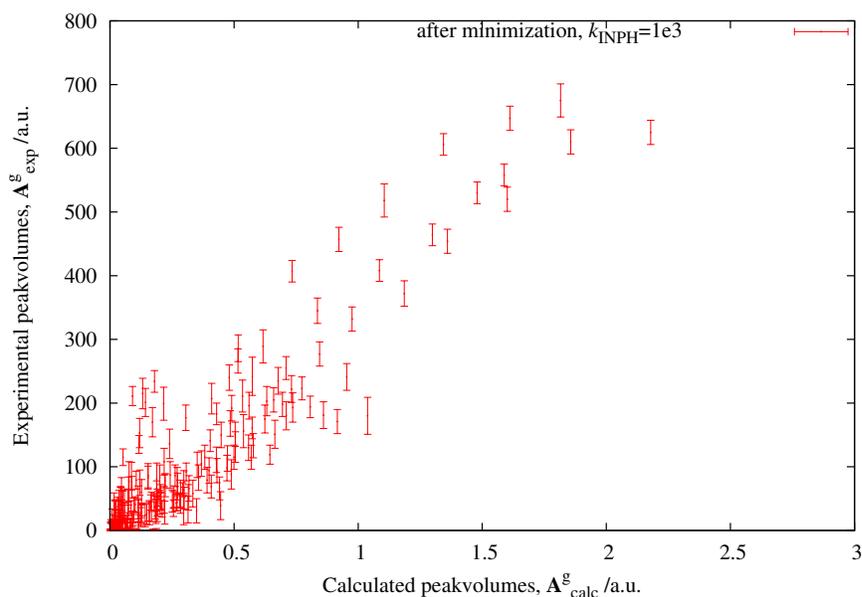
**Figure 4.26:** Evolution of all active energy terms during INPHARMA minimization with  $k_{\text{INPH}} = 1,000,000$ .



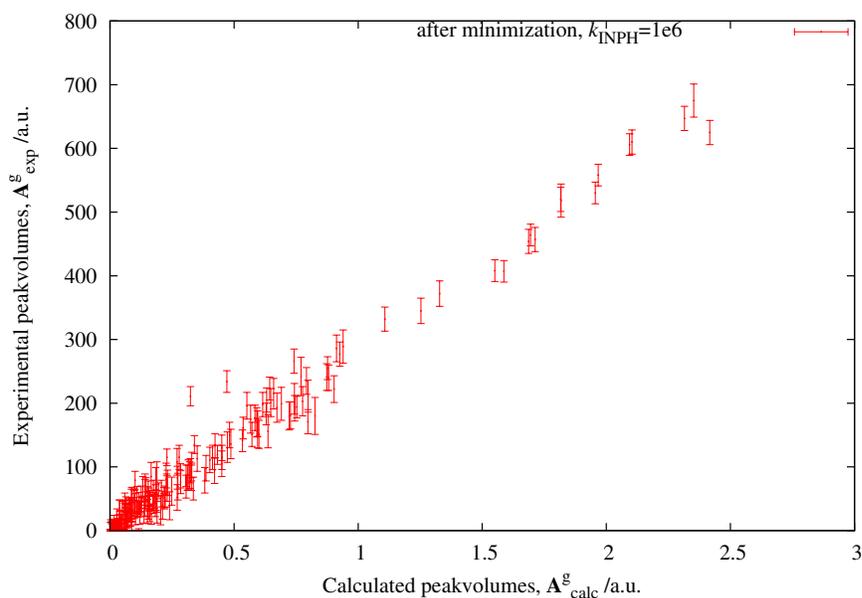
**Figure 4.27:** Correlation plot of experimental and calculated grouped peak volumes before minimization of the macromolecular test system (fig. 4.31, 4.32). The scaled INPHARMA energy is  $0.209 \text{ kcal mol}^{-1}$ . The complex models correspond to the pair published previously,<sup>12</sup> thus the correlation is already good.



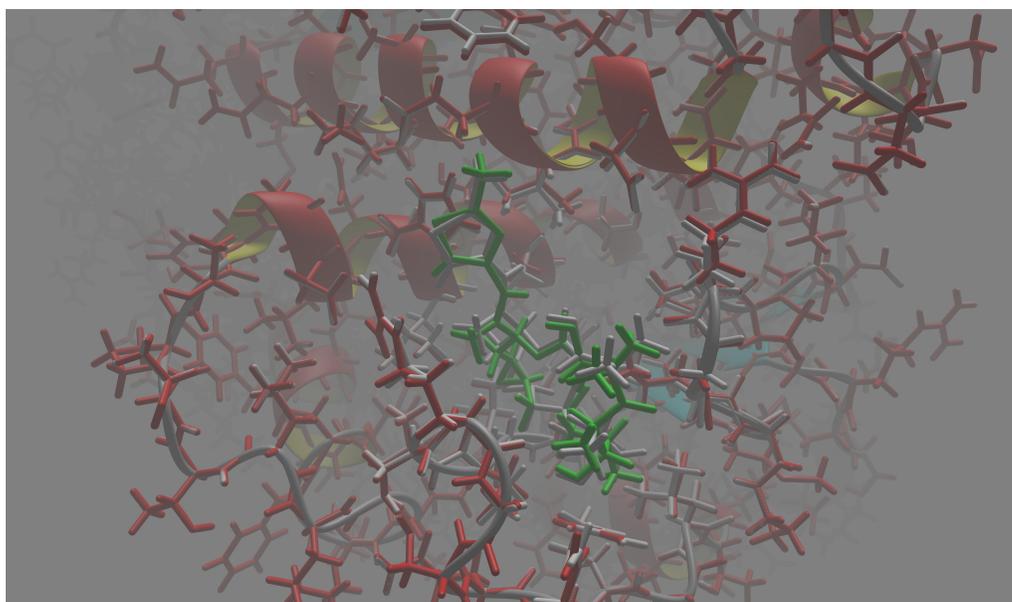
**Figure 4.28:** Correlation plot of experimental and calculated grouped peak volumes after minimization with  $k_{\text{INPH}} = 100$ . The scaled INPHARMA energy is  $0.179 \text{ kcal mol}^{-1}$ . Using this low INPHARMA weight only minor correlation improvements are visible.



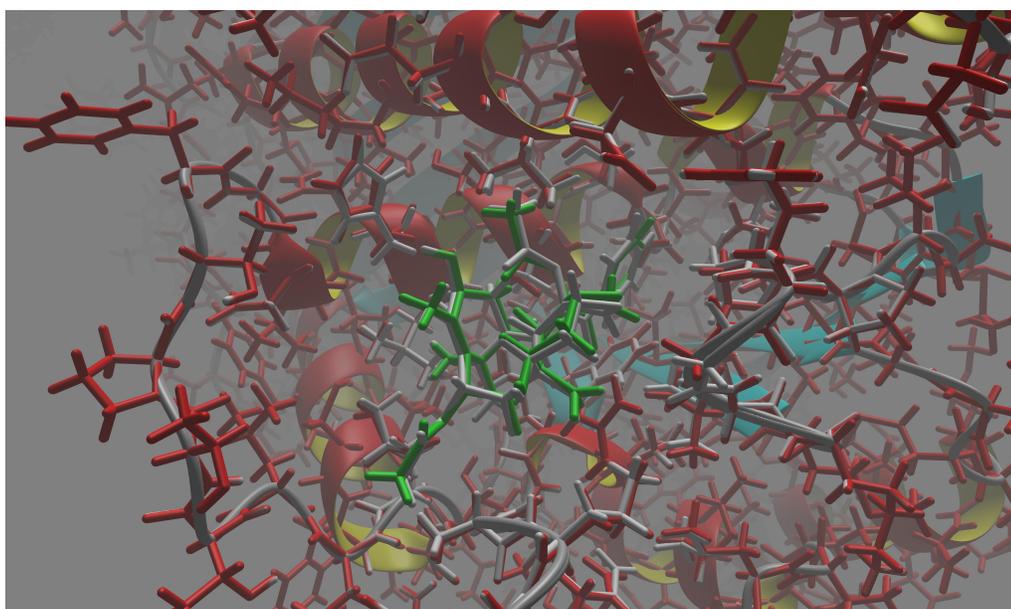
**Figure 4.29:** Correlation plot of experimental and calculated grouped peak volumes after minimization with  $k_{\text{INPH}} = 1000$ . The scaled INPHARMA energy is  $0.088 \text{ kcal mol}^{-1}$ .



**Figure 4.30:** Correlation plot of experimental and calculated grouped peak volumes after minimization with  $k_{\text{INPH}} = 1,000,000$ . The scaled INPHARMA energy is  $0.022 \text{ kcal mol}^{-1}$ .



**Figure 4.31:** Close up view of the epothilone A models before (white) and after (colored) minimization using  $k_{\text{INPH}} = 1,000,000$ . The total RMSD, including the complex with baccatin, was  $0.157 \text{ \AA}$ . The shown complex models correspond to the ones published previously, explaining virtually all SAR, mutational data and INPHARMA data.<sup>12</sup>



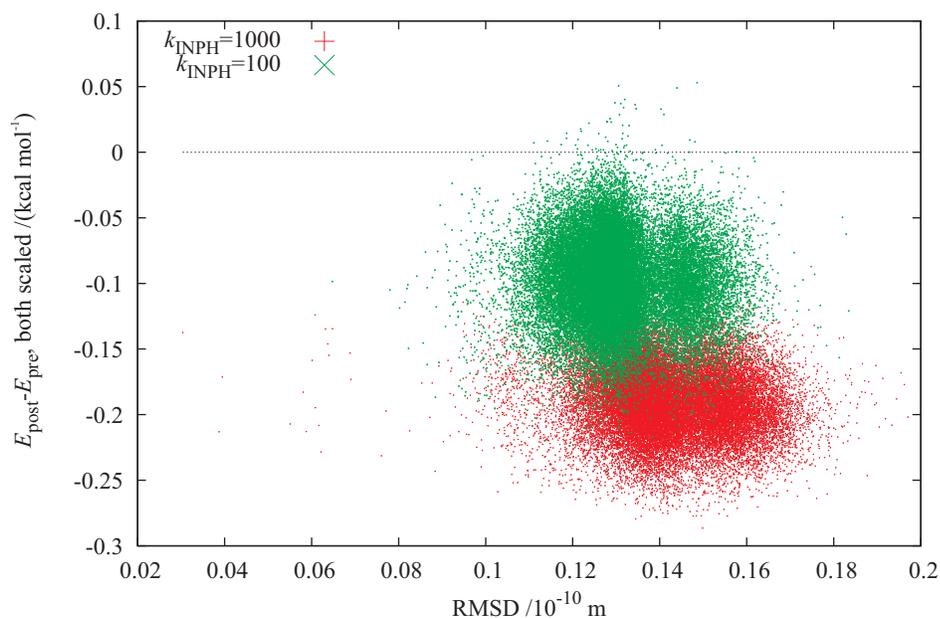
**Figure 4.32:** Close up view of the baccatin III models before (white) and after (colored) minimization using  $k_{\text{INPH}} = 1,000,000$ . The total RMSD, including the complex with epothilone, was  $0.157 \text{ \AA}$ . The shown complex models correspond to the ones published previously, explaining virtually all SAR, mutational data and INPHARMA data.<sup>12</sup>

The next step was to check, in a first test case, if INPHARMA minimization can be used to easily score the binding modes and to select the correct binding mode. To achieve this, the best 37,000 models of the INPHARMA-scored docking modes have been minimized according to the protocol described in this chapter. In fig. 4.33 the structural change over minimization is plotted against the change in scaled INPHARMA energy, with  $E_{\text{INPH,scaled}} = E_{\text{INPH}}/k_{\text{INPH}}$ . For a given INPHARMA weight the structural change is not strongly correlated with the improvement of INPHARMA energy (This view might change if less atoms, just the ones being close to the ligands, are considered). Increasing the weight will increase both, the structural change as well as the INPHARMA energy improvement. The scaled INPHARMA energy is improved twice as well with the stronger than with the lower weight, by either -0.19 or -0.098 kcal mol<sup>-1</sup>, respectively. The structural change to accomplish this extra improvement seems minute as the average RMSD only increases from 0.130 Å to 0.142 Å.

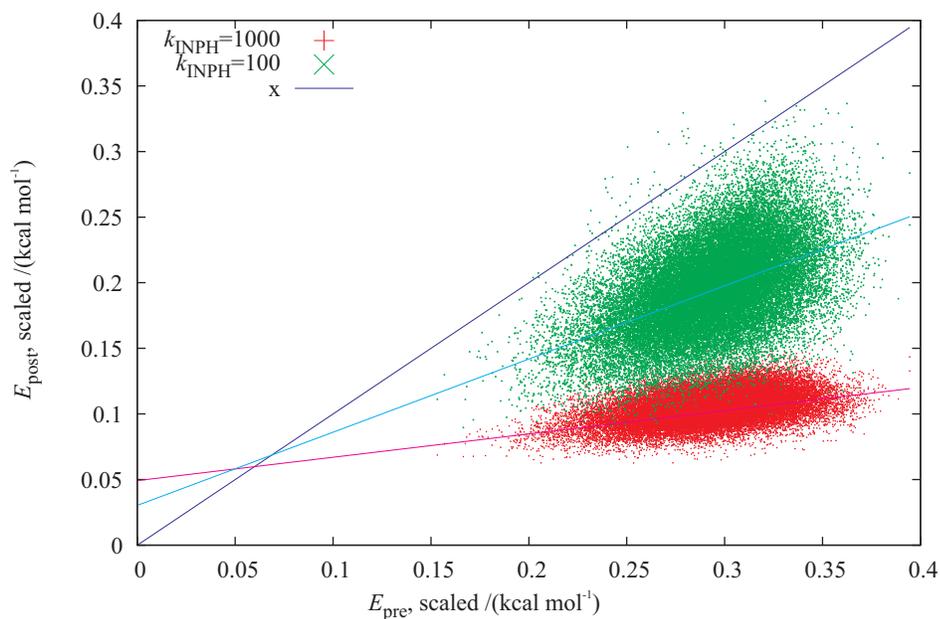
There is a well visible correlation between the pre-minimization and post-minimization energies (fig. 4.34). Nevertheless, models which just perform sub-average before minimization, having a scaled INPHARMA energy in between 0.2 and 0.3, can get top scoring models after minimization. This means: To be sure to really find all potentially top scoring models after minimization, it is necessary to consider of a lot of initial models and to perform the minimization on them. The plot also reveals that the INPHARMA energy can actually get *worse* during minimization with  $k_{\text{INPH}} = 100$ . This can be rationalized by the fact that here the INPHARMA energy will have just a small contribution to the total energy and is not dominating the minimization process (as indicated by the energy changes discussed above).

The scaled INPHARMA energies,  $E_{\text{INPH,scaled}} = E_{\text{INPH}}/k_{\text{INPH}}$ , have been sorted to rank the models. The energies are plotted over rank in fig. 4.35. The better fit of experimental and calculated data is clearly visibly for the results with higher INPHARMA-weight.

The question arises if all the models having a low energy will have the same structure and if only these would be able to explain the INPHARMA data correctly. To check this hypothesis, in a first attempt, the 10 best post-minimization models have been selected for both INPHARMA weights and were visually inspected (figs. 4.36-4.37). Here, the first 10 best ligands, using the pre-minimization models for simplicity, are superimposed on the best receptor. The relative orientation of the ligands is ambiguously emphasized by marking one of the exterior atoms with a green sphere. In the figures less than 10 different ligand orientations might be distinguishable as some of the initial models are actually identical. For both weights and for both ligands a large structural variation in binding mode can be observed,



**Figure 4.33:** Plot of the scaled INPHARMA energy improvements over the structural changes, expressed by the RMSD of the movable atoms, over the course of the used minimization protocol. Green:  $k_{\text{INPH}} = 100$ ; red:  $k_{\text{INPH}} = 1000$ .

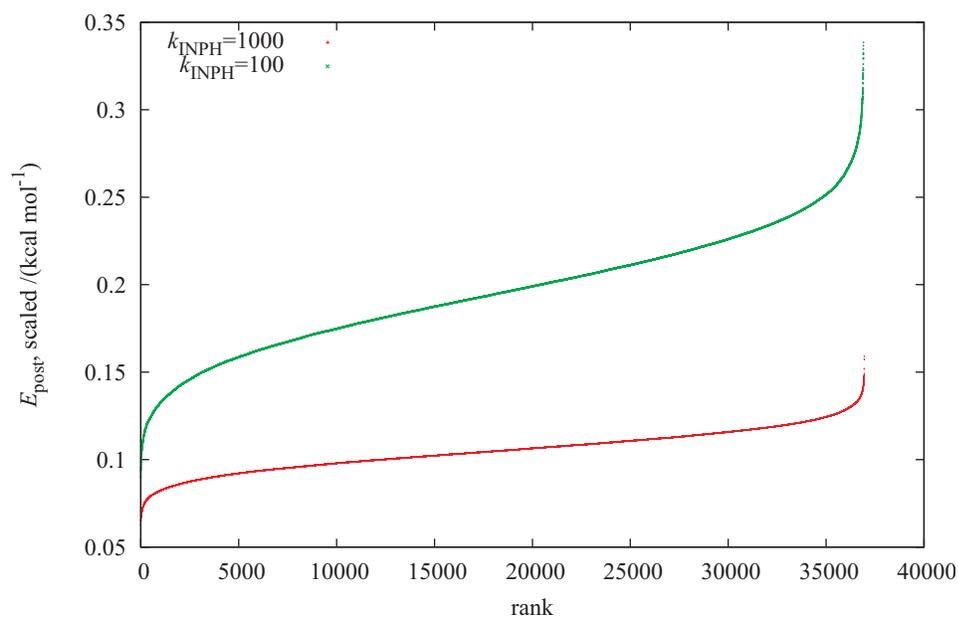


**Figure 4.34:** The scaled post-minimization INPHARMA energy is plotted over the scaled pre-minimization INPHARMA energy. Although a correlation is visible, even average pre-minimization models can perform very well after minimization. Green:  $k_{\text{INPH}} = 100$ ; red:  $k_{\text{INPH}} = 1000$ .

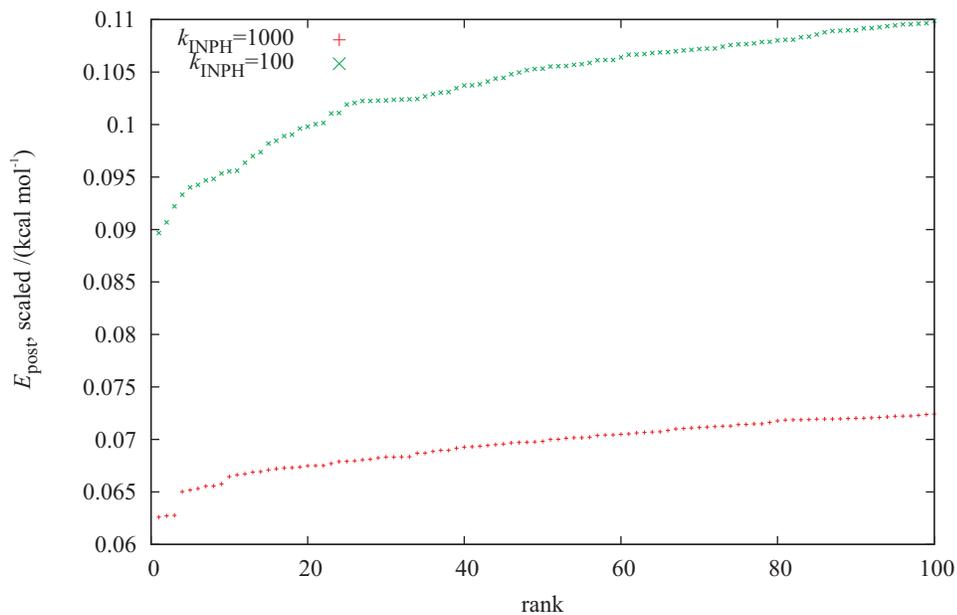
even in the 10 best scoring pairs. All of these models have an excellent fit to the INPHARMA data. A selection of the best binding mode, in this first attempt, based exclusively on INPHARMA data and not considering any other energy term, is not possible.

The inclusion of further experimental data, as the remaining diagonal and off-diagonal non-INPHARMA NOE peaks, and by this refining also the bound conformations of the models, should be considered to improve the degree of convergence to a single model. Furthermore, the use of other physical models to simulate the spectra, considering the dynamics of the system, should be considered.

To test the possibility to get even better refinement of the INPHARMA spectra, and possibly getting a better selectivity as well, a simple simulated annealing protocol has been implemented and tested as described in the next chapter.

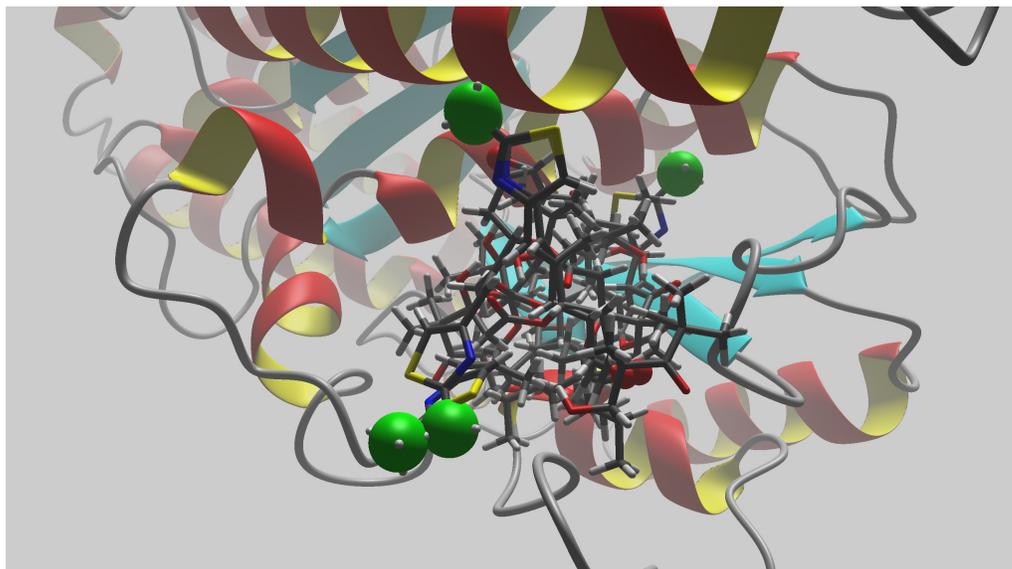


(a) All complex pairs.

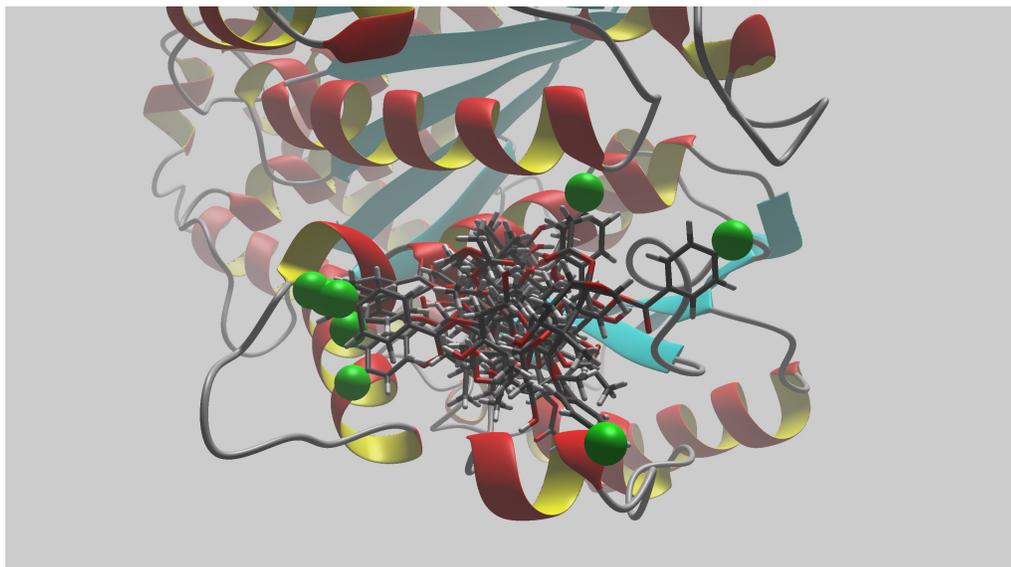


(b) Best 100 complex pairs.

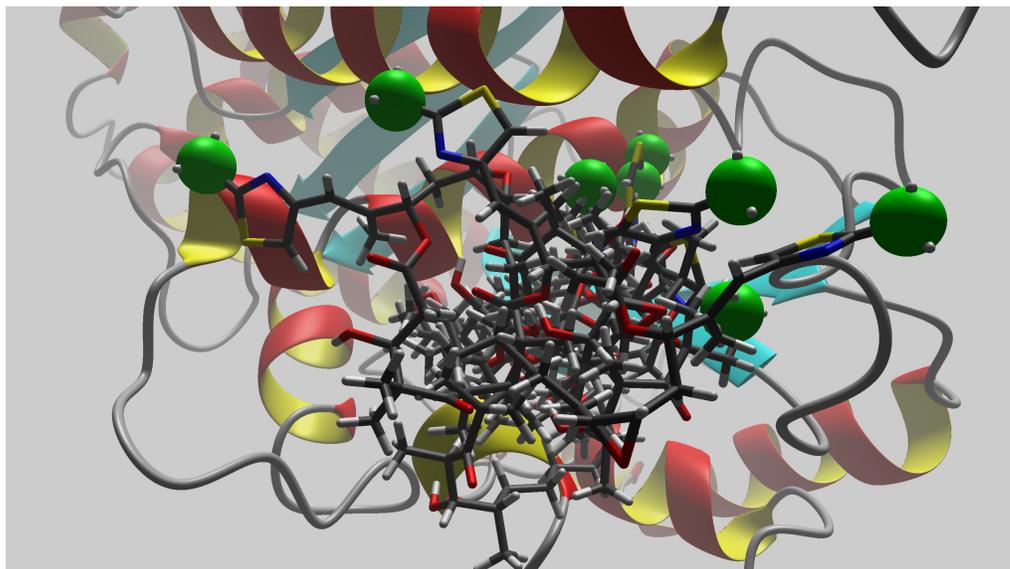
**Figure 4.35:** Scaled INPHARMA energies  $E_{\text{INPH, scaled}} = E_{\text{INPH}}/k_{\text{INPH}}$  after minimization over the model rank. For  $k_{\text{INPH}} = 1000$  the improved fit of calculated and experimental data can be deduced from the lower scaled INPHARMA energy values. Green:  $k_{\text{INPH}} = 100$ ; red:  $k_{\text{INPH}} = 1000$ .



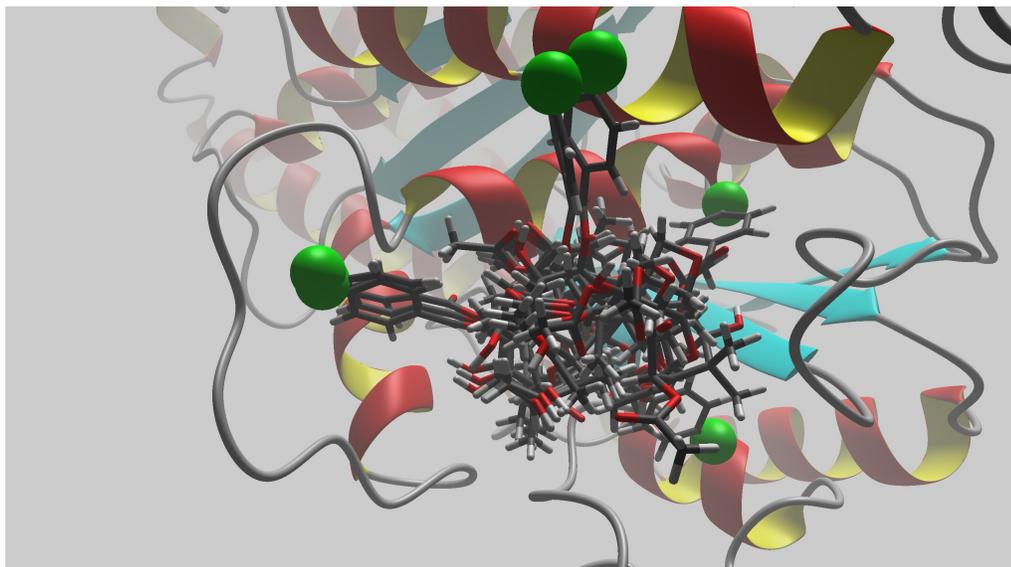
**Figure 4.36:** Representative models of the 10 best post-minimization epothilone A - tubulin complexes. The minimization was run with  $k_{\text{INPH}} = 100$ . The models were sorted by scaled INPHARMA energy. The shown, representative models, are actually the pre-minimization models (superimposed on the heavy backbone atoms of the receptor). This is why less than 10 models are actually distinguishable in this picture. The relative ligand orientation is ambiguously highlighted by marking one exterior atom with a green sphere. Only the best-scoring receptor is shown.



**Figure 4.37:** Representative models of the 10 best post-minimization baccatin III - tubulin complexes. The minimization was run with  $k_{\text{INPH}} = 100$ . The models were sorted by scaled INPHARMA energy. The shown, representative models, are actually the pre-minimization models (superimposed on the heavy backbone atoms of the receptor). This is why less than 10 models are actually distinguishable in this picture. The relative ligand orientation is ambiguously highlighted by marking one exterior atom with a green sphere. Only the best-scoring receptor is shown.



**Figure 4.38:** Representative models of the 10 best post-minimization epothilone A - tubulin complexes. The minimization was run with  $k_{\text{INPH}} = 1000$ . The models were sorted by scaled INPHARMA energy. The shown, representative models, are actually the pre-minimization models (superimposed on the heavy backbone atoms of the receptor). This is why less than 10 models are actually distinguishable in this picture. The relative ligand orientation is ambiguously highlighted by marking one exterior atom with a green sphere. Only the best-scoring receptor is shown.



**Figure 4.39:** Representative models of the 10 best post-minimization baccatin III - tubulin complexes. The minimization was run with  $k_{\text{INPH}} = 1000$ . The models were sorted by scaled INPHARMA energy. The shown, representative models, are actually the pre-minimization models (superimposed on the heavy backbone atoms of the receptor). This is why less than 10 models are actually distinguishable in this picture. The relative ligand orientation is ambiguously highlighted by marking one exterior atom with a green sphere. Only the best-scoring receptor is shown.

### 4.3.2 Simulated Annealing

The possibility of the XPLOR-NIH INPHARMA implementation to optimize models with simulated annealing was tested with the protocol described in the following. Simulated annealing is an important optimization technique for structure determination in NMR.

The molecular system will be modeled using an empirical molecular force-field. Its all-atom-dynamics will be simulated by solving the equations of motion considering a coupling to a heat bath.<sup>43</sup> The kinetic energy of the atoms will allow transitions across energy barriers of the energy landscape. The higher the temperature the higher energy barriers can be overcome. For structure determination experimental energy terms will be implemented adding a penalty energy on all conformations which don't fulfill the experimental restraints. The total energy of the system should be minimal when the conformation of the system is physically meaningful - e.g. the empirical force field energies are low - and when the experimental restraints are fulfilled. In this state the conformation of the model will correspond to the actual conformation of the real molecule, with a quality depending on the quality of the experimental data and their model as well as the used empirical forcefield. Infinitely slow simulated annealing of a system will yield convergence to the global minimum of the energy landscape<sup>26</sup> and find the model in best possible agreement with the experimental data.

The simulated annealing protocol was based on the minimization protocol described above (ch. 4.3). The only used INPHARMA force constant was  $k_{\text{INPH}} = 1,000,000$ . All energy terms, restraints and the sets of movable atoms were handled just as above.

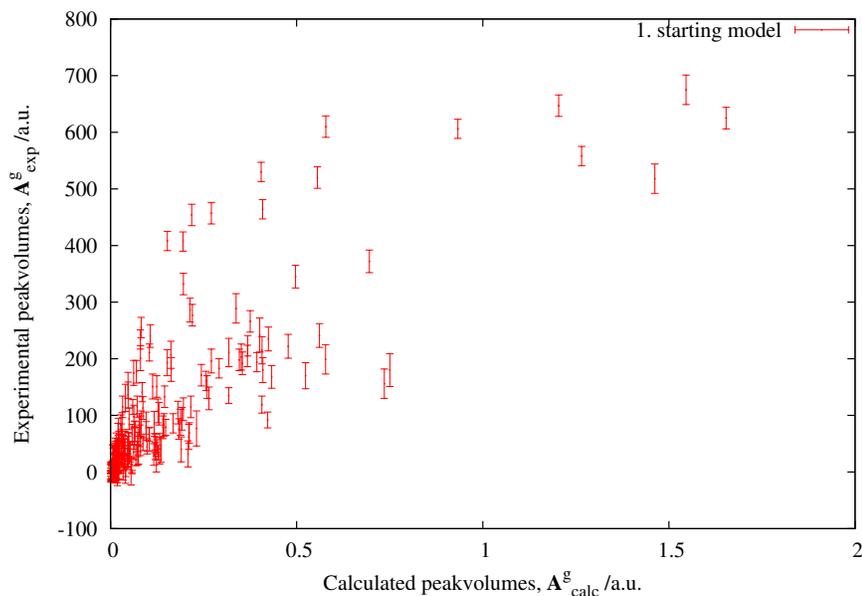
The simulated annealing protocol had several stages and can be found in the appendix (ch. 10.3.2, p. 181). After every step the simulated INPHARMA peaks, the regression results and all active energies were printed.

1. The starting model was initialized from two HADDOCK docking results, complexes of tubulin with either epothilone A or baccatin III.
2. Initial minimization using all energy terms, including INPHARMA, 50,000 steps.  
Minimization was automatically stopped after convergence, when the route mean square gradient of all forces was below  $10^{-9}$  kcal mol<sup>-1</sup>, Å<sup>-1</sup> or no further energy decrease was possible (generally the latter applies).
3. The system was heated to 250 K starting at 150 K from a initialized Maxwell distribution of velocities.

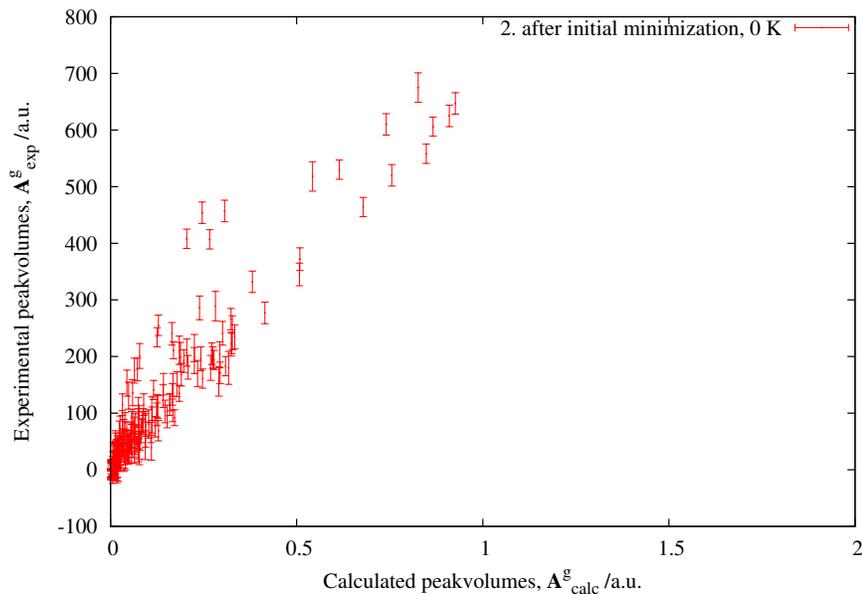
The initial time step was 0.1 fs, the center of mass motion was removed every 100 integration steps, every 50 steps the temperature was increased by 2.5 K, the temperature coupling factor FBETA was set to 500, for providing tight temperature control. The atom mass was set to 100 u for all atoms.

4. The system was cooled to a target temperature of 35 K. After the first 40, 60, 80, and 100 steps the time step was increased by 0.1 fs to a final value of 0.5 fs. Every 20 integration steps the target temperature was reduced by 1 K. The center of mass motion was removed every 10 integration steps.
5. The system was Powell-minimized 50 times using all energy terms and 5 steps per trial.
6. The system was Powell-minimized during 10,000 steps using all but the INPHARMA energy term. Minimization was stopped at convergence of the rms gradient to less than  $1 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  or when no further energy decrease was possible (generally the latter applied). Finally the INPHARMA energy term was switched on for re-scoring the INPHARMA peaks.

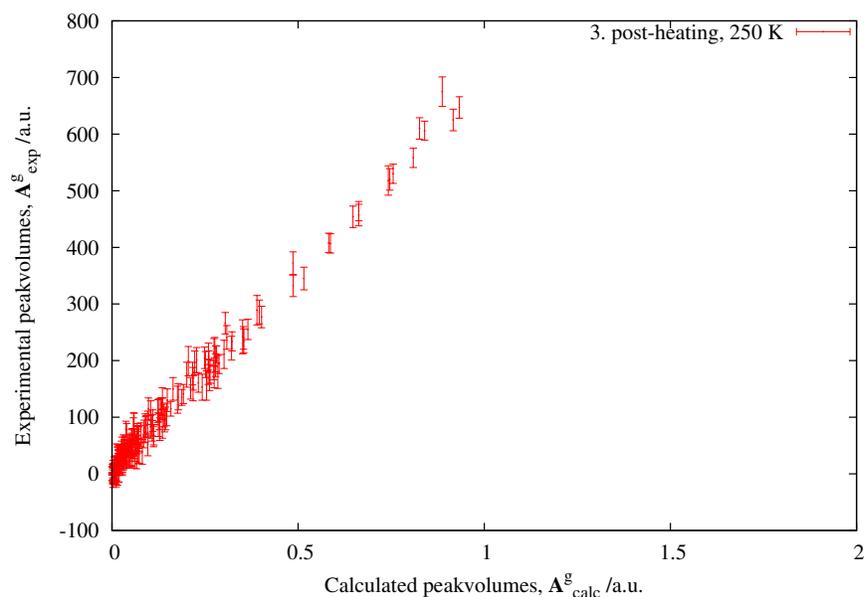
Examples of correlation plots of experimental and calculated data for all stages of the protocol can be found in the figures 4.40-4.45, p. 88 ff., clearly showing the possible improvements.



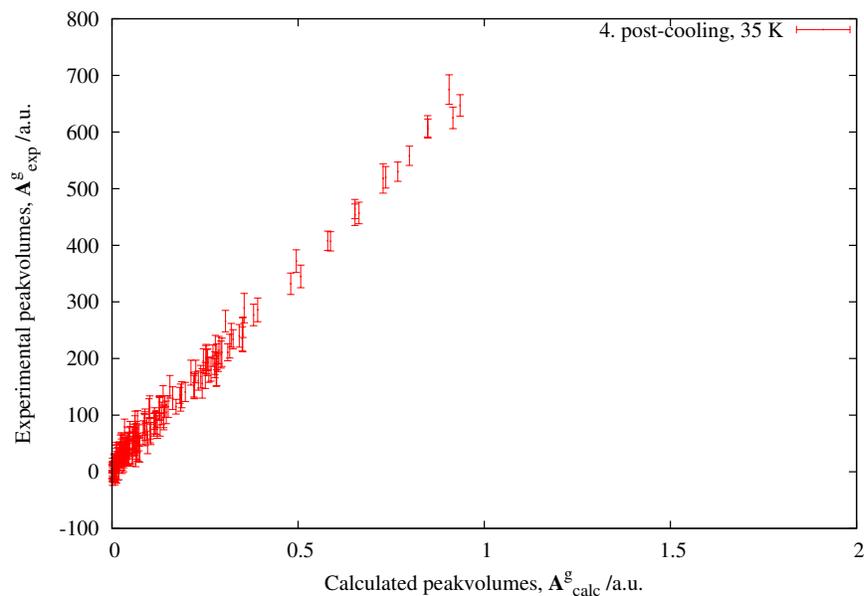
**Figure 4.40:** Correlation plot of experimental and calculated grouped peak volumes at stage 1, before initial minimization of the macromolecular test system (fig. 4.31, 4.32). The scaled INPHARMA energy is  $0.258 \text{ kcal mol}^{-1}$ . The following figures 4.41-4.45 refer to the same test system.



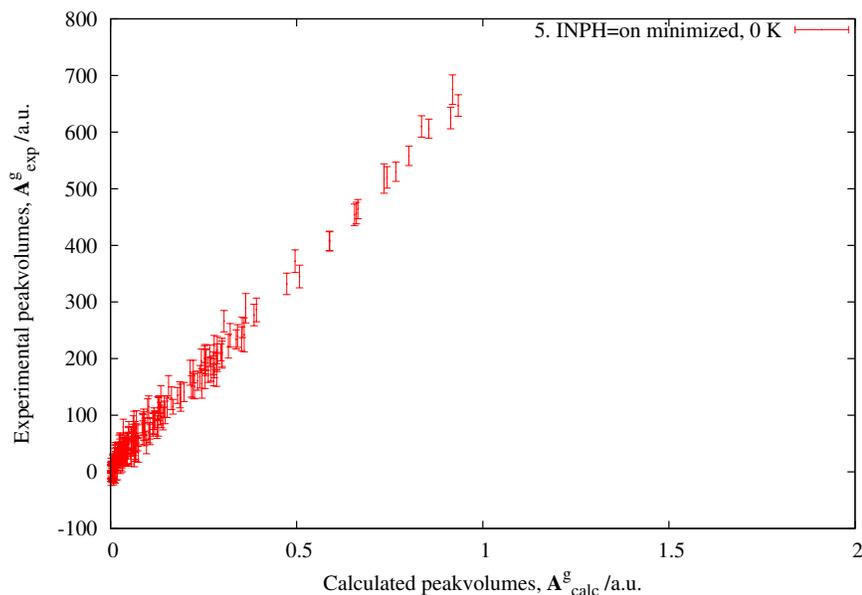
**Figure 4.41:** Correlation plot of experimental and calculated grouped peak volumes at stage 2, after initial minimization. The scaled INPHARMA energy is  $0.045 \text{ kcal mol}^{-1}$ .



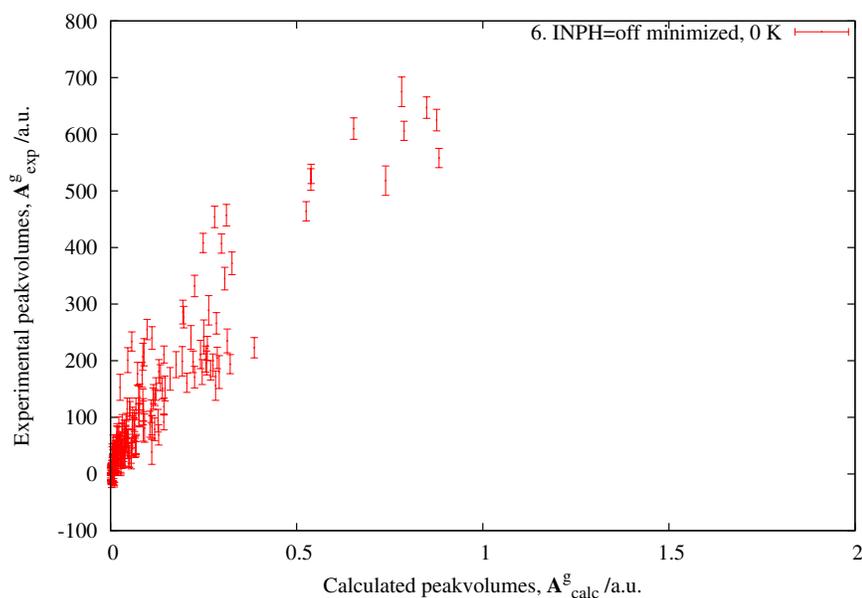
**Figure 4.42:** Correlation plot of experimental and calculated grouped peak volumes at stage 3, after heating. The scaled INPHARMA energy is  $0.014 \text{ kcal mol}^{-1}$ .



**Figure 4.43:** Correlation plot of experimental and calculated grouped peak volumes at stage 4, after cooling. The scaled INPHARMA energy is  $0.0115 \text{ kcal mol}^{-1}$ .



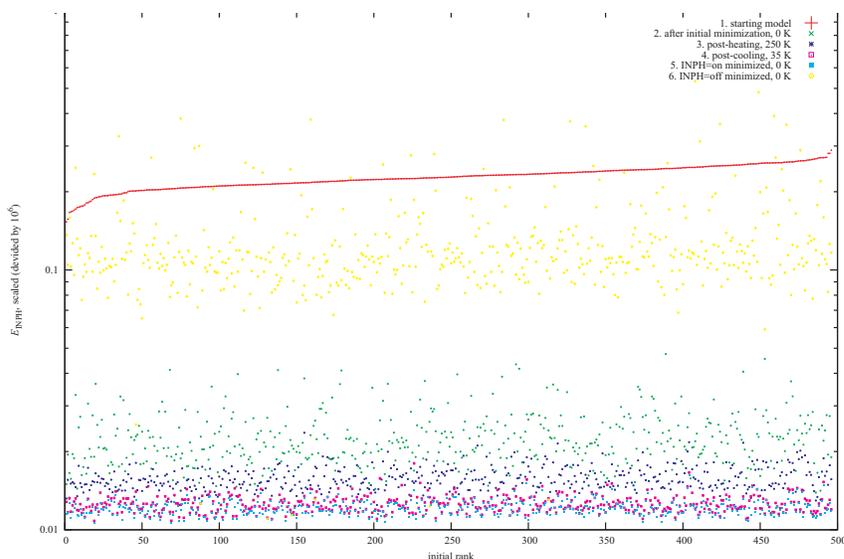
**Figure 4.44:** Correlation plot of experimental and calculated grouped peak volumes at stage 5, after cooling and minimization with INPHARMA *activated*. The scaled INPHARMA energy is  $0.0113 \text{ kcal mol}^{-1}$ .



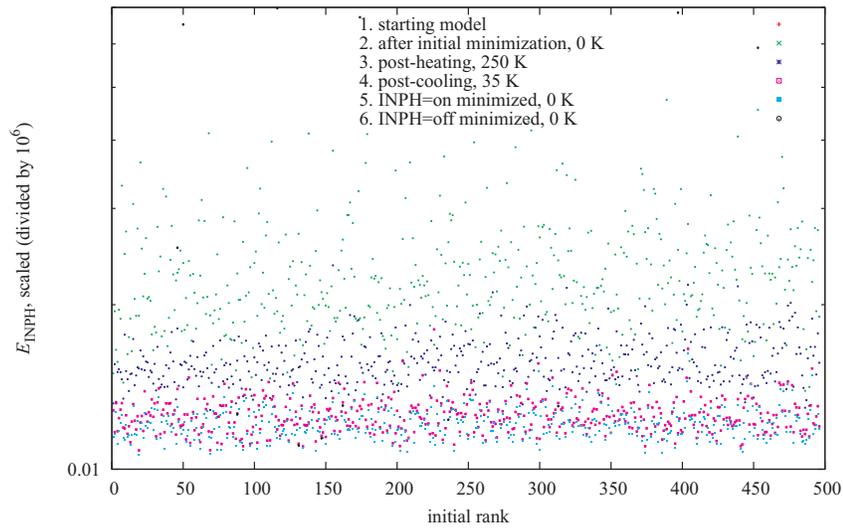
**Figure 4.45:** Correlation plot of experimental and calculated grouped peak volumes at stage 5, after cooling and minimization with INPHARMA *deactivated*. The scaled INPHARMA energy is  $0.059 \text{ kcal mol}^{-1}$ .

This protocol was run on the best 500 complex pairs of the 27 million HADDOCK docking pairs generated during my diploma thesis.<sup>11,12</sup> The ranking was based on the sorted regression coefficient  $R_0^2$  to the fit function  $f(x) = b \cdot x$ .

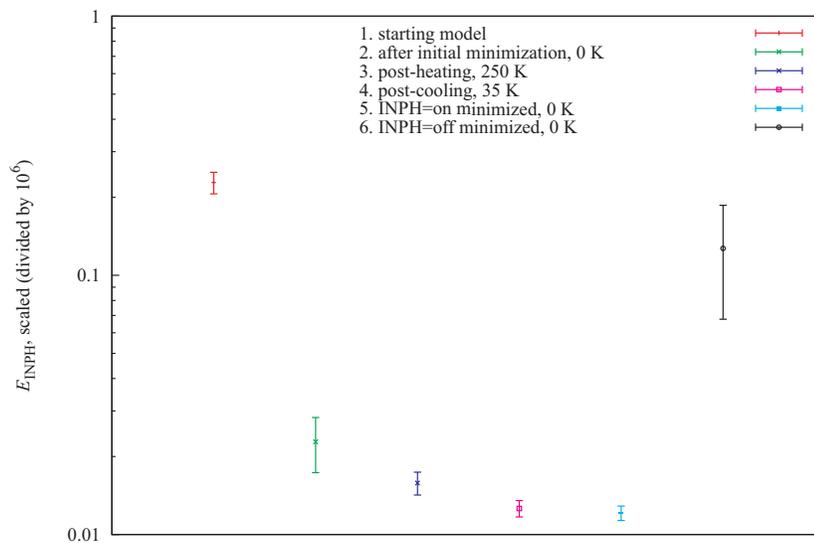
The scaled INPHARMA energy for all pairs and the various stages of the protocols are plotted as a function of initial rank in the figures 4.46 and 4.47. It can be seen that in essentially every stage of the protocol and for all models the INPHARMA energy is further minimized. Even after releasing the INPHARMA restraint in step 6 most of the final models have a lower INPHARMA energy than the initial model, although there are a few cases where the models from stage 6, being minimized without INPHARMA, actually become worse than the initial model. The average energies and their standard deviations at various steps of the protocol are given in fig. 4.48. The average energy convergence is clearly visible.



**Figure 4.46:** Scaled INPHARMA energies for 500 models in several stages of the simulated annealing protocol. The rank is sorted by initial INPHARMA energy.



**Figure 4.47:** Close up of the scaled INPHARMA energies for 500 models in several stages of the simulated annealing protocol (see fig. 4.46). The rank is sorted by initial INPHARMA energy.



**Figure 4.48:** Scaled INPHARMA energy averages and their standard deviations for 500 models in several stages of the simulated annealing protocol.

In figure 4.49 the evolution of a simulated annealing run, also including the minimization stages, is depicted. Shown is the temperature (red), total energy (green), total rms gradient (blue), and the INPHARMA energy

(magenta). The respective correlation plots of experimental and calculated data, for the different stages, can be found in figures 4.40-4.45.

In the beginning of the run, just after initial minimization (fig. 4.50), the temperature is not equilibrated yet and shortly reaches values of up to  $\approx 5500$  K (figs. 4.49, 4.51).

This kind of behaviors is expected after initial minimization as, due to the temperature initialization to 150 K, there is virtually an instantaneous temperature jump from 0 K to 150 K, leading to instabilities. Due to the short time step of 0.1 fs, the high atomic mass of 100 u, and the strong coupling to the temperature bath this initial temperature jump quickly decays. In general, for structure determination using experimental restraints, this kind of initial temperature jumps might be permissible. Initializing the temperature at a high value will save computer time and was justified for the initial protocol tests here.

Not only during the initial minimization but even during the ultra high temperature regime (caused by the temperature initialization) of the MD INPHARMA restraints get better fulfilled, as indicated by the steeply decreasing INPHARMA energy (fig. 4.51). The drop of INPHARMA energy is continuing during heating to 250 K, though on a much slower time scale.

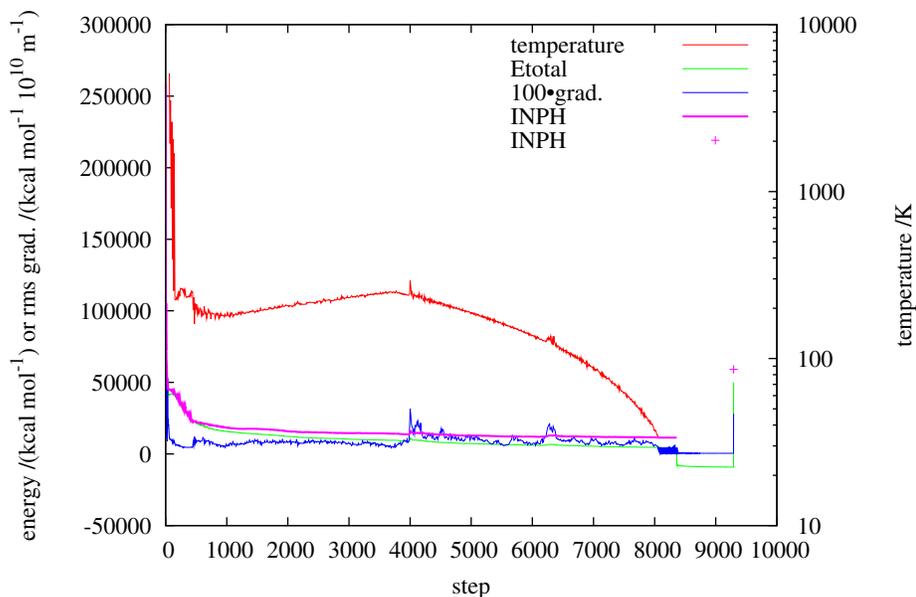
After heating the time step was step-wise increased to 0.5 fs. The system seems to react with an increased rms gradient and slightly higher fluctuations in all energy terms (fig. 4.51) (probably as it is not properly equilibrated which was not expected for this short run time of only approx. 370 fs for heating). In general the INPHARMA restraints are very strong in this example, using  $k_{\text{INPH}} = 1,000,000$ . This might make it necessary to use a very short integration time step due to the strong forces, and so harmonic frequencies, induced by the INPHARMA restraints.

Nevertheless the MD is stable and the temperature stays under control. During the annealing stage the two energies, INPHARMA and total, keep mainly decreasing (fig. 4.51). Also later, during minimization after annealing, including the INPHARMA term, the energies keep dropping (fig. 4.52). Here, the INPHARMA energy reaches its absolute minimum during this protocol. The scaled INPHARMA energy is just  $E_{\text{INPH,scaled}} = 11,400/1,000,000 = 0.014$ , indicating a very good fit of experimental and calculated data. Further minimizing, without using the INPHARMA energy, leads to a considerable drop in energy of about  $2000 \text{ kcal mol}^{-1}$  (fig. 4.52).

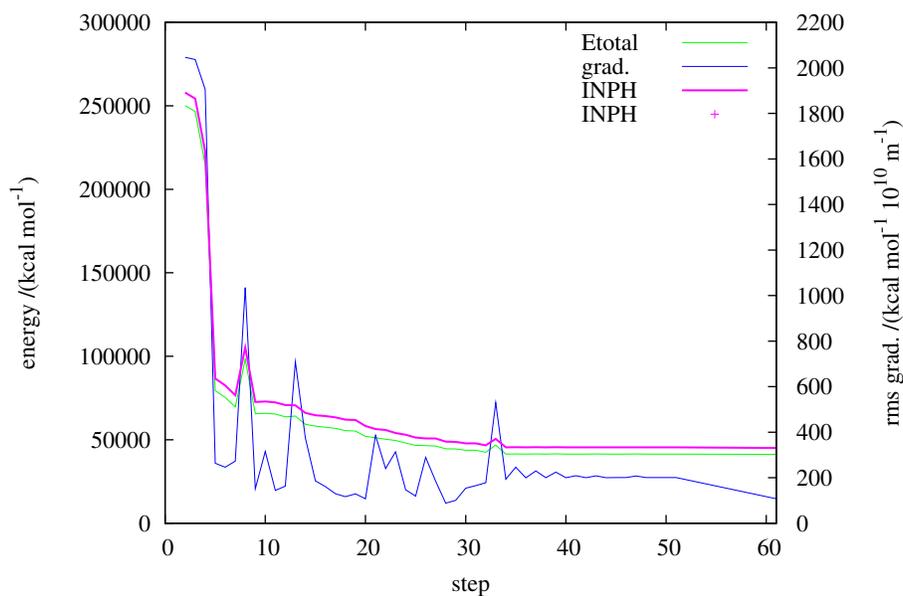
After this final minimization the INPHARMA restraints are not as well fulfilled (fig. 4.49, magenta cross) as during most steps in the course of the protocol. Nevertheless, the initial scaled INPHARMA energy was reduced

from 0.258 to 0.059 kcal mol<sup>-1</sup> yielding an all-over scaled energy drop of 0.199 kcal mol<sup>-1</sup>.

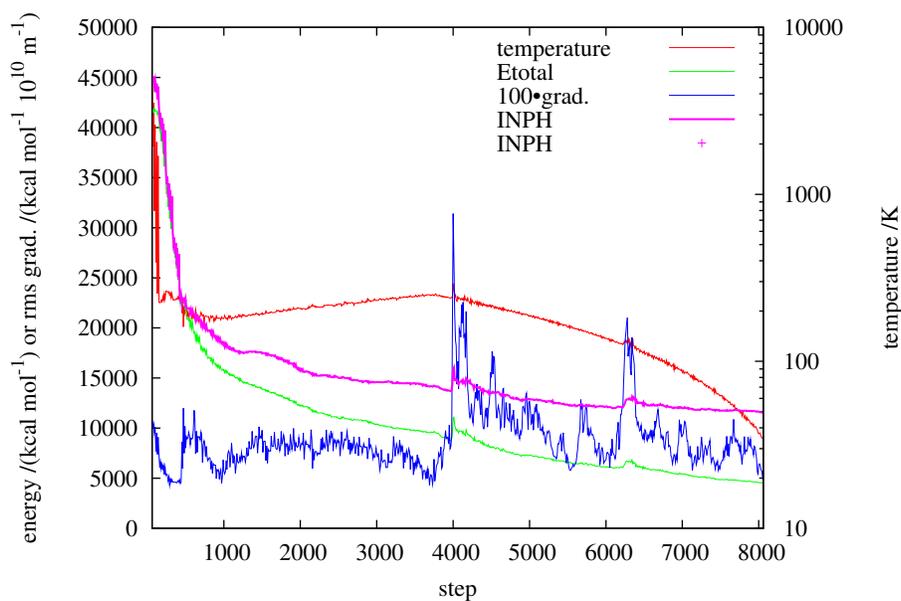
Even the initial minimization yielded an higher improvement of energy by reducing the scaled INPHARMA energy by 0.212 to 0.045 kcal mol<sup>-1</sup>. The difference is that the post-annealing model fulfills these restraints without actually having them applied, or being active in any way. No further artificial forces are active to keep an optimal conformation, in contrast to the model coming from initial minimization.



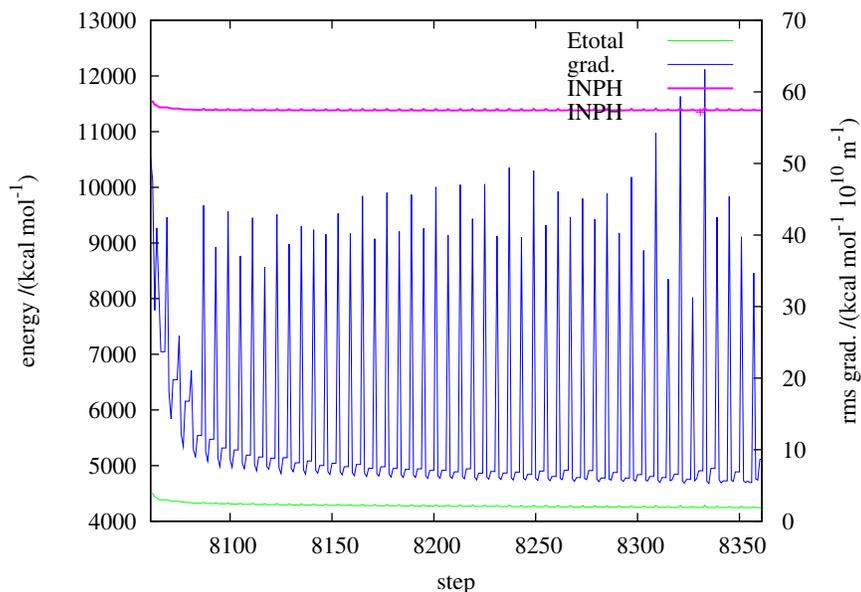
**Figure 4.49:** Overview over the complete simulated annealing run indicating the temperature (red), the total energy (green), the INPHARMA energy (magenta), and the rms gradient (blue). The temperature is only defined during molecular dynamics stages and is missing during minimization. The discrete jump at the end of the protocol, after the final minimization with INPHARMA *deactivated*, is caused by reactivating INPHARMA for scoring the final conformation.



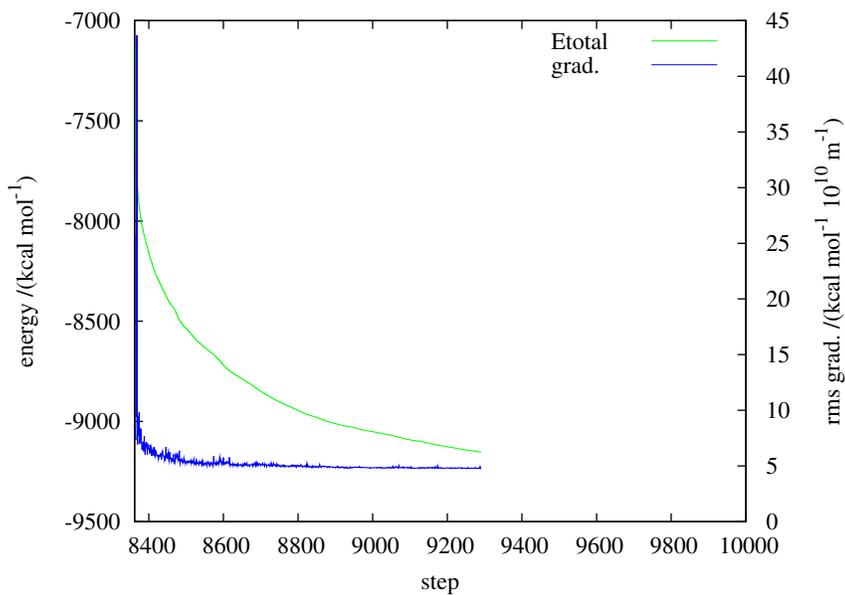
**Figure 4.50:** Evolution of selected variables during the initial minimization stage of the protocol. The corresponding initial and final correlation plots are given in the figs. 4.40 and 4.41.



**Figure 4.51:** Evolution of selected variables during the heating and cooling stages of the protocol. The corresponding correlation plot are given in the figs. 4.41, 4.42 and 4.43.

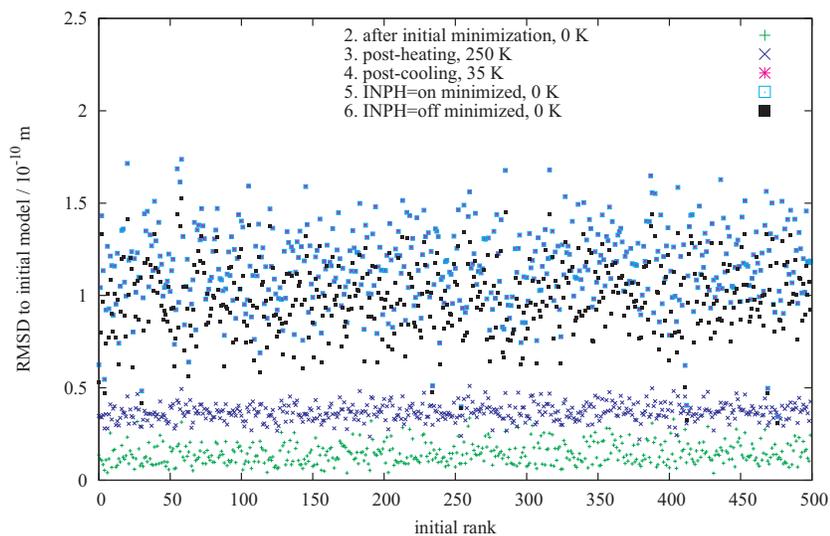


**Figure 4.52:** Evolution of selected variables during the post-annealing minimization stage of the protocol, having INPHARMA *activated*. The apparent peaks come from restarting minimization every 5 steps, requiring initial step size adaptation. The corresponding correlation plots are given in the figs. 4.43 and 4.44.

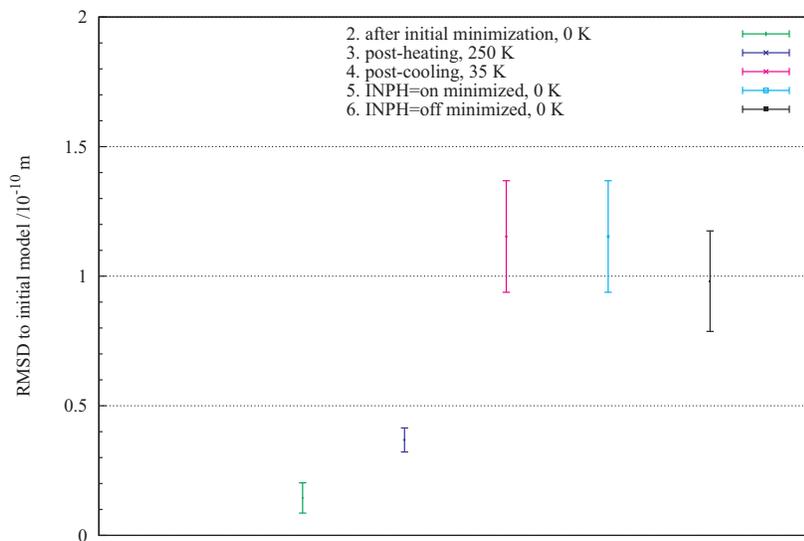


**Figure 4.53:** Evolution of selected variables during the post-annealing minimization stage of the protocol, having INPHARMA *deactivated*. The corresponding correlation plots are given in the figs. 4.44 and 4.45.

Allover, the structural changes during the run of this protocol are not very large. To give an impression on the extend of structural change during the discussed run, the figures 4.56 and 4.57 are provided. The evaluation of all RMSD changes of the present model with respect to the initial models (figs. 4.54, 4.55) shows an increasing distance to the initial model during all but the last stage of the protocol. The structures have *not* been fitted to each other before RMSD calculation and kept superimposed on the not movable atoms of the complexes. The movable atoms (all atoms of the residues having at least one atom closer than 14 Å to any of the ligands) of both complexes are considered for the RMSD calculation. The by far biggest changes in structure happen during the cooling stage. This correlates with a rather small improvement of the INPHARMA energy during the same stage (fig. 4.46). In the last stage, post-annealing minimization with INPHARMA *deactivated*, the structures become more similar to the initial structures again, going together with a decreasing fit-quality (fig. 4.46).

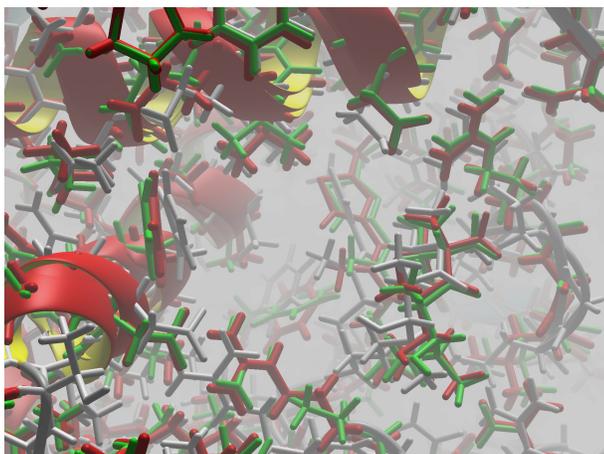


**Figure 4.54:** RMSD changes (including both complexes) with respect to the initial models for all 500 models in several stages of the simulated annealing protocol.

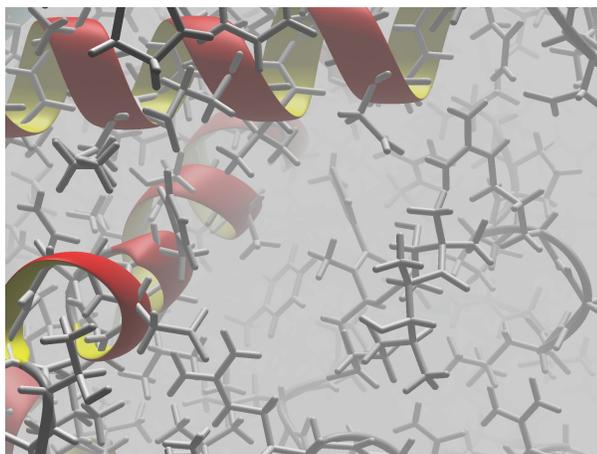


**Figure 4.55:** RMSD changes in respect to the initial models and their standard deviations for 500 models in several stages of the simulated annealing protocol. The structural change during stage 5, in respect to the previous stage, is vanishingly small. The RMSDs

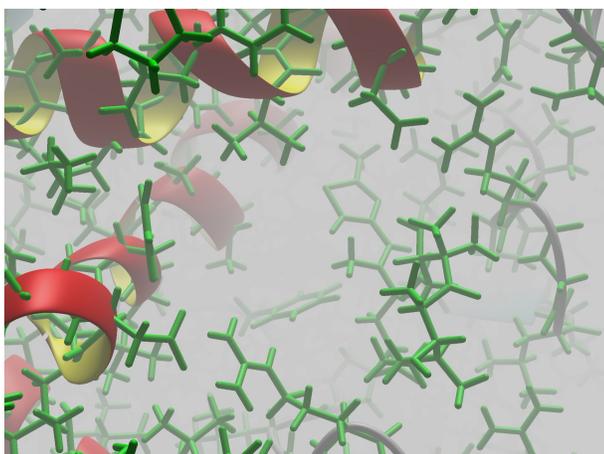




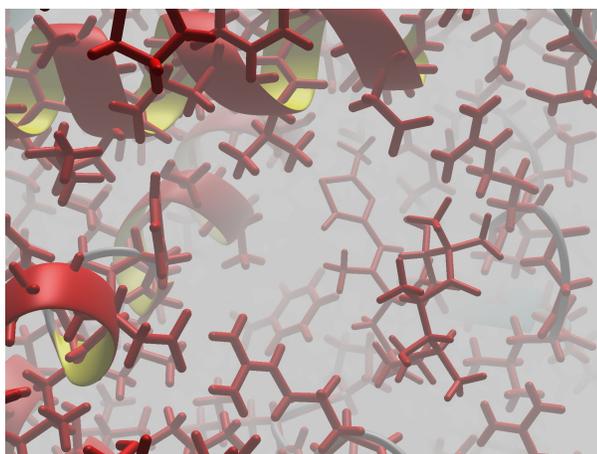
(a) Superposition of three epothilone A - tubulin models coming from different stages of the protocol, see further sub-figures.



(b) Initial model, after stage 1

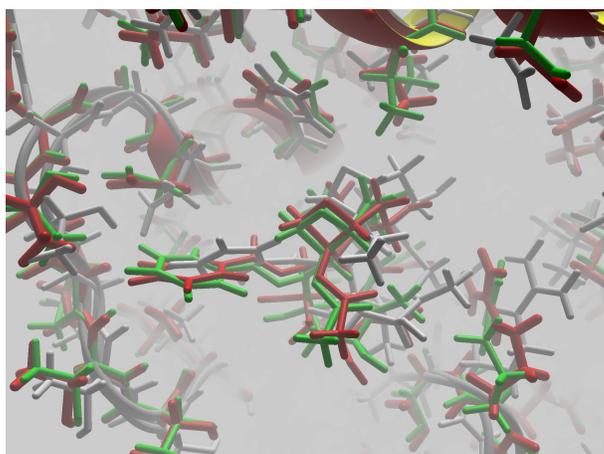


(c) Post-annealing model minimized with IN-PHARMA *activated*, after stage 5

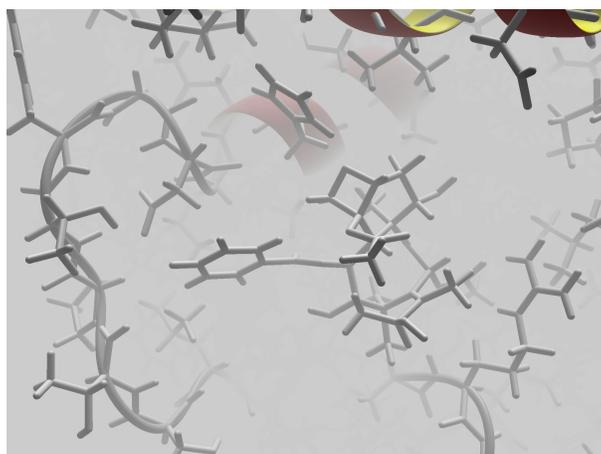


(d) Post-annealing model minimized with IN-PHARMA *deactivated*, after stage 6

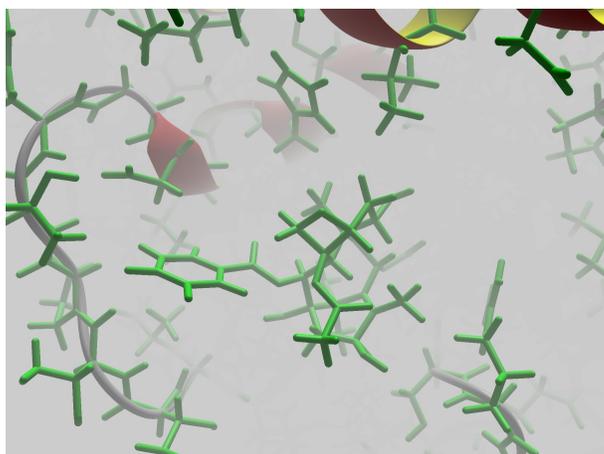
**Figure 4.56:** Here and in fig. 4.57 the structural influence of the simulated annealing run on the models is illustrated. The pair of complexes was picked randomly and is the one corresponding to the correlation plots (figs. 4.40-4.45) and the trajectory plots (figs. 4.49-4.53).



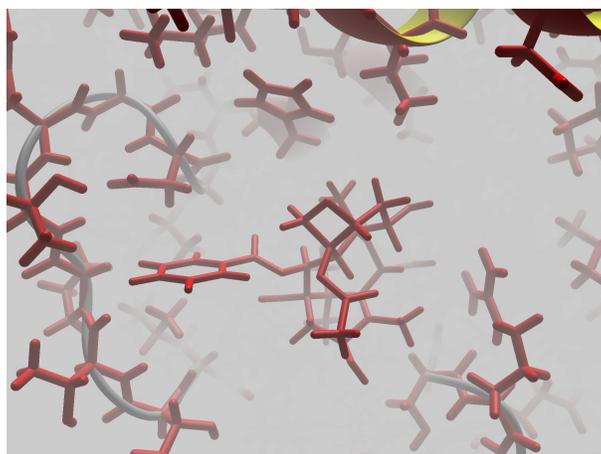
(a) Superposition of three BAC II - tubulin models coming from different stages of the protocol, see further sub-figures.



(b) Initial model, after stage 1



(c) Post-annealing model minimized with INPHARMA *activated*, after stage 5



(d) Post-annealing model minimized with INPHARMA *deactivated*, after stage 6

**Figure 4.57:** Here and in fig. 4.56 the structural influence of the simulated annealing run on the models is illustrated. The pair of complexes was picked randomly and is the one corresponding to the correlation plots (figs. 4.40-4.45) and the trajectory plots (figs. 4.49-4.53).

It is not easily possible to unambiguously pick a "correct" binding mode from the models generated by simulated annealing. Sorting them by the INPHARMA energy after stage 6, post-annealing minimization *excluding* INPHARMA, yields 10 best model pairs showing a large structural variation (fig. 4.58).

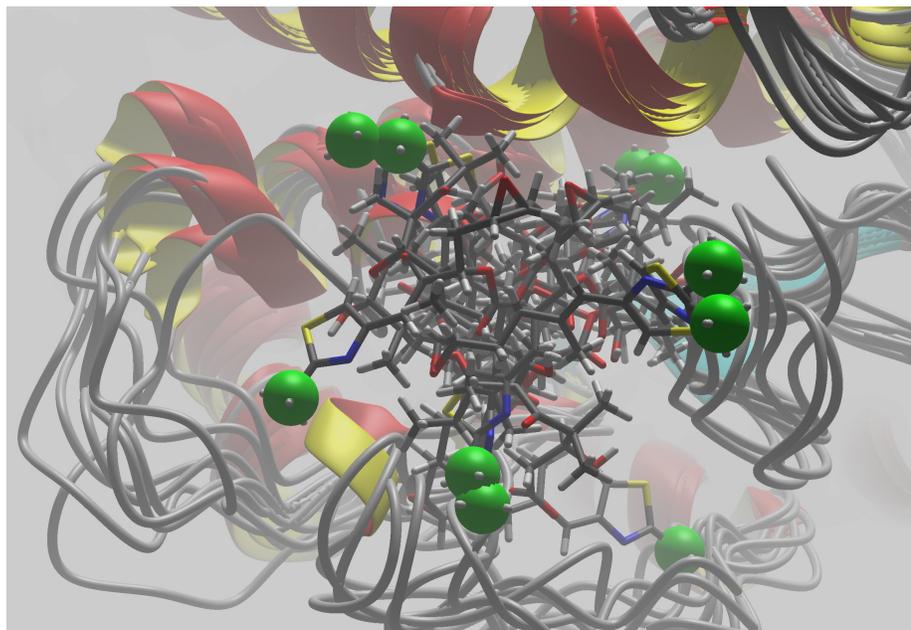
Sorting the models by the INPHARMA energy after stage 5, post-annealing minimization *including* INPHARMA, yields 10 different best model pairs (fig. 4.59). Here the structural variety seems to be smaller and more promising (see fig. 4.58). Still, based only on the INPHARMA energy, no model can be selected unambiguously as the correct complex.

In conclusion I believe that in the present state of structural refinement it is not possible to select the correct binding mode of the ligands exclusively based on INPHARMA. It might be that considering further data, like structure activity relationship (SAR) or mutational data, can be successfully used to further refine the number of possible binding modes. A further possibility would be to include further experimental NMR data. The easiest would be to take all remaining information from the measured INPHARMA spectra. Here, the diagonal peaks, intra-molecular cross peaks and peaks of vanishing volume were not considered. Only inter-ligand cross peaks, the INPHARMA peaks, were used for back calculation and for comparison with the experiment.

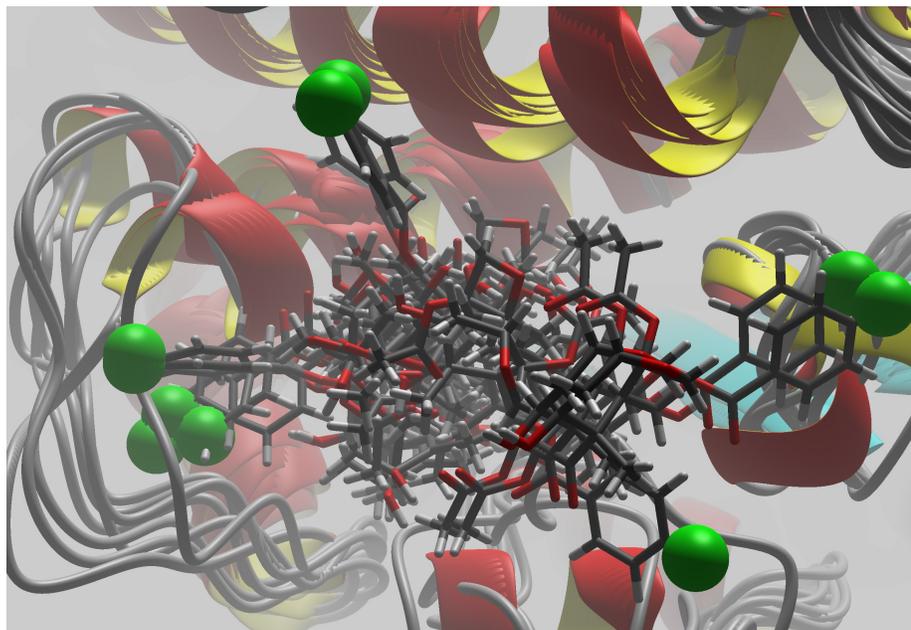
A further draw back in the present stage of the simulation is the lack of a state-of-the-art model of dynamics. The molecule is considered as rigid, without any internal dynamics. Implementation of appropriate dynamical models and the determination of appropriate parameters would be beneficial for using INPHARMA for structural refinement of receptor-ligand complexes.

The average CPU time required for one of the 500 simulated annealing protocols was 2:56 CPU hours. The longest run took 3:35 CPU hours and the shortest 2:19 hours. 4 cores of the type Xeon 5160 (Woodcrest, Dual Core, 3 GHz, 4 MB L2 Cache,  $\approx$  12 GFLOPS/core) have been used on a cluster of dual CPU (quad core) shared memory machines. The average CPU load was about 80%, which is likely close to the scaling factor of parallelization. XPLOR-NIH was compiled with the Intel compiler version 11.0. Highest optimization and auto-parallelization had been activated to create parallelized code for the non-INPHARMA parts of XPLOR-NIH. The INPHARMA part is mainly parallelized via the used Intel-MKL library 10.2.1.017.



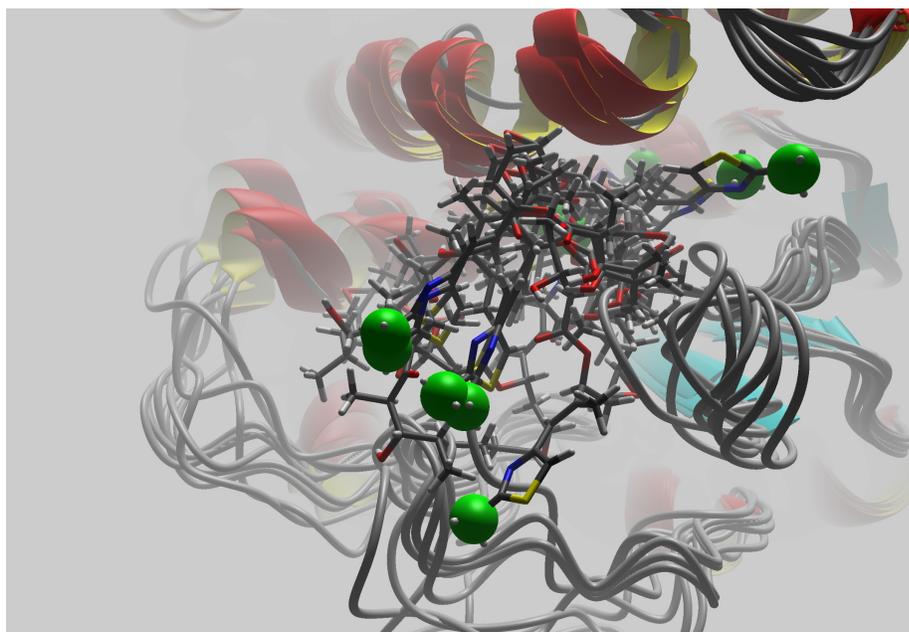


(a) Ten best binding modes of epothilone A with tubulin after simulated annealing and minimization *excluding* INPHARMA restraints.

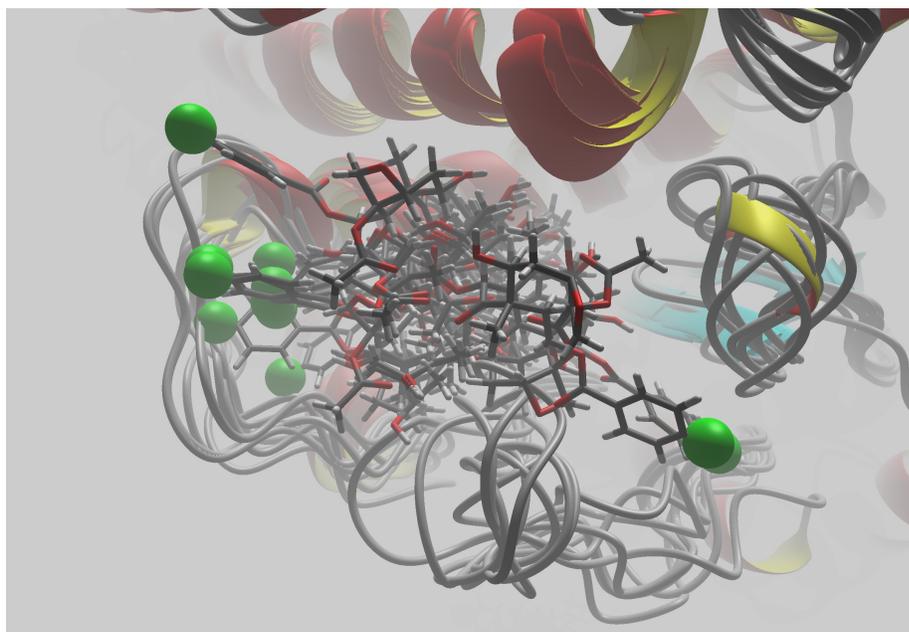


(b) Ten best binding modes of baccatin III with tubulin after simulated annealing and minimization *excluding* INPHARMA restraints.

**Figure 4.58:** The 10 best models after stage 6, post-simulated annealing minimized *without* INPHARMA. An exterior atom is highlighted in green to ambiguously emphasize the ligand orientations.



(a) Ten best binding modes of epothilone A with tubulin after simulated annealing and minimization *including* INPHARMA restraints.



(b) Ten best binding modes of baccatin III with tubulin after simulated annealing and minimization *including* INPHARMA restraints.

**Figure 4.59:** The 10 best models after stage 6, post-simulated annealing minimized *including* INPHARMA. An exterior atom is highlighted in green to ambiguously emphasize the ligand orientations.



# Chapter 5

## Conclusion and Outlook

There are several established NMR techniques based on the Overhauser effect to probe for the interaction of ligands with their receptor.<sup>1-3</sup> Techniques of special interest to structure determination are the transferred nuclear Overhauser effect (trNOE),<sup>5-7</sup> inter ligand Overhauser effect (ILOE)<sup>8,9</sup> and inter ligand NOEs for pharmacophore mapping (INPHARMA).<sup>10-14</sup> These methods make use of the NOESY experiment<sup>4</sup> under the condition of chemical exchange of the ligands. On a general level they are mathematically equivalent and all of these experiments are sensitive to both, the bound conformation of the ligands and of the receptor.

Until now the use of these methods for structure determination is limited. Spectra can be simulated from a given structural model and compared to an experiment to judge the validity of the given model in respect to the experimental data. To my knowledge, for the experiments above, all involving chemical exchange of the ligands, no publications exist which demonstrate systematical improvement of the quality of a given complex model, including the receptor conformation, or ways to find the best possible structural model explaining the experiment.

Here I have shown, for the example of INPHARMA, one possible implementation of such structural refinement protocols against chemical exchange and spin diffusion determined NOESY intensities.

I have introduced an R-factor based penalty energy to judge the agreement of a given conformation with the experimental data. Then I calculated the analytical derivative of this energy in respect to the coordinates of all involved nuclei. The theory considers the details of chemical exchange, the initial concentrations of all species, the chemical shift overlap of protons, and the necessity to symmetrize the effective relaxation matrix for efficient matrix diagonalization.

The theory can be easily applied to trNOE, ILOE, and NOESY spectra with conformational exchange by substituting the relevant matrices to the corresponding ones required for the simulation of the respective spectrum.

To evaluate the method the analytical INPHARMA gradient was implemented into the molecular dynamics and structure determination program XPLOR-NIH.<sup>36,37</sup> Simple 6 and 12-spin systems have been used for initial tests (ch. 4.1,4.2). The validity of the analytical INPHARMA gradient was demonstrated in comparison with a numerical gradient (ch. 4.1.2); the shape of the energy landscape has been exemplified, different versions of penalty functions have been tested and evaluated, and the efficiency of steepest decent minimization has been demonstrated in a grid search of the energy landscape (ch. 4.1.3; figs. 4.13, 4.14) or from random coordinates (ch. 4.2). Finally, using the macromolecular system of tubulin, epothilone A and baccatin III, with experimental INPHARMA data, the technical applicability and reliability of XPLOR-NIH implemented INPHARMA restraints for minimization (ch. 4.3.1) and simulated annealing protocols (ch. 4.3.2) has been demonstrated.

Although the new XPLOR-NIH implemented INPHARMA restraints work technically reliably there is still space for improvement:

Obviously, the presented protocols for minimization and simulated annealing need to be optimized and evaluated. The heating and cooling scheme as well as the kind and strength of necessary restraints have to be understood. Experimental peaks have to be left out for stability analysis (R-free calculations). An ideal system for these evaluation would be protein kinase A (PKA).<sup>13</sup> Here high-resolution crystal structures as well as high quality INPHARMA spectra are available. The result of INPHARMA could directly be compared to the reference structures from X-ray. The resolution of INPHARMA, using analytical gradients, could be determined and models of internal dynamics could be evaluated and optimized.

Until now only the most simple "dynamical" model, the isotropically tumbling rigid molecule, is considered. This is not a general limitation. For simple dynamical models, not depending of internuclear distances, only the NOE relaxation matrix needs to be adapted. For distance or angle depended models it should be straight forward, in most cases, to adapt the derivate  $\partial \mathbf{R}_{mn} / \partial r_{ij}$  in formula (2.32) to the new model. Also joint atom models, like for methyl groups, can and should be considered.

The first simulated annealing protocols, presented here, refine the structures beyond the next local minimum. Nevertheless, the protocols are just first trials and need to be optimized and evaluated, favorably for complexes of known structure. It is expected that the resolution and accuracy of INPHARMA, and similar techniques, will be significantly improved beyond

previous levels.

The speed of the XPLOR-NIH implementation can likely be increased significantly. A problem of the present approach is that it relies on the diagonalization of a rather large matrix, comprising all ligand protons and the receptor protons in the vicinity of the ligands. In principle it should be possible to reduce the size of these matrices by choosing smaller subsystems around nuclei of interest. Due to the  $\mathcal{O}(N^3)$  nature of the matrix diagonalization it might be favorable to rather compute many small diagonalizations than a single large. This concept is well established for the simulation of NOESY spectra of single molecules.<sup>22</sup> It needs some thought how to reliably transfer the concept to a system under chemical exchange, where the spatial environment of a single spin is not clearly defined.

Further gain in speed could also be achieved by implementing some update criteria for the matrix derivative calculation. In general the matrix elements and the energy change continuously and smoothly with the change of coordinates. It might be sufficient to evaluate the gradient only every few steps of a protocol; also simple linear or nonlinear forward prediction could be used. Furthermore matrix diagonalization might be completely avoidable if the gradient calculation can be transferred to an "integral expression for the gradient".<sup>29</sup> Matrix diagonalization (via SVD) and matrix multiplications might also be shifted to the GPU (graphics processor unit) of a graphics card using CUDA.<sup>44,45</sup> As these GPUs are powerful vector computers with  $\approx 200$  gigaFLOPS double precision floating point performance (AMD FireStream 9250 GPU; June 16, 2008),<sup>46</sup> compared to about 12 gigaFLOPS for a core used here (Intel Xeon 5160, Dual Core, 3 GHz, 4 MB L2 Cache), a significant speed up might be achievable.

Using the presented example of calculating the gradient of the INPHARMA energy, it is to be seen if this approach can be extended to other NMR spectra using the nuclear Overhauser effect in a chemical exchange situation. This includes the STD (saturation transfer difference),<sup>17</sup> trROE (transferred rotating frame Overhauser effect)<sup>7</sup> and WaterLOGSY<sup>18</sup> experiments.

The fast and accurate determination of structural models of drug-target interactions is essential for increasing the speed and effectiveness in structure-based drug design. I hope that the work presented here will be a good starting point for further progress in structure based drug design using NMR techniques relying on the Overhauser effect.



# Part II

## DNP



# Chapter 6

## Introduction

Nuclear magnetic resonance is a very powerful and versatile tool for the analysis of chemical compounds. The applications of NMR in chemistry and biochemistry cover the identification, structure determination and dynamical analysis of molecules of sizes ranging from small molecules with a molecular weight of a few Dalton to large macromolecular complexes in the megadalton range. Magnetic resonance imaging (MRI) is an extremely powerful tool for scientific research and medical diagnosis. It gives insight to the inner structure and function of living organisms without interference with physiological processes.

A general drawback of all NMR techniques is their low sensitivity due to the small magnetic moment of nuclear spins. The small Zeeman energy in a magnetic field results in only a very small net magnetization in most situations. Given a magnetic field of 14 T, corresponding to 600 MHz proton Larmor frequency, one out of 20,000 protons can be effectively detected at 25°C.

Dynamic nuclear polarization (DNP) is an umbrella term for various methods for increasing the nuclear polarization of a sample, normally by polarization transfer from inherently higher polarized paramagnetic centers. DNP is going back to the beginnings of magnetic resonance and has first been predicted by Overhauser in 1953.<sup>47</sup> Not much later in the same year Carver and Slichter<sup>48</sup> observed DNP, for the first time, in a system of metallic lithium under electro-magnetic irradiation of the conduction electrons. In principle, for a nucleus-electron system, the complete saturation of the electron transitions can lead to a dynamic nuclear polarization equal to the Boltzmann polarization of the electron. The term "dynamic" in "dynamic nuclear polarization" describes that the nuclear polarization is off-equilibrium, and thus decays back to its Boltzmann value after irradiation.

Recently, in the context of the structural biology and magnetic resonance

imaging, a lot of attention has been drawn to DNP for the application to sensitivity enhancement of nuclear magnetic resonance in liquids and solids. The nucleation point was the impressive signal enhancements achieved in Bob Griffin's laboratory at MIT for solid state NMR using magic angle spinning (MAS).<sup>49-51</sup> Until now, they have observed large signal enhancements at high field, like recently a factor of 250 using a the biradical bis-TEMPO-bisketal,<sup>52</sup> opening completely new perspectives for solid state NMR in structural biology.

The group of Golman at GE Healthcare has demonstrated the utilization of very high spin polarization created in the solid state at very low temperatures for liquid state NMR.<sup>53</sup> For detection, the sample was quickly dissolved and flushed to a high-resolution NMR magnet. A large fraction of the high polarization on  $^{13}\text{C}$  and  $^{15}\text{N}$  is conserved. Enhancement factors  $>10,000$  were observed for small molecules like pyrovate- $^{13}\text{C}$ , inducing particular interest of parts of the in vivo MRI community.

These achievement from the Griffin and Golman laboratories stimulated several groups to dedicate them to further studies on all aspects of DNP.

The project presented here is part of the EU Framework 6 "Bio-DNP Design Study". The scope of this study is DNP for high-resolution NMR in liquids and structural biology. It covers the aspects of hardware construction and optimization of DNP by understanding its limits using experimental, theoretical and numerical methods.

DNP requires the saturation of the electron transitions in the sample, normally by the use of microwaves of more than 9 GHz frequency. As polar liquids (like water) show severe dielectric absorption of microwaves, DNP is hampered for biological samples in liquids.

Several approaches have been developed for creating hyperpolarized solutes in liquid solution:

The approach by Golman<sup>53</sup> uses a dedicated external polarizer for polarizing the sample at the temperature of 1.2 K at a field of 3.35 T. A quick dissolution and transfer step to high resolution NMR detection follows. This approach is limited by the magnetization losses during sample transfer, especially on protons. Other nuclei like  $^{13}\text{C}$  and  $^{15}\text{N}$  show strong polarization after transfer. The observed enhancement factors were  $\epsilon=44,000$  for  $^{13}\text{C}$  and  $\epsilon=23,500$  for  $^{15}\text{N}$ , compared to the 9.4 T Boltzmann signal at room temperature. These large enhancement factor were only possible as polarization is performed at the low temperature of 1.2 K. Here the electron polarization at 3.35 T is 95.4% or 127 times higher than at the same field at room temperature (298 K). In these experiments, after sample transfer, the enhancement is very unequally distributed over the spins in the molecule. Sometimes

one even observes different signs of enhancement, depending on the specific nucleus. Conventional multidimensional NMR spectra require the repeated application of pulse sequences on the very same sample. As the polarization life time in dissolution DNP is limited several techniques have been developed using only a fraction of the magnetization for each experiment<sup>54</sup> or by performing all experiments simultaneously on slices of the sample.<sup>55</sup> Nevertheless, the application of the dissolution method to biological macromolecules has not been demonstrated. The required dissolution step, using hot solvents, will probably be incompatible with biological macromolecules.

In the case of the *in-situ* temperature-jump method, developed in Bob Griffin's laboratory, low-temperature DNP and liquid-state NMR detection are performed in the same place.<sup>56,57</sup> The polarization is created in the frozen solid at 80 K and NMR is detected at 300 K after *in-situ* melting, using an IR laser. After NMR detection the sample gets frozen again and the process is restarted with another polarization period. The repetition time is about 1 min. Enhancement factors of approximately 400 have been observed for  $^{13}\text{C}$ <sup>56</sup> and the possibility to perform 2D  $^{13}\text{C}$ - $^{13}\text{C}$  spectroscopy has been demonstrated.<sup>57</sup> Although the method might be applicable to some proteins, in general the sample will suffer from the repeated heating and cooling.

Walter Köckenberger from the University of Nottingham has orally reported<sup>58</sup> on an integrated hardware system for dissolution DNP. Here a dual magnet is used, comprising two superconducting coils operating at 3.4 T and 9.4 T, respectively. The sample is polarized at the low temperature of <1 K at low field. Then it is transferred in a few hundred milliseconds to the high resolution part of the magnet and dissolved for immediate detection. During transfer the sample is in the frozen state and is passing through still high magnetic stray fields. The magnetization loss during sample transfer is significantly reduced compared to the Goldman approach. Both, the high field and the sample being frozen, contribute significantly to reduced magnetization losses. The sample will be dissolved just above the NMR probe and is immediately flushed to the probe head. Enhancement factors of many thousands have been reported for  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  for various samples, not including macromolecules.

Although high dimensional spectra can be recorded with the previously mentioned methods of dissolution DNP, extended repeated acquisition on identical samples, necessary for acquiring high dimensional NMR spectra, is not possible for any of them, yet.

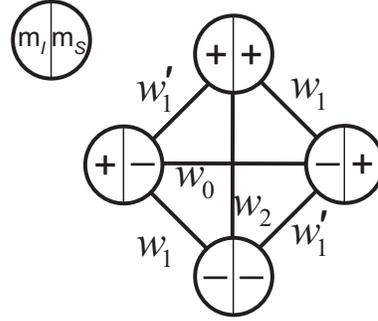
Direct polarization of liquid samples at the desired temperature and high magnetic field, suitable for high-resolution NMR, is possible, as demonstrated by Thomas Prisner's group at the university of Frankfurt.<sup>59</sup> En-

hancement factors for the protons of water molecules of more than 10 have been reported in a magnetic field strength of 9.2 T (260 GHz EPR, 400 MHz  $^1\text{H}$  NMR). To minimize the amplitude of the electrical field at the sample, which is responsible for heating, a helical resonator was used and the sample was placed axially on a node line of the electrical field. The diameter of the sample had to be restricted to 0.05 mm, a small fraction of the microwave wave length of 1.15 mm. Only very small sample volumes of 3-4 nL were available. Even if proteins could be effectively enhanced in this setup, normally much more sample would be available than usable. Using all of it (without DNP) would easily be beneficial. It keeps to be seen if larger molecules can be effectively polarized in this setup. The Overhauser effect depends strongly on the field and timescale of the electro-nuclear distance variation (ch. 6.1). Smaller molecules and lower fields show stronger enhancement.<sup>60</sup> It remains to be seen if larger molecules can get enhanced as well.

As later explained in more detail, we followed the approach of high resolution shuttle DNP, polarizing the sample at low field (0.34 T) and then transferring the sample quickly to high field (14.09 T) for detection. Due to the relatively low microwave frequency of only 9.7 GHz at low field and a large wave length of 31 mm the sample volume was bigger, compared to polarizing at high field. The microwave components for the used X-band frequency range are relatively easy to obtain and to handle and higher enhancement factors can be achieved at lower fields.<sup>60</sup> Significant signal enhancement factors, compared to the 14 T Boltzmann magnetization, on several nuclei and solvents were reported.

## 6.1 Brief Theory

DNP in liquids is governed by the Overhauser effect (dipolar cross relaxation) and scalar cross-relaxation.<sup>60</sup> The system can be described by a simple 2-spin system of the spins  $I$  and  $S$  (fig. 6.1), corresponding to a nuclear and electron spin in solution, respectively. The relaxation rates between the states decide on the possible size and sign of DNP for a given system.



**Figure 6.1:** Symbolic scheme of the electronic ( $S$ ) and nuclear ( $I$ ) Zeeman levels with their transition rates.

From these rates the enhancement factor  $\epsilon$  can be derived as a function of several parameters:<sup>60</sup>

$$\epsilon = \frac{\langle I_z \rangle}{I_0} = 1 + \xi f s \frac{\gamma_S}{\gamma_I} = 1 - \xi f s \frac{|\gamma_S|}{\gamma_I}. \quad (6.1)$$

Here  $I_z$  and  $I_0$  are the dynamic and Boltzmann polarization, respectively,  $\xi$  is the coupling factor,  $f$  the leakage factor, and  $s$  the saturation factor.  $\gamma_S$  and  $\gamma_I$  are the gyromagnetic ratios of the electron and nucleus, respectively. The coupling factor depends on the involved relaxation mechanisms, having a value of "-1" for complete scalar relaxation and of "+0.5" for dipolar relaxation.

In detail, the parameters are defined as following:

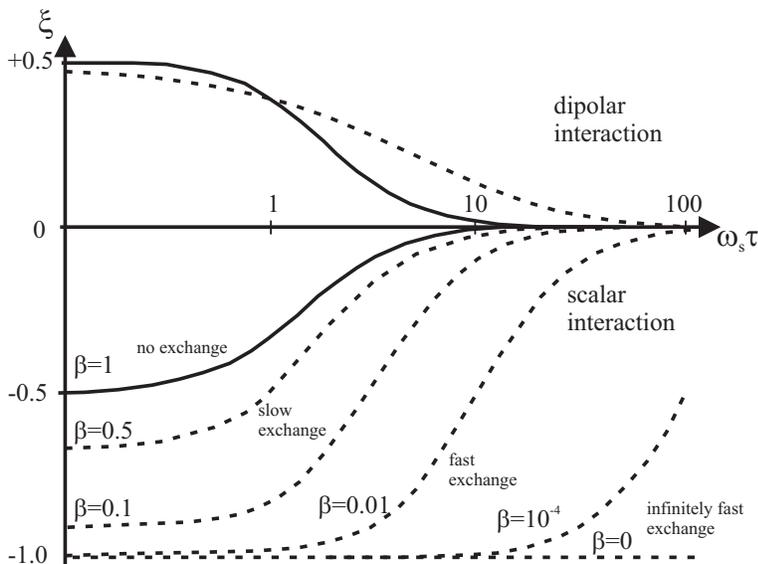
$$\begin{aligned} \xi &= (w_2 - w_0) / (w_0 + 2w_1 + w_2) = \sigma / \rho \\ f &= (w_0 + 2w_1 + w_2) / w^t = \rho / (\rho + w^0) \\ s &= (S_0 - \langle S_z \rangle) / S_0 \end{aligned} \quad (6.2)$$

where  $w_0$ ,  $w_1$ , and  $w_2$  are the zero, single, and double quantum transition rates, respectively. It is defined:  $\sigma = w_2 - w_0$  and  $\rho = w_0 + 2w_1 + w_2$ .  $w^0$  is the sum of all relaxation rates not involving the electron spin  $S$ .

The leakage factor  $f$  is approaching the value of "1" for high radical concentrations, as here non-electronic contributions  $w^0$  to the relaxation rates can be neglected.

The saturation factor  $s$  approaches the value of "1" for higher microwave irradiation powers, resulting in higher  $B_1$  fields.

The coupling factor depends on the dynamics of the electron-nucleus distance and on the field strength  $B_0$ , as shown in fig. 6.2. It is generally lower for higher fields and slower motions. The latter fact means that smaller radicals are better polarizers and that smaller target molecules (solutes or solvents) get better polarized.



**Figure 6.2:** The coupling factor  $\xi$  depends on the electron Larmor frequency  $\omega_S$  and on the correlation time of the electron-nuclear distance  $\tau$ . Here it is plotted as a function of  $\omega_S \tau$  for the dipolar and scalar relaxation mechanisms. For the dipolar mechanism (top) two models of motion are considered (dashed line: translational motion; solid lines: rotational motion). For higher fields and slower motions the coupling factor is lowered, reducing the possible enhancement factor. For the scalar mechanism (bottom) the correlation time of chemical exchange (binding) of the paramagnetic center to the polarized nucleus is considered in the factor  $\beta$ . For  $\beta = 1$  the radical will permanently bind to the polarized species. This figure is adapted from reference [60].

This size and field dependence can be illustrated by the fact that the dipolar and scalar cross relaxation, and thus DNP in liquids, is induced by the randomly fluctuating electro-magnetic fields created by the moving molecules (carrying spins and so magnetic moments) in solution. Translations, rotations and vibrations of a certain frequency will randomly excite state transitions of the same frequency. For too high fields or too large molecules these transitions can not get excited anymore by translational or rotational global motions as no corresponding, fast motion is present anymore. Intramolecular motions like methyl group rotations or bond vibrations will still provide possible relaxation pathways.

The rather short effective correlation time for the dipolar interaction of the protons of water to the electron spins of the radical TEMPOL is in the range of 15-20 ps.<sup>61</sup> Using the oversimplified model that this is the time needed for a rotation of 1 radian, the corresponding frequency would be 8-11 GHz which is the range of the here used 9.7 GHz electron Larmor frequency at X-band (0.34 T). Complete descriptions of the motions and the resulting amplitudes and frequencies of the random fields involve spectral density functions, defining the relaxation rates.



# Chapter 7

## Summary

In this chapter the DNP relating results obtained during my dissertation work will be summarized. Most of the results on DNP have been published in peer reviewed articles or are submitted for publication, as listed in the following . These articles or submitted manuscripts are available to my doctorate examiners, including supplementary information. The articles will be made available in the published version of this thesis, depending on the agreement of the respective copy right owner. A list of my specific contributions can be found in the appendix (ch. 11, p. 197).

- P. Höfer, G. Parigi, C. Luchinat, P. Carl, G. Guthausen, M. Reese, T. Carlomagno, C. Griesinger, and M. Bennati. Field Dependent Dynamic Nuclear Polarization with Radicals in Aqueous Solution. *Journal of the American Chemical Society*, 130:3254-3255, 02 2008. (ref. 61)
- P. Höfer, P. Carl, G. Guthausen, T. Prisner, M. Reese, T. Carlomagno, C. Griesinger, and M. Bennati. Studies of Dynamic Nuclear Polarization with Nitroxides in Aqueous Solution. *Applied Magnetic Resonance*, 34:393-398, 2008. (ref. 62)
- M. Reese, D. Lennartz, T. Marquardsen, P. Höfer, A. Tavernier, P. Carl, T. Schippmann, M. Bennati, T. Carlomagno, F. Engelke, and C. Griesinger. Construction of a Liquid-State NMR DNP Shuttle Spectrometer: First Experimental Results and Evaluation of Optimal Performance Characteristics. *Applied Magnetic Resonance*, 34:301-311, 2008. (ref. 63)
- M. Reese, M.-T. Türke, I. Tkach, G. Parigi, C. Luchinat, T. Marquardsen, A. Tavernier, P. Höfer, F. Engelke, C. Griesinger, and M.

Bennati.  $^1\text{H}$  and  $^{13}\text{C}$  Dynamic Nuclear Polarization in Aqueous Solution with a Two-Field (0.34 T/14 T) Shuttle DNP Spectrometer. *Journal of the American Chemical Society*, 131:15086-15087, 2009. (ref. 64)

- A. Krahn, P. Lottmann, T. Marquardsen, A. Tavernier, Türke M.-T., M. Reese, A. Leonov, M. Bennati, P. Höfer, F. Engelke, and C. Griesinger. Shuttle DNP Spectrometer with a Two-Center Magnet. *Physical Chemistry Chemical Physics*, 2010, submitted. (ref. 65)
- M.-T. Türke, I. Tkach, M. Reese, P. Höfer, and M. Bennati. Optimization of Dynamic Nuclear Polarization Experiments in Aqueous Solution at 15 MHz/9.7 GHz: A Comparative Study with DNP at 140 MHz/94 GHz. *Physical Chemistry Chemical Physics*, 2010, submitted. (ref. 66)

The shuttle-DNP project for the application of DNP to high-resolution liquid state NMR was designed as pilot study to examine the feasibility and limits of the concept. Experimental test setups were to be created and optimal design suggestions were to be made. The general requirement for our setup was to repeatedly and reproducibly polarize one and the same liquid, aqueous sample via DNP in a field lower than the final high-resolution detection field. Comparable and reproducible measurements should be allowed on the very same sample. This is important for high dimensional NMR spectroscopy, crucial to NMR in structural biology, and is not possible for other polarization techniques, like the approaches by Golman or Köckenberger. Furthermore, it should be possible to allow for proton enhancement and proton detection because this abundant nucleus has one of highest gyromagnetic ratios and so one of the highest intrinsic sensitivities.

## 7.1 Field Dependent Dynamic Nuclear Polarization with Radicals in Aqueous Solution, Peter Höfer et al., 2008

Before constructing a shuttle DNP spectrometer, in a first step, the magnetic field strength at which DNP should be performed was investigated. As the irradiation and saturation of the electron transitions is necessary to create DNP, several standard field strengths of EPR were considered. These include X-band ( $\nu_e=8-12$  GHz,  $\nu_H=12-18$  MHz,  $B_0=0.28-0.43$  T) and W-band ( $\nu_e=75-110$  GHz,  $\nu_H=114-167$  MHz,  $B_0=2.63-3.86$  T). From literature

only limited information was available and the experimentally observed enhancements were limited by the applied microwave field strength  $B_1$ . The values for infinite power had been extrapolated,<sup>67</sup> but never experimentally determined.

The first experiments were performed on commercial Bruker EPR spectrometers at 9.6 GHz (X-band) or 94 GHz (W-band).<sup>61</sup> For the X-band experiment a dielectric resonator was used in conjunction with a 5 W CW microwave amplifier, achieving a  $B_1$  field strength of up to 3 MHz (electron Rabi frequency). NMR was detected by a Bruker Minispec NMR spectrometer, using the externally tuned and matched NMR coil of the used ENDOR probe head. For the W-band experiment a cylindrical cavity was used as microwave resonator. A microwave power of 0.1 W could achieve a  $B_1$  field of 8 MHz (electron Rabi frequency). Here, the NMR signal was detected with a Bruker Avance console, using the tuned and matched NMR coil of the ENDOR probe head.

For the first time at X-band, non-extrapolated enhancement factors of approximately  $-100$  were observed for water protons at ambient conditions. The used radical was 4-hydroxy-TEMPO-<sup>14</sup>N (TEMPOL) at concentrations ranging between 5-10 mM. These samples achieved enhancement factors of up to  $-20$  in the W-band setup. Higher concentration led to reduced enhancement factors, due to the increased relaxation rates and a reduction of the saturation factor by insufficient microwave power.

Using the trityl radical TAM, only a factor 2.5 to 4 times smaller enhancement factors could be achieved using comparable concentrations. The reasons are the lower coupling and leakage factors for trityl. One reason for these low factors is the position of TAM's unpaired electron, lying in the center of the globular radical, isolating it from the solvent. The other is TAM's larger molecular mass and the resulting longer correlation time (TEMPOL: 15-20 ps,<sup>61</sup> TAM: 142 ps<sup>68</sup>).

Relaxation dispersion data, measuring the  $T_1$  relaxation time of radical doped water solutions over a large magnetic field range, were used to determine the coupling and leakage factors of the radicals at the given fields. From these and the observed enhancements the full-saturation (infinite microwave power) enhancement factors were obtained. For X-band this was about  $-220$  and for W-band about  $-27$ .

The observed enhancement factor of  $-100$  for X-band (0.34 T) translates to a theoretical post-shuttle enhancement factor of  $\epsilon_{\text{HF}} = -2.4 = -100 \cdot 0.34 \text{ T} / 14.09 \text{ T}$ , compared to the Boltzmann magnetization at 14.09 T, assuming no magnetization losses during sample transfer.

For W-band the same consideration for the observed enhancement factor

of  $-20$  leads to a post-shuttle enhancement factor of  $\epsilon_{\text{HF}} = -4.7 = -20 \cdot 2.2 \text{ T}/14.09 \text{ T}$ .

The seemingly larger possible post-shuttle enhancement factor for the W-band setup is compromised by the required small sample volume. The sample diameter for X-band was 1 mm and for W-band it was 0.15 mm; a total sample volume of 8000 nL or 45 nL was used, respectively. Using larger diameters would lead to significant heating.

In conclusion, it was shown that polarization at X-band and W-band could both provide enough magnetization to effectively enhance the signal of high-resolution NMR at 14.09 T.

## 7.2 Studies of Dynamic Nuclear Polarization with Nitroxides in Aqueous Solution, Peter Höfer et al., 2008

Further experiments were performed to systematically investigate the enhancement limits at X-band. Radical solutions in  $\text{H}_2\text{O}$  with 4-hydroxy-TEMPO- $^{14}\text{N}$  (TEMPOL), 4-oxo-TEMPO- $^{14}\text{N}$  (TEMPONE) and 4-oxo-TEMPO-D, $^{15}\text{N}$  (TEMPONE-D, $^{15}\text{N}$ ) were used at concentration of 5, 10 and 25 mM. Although similar experiments had been performed by several groups<sup>67</sup> we could measure the highest enhancements ever reported. The experimental setup was identical to the one described in the previous chapter, performing DNP at 0.34 T (9.7 GHz), using a standard ENDOR probe head for EPR irradiation and NMR detection. 5 W of microwave power could be applied. The sample diameter was 0.9 mm.

In the experiments the enhancement factor of the various radicals and concentrations have been measured as a function of microwave power. In general, the difference in enhancement for the unlabeled radicals TEMPOL and TEMPONE was small. For all solutions the enhancement increased by increasing the microwave power, as anticipated due to an increasing saturation factor. For low powers the increase in enhancement was still linear to the applied power. For larger powers the enhancement converged towards saturation. This was especially visible for the lowest used concentration of 5 mM. Here, enhancements of  $-100$  (close to the full-saturation maximum) have been observed for all radicals. This observation indicated that the enhancement was concentration limited.

In the case of the unlabeled  $^{14}\text{N}$ -radicals, it was not possible to observe larger enhancements with larger concentrations. This was due to the increased electronic relaxation rates at higher radical concentrations, induced

by increased inter-electronic Heisenberg spin exchange. This increase in relaxation rate was reducing the achievable electron saturation at the given microwave power. Heisenberg spin exchange is a inter-electronic relaxation process [69, and refs. there in], leading to cross-relaxation ("cross-talk") of the irradiated (resonant) and non irradiated (off-resonant) radical ensembles.

Nevertheless, for the isotope labeled compound TEMPONE-D,<sup>15</sup>N higher enhancement factors of  $-130$  and  $-140$  were possible, using higher concentrations of 10 mM and 25 mM, respectively.

The additional enhancement (after increasing the concentration) was still possible due to the fact that the <sup>15</sup>N-labeled (<sup>15</sup>N is spin 1/2) radical only shows two lines in the EPR spectrum, in contrast to three lines of the <sup>14</sup>N compound. Although Heisenberg spin exchange will lead to intermolecular cross-saturation of the other transitions, direct saturating is the dominating effect. Initially, only one line could be irradiated and better saturation was expected for the two line system (TEMPONE-D,<sup>15</sup>N). Furthermore, the substitution of protons against deuterons (being quadrupolar and having a 6.7 fold lower gyromagnetic ratio than protons) in the labeled radical led to longer electron life times, facilitating saturation with a given microwave power. The increased electron life time was confirmed by line width measurements using pulse EPR at a X-band pulse spectrometer [62, fig 1.].

For the solutions with higher concentration (10 mM and 25 mM) of TEMPONE-D,<sup>15</sup>N full saturation of the electrons was not achieved, as indicated by the fact that the enhancement factors do not reach saturation by increasing the power [62, Fig. 1c]. Nevertheless, the enhancement factor of  $-140$  for the 25 mM solution was the highest enhancement ever observed for water protons polarized by nitroxide radicals.

Further DNP experiments were undertaken to irradiate *both* EPR lines of TEMPONE-D,<sup>15</sup>N, being the first of their kind reported. Unfortunately, due to the power limit of the second amplifier (0.5 W max) and the low resonator quality on the resonance of the second line, no higher enhancements than with single line irradiation were achieved.

The observed low-field enhancement factor of  $-140$  for the 25 mM TEMPONE-D,<sup>15</sup>N aqueous solution translates to a theoretical high-field (post-shuttle) enhancement factor (relative to the 14.09 T Boltzmann polarization) of  $\epsilon_{\text{HF}} = \epsilon_{\text{LF}} \cdot 0.34 \text{ T} / 14.09 \text{ T} = -3.4$ , assuming no loss of magnetization during sample transfer.

The results indicated, together with the results of the the previous section,<sup>61</sup> that considerable DNP enhancements at 14.09 T could be created by polarizing at low field (0.34 T) and quickly transferring the sample to

the high-field magnet. Furthermore, it became evident that the highest enhancements can be gained using the  $^{15}\text{N}$ - and  $^2\text{H}$ -labeled TEMPONE. It had the advantage of longer relaxation times, facilitating higher saturation factors. Anyways, the available power of 5 W for microwave irradiation was not sufficient to obtain the maximum possible enhancement factors. More powerful amplifiers or alternatively higher conversion factor resonators were expected to provide higher  $B_1$  fields, saturation and enhancement factors.

### 7.3 Construction of a Liquid-State NMR DNP Shuttle DNP Spectrometer: First Experimental Results and Evaluation of Optimal Performance Characteristics, Marcel Reese et al., 2008

Based on our results in low-field DNP at 0.34 T and 3.3 T (ch. 7.1, 7.2)<sup>61,62</sup> a first prototype of a shuttle DNP spectrometer was constructed and initial experiments were performed.

The polarizing field was chosen to be 0.34 T (X-band). This was based on the fact that high enhancement factors of  $-140$  could be gained by X-band DNP (ch. 7.2)<sup>62</sup> and that the sample diameter (0.9 mm) and volume (10  $\mu\text{L}$ ) were large enough for simple sample handling and NMR detection. Furthermore, the EPR hardware at X-band (9.7 GHz) is relatively easy to handle and to obtain, compared to W-band (94 GHz).<sup>61</sup> For example, we could use a permanent magnet of only 80 kg mass to create the required polarization field of 0.34 T.

The first shuttle DNP setup [63, figs. 1,2] consisted of a 0.34 T permanent magnet with its center 1.5 m above the center of a 14.09 T superconducting magnet (figs. 12.1-12.4, p. 204 ff). While the field  $B_0^{\text{HF}}$  (high field) of the superconducting magnet was oriented vertically (along the  $z$ -axis) the field  $B_0^{\text{LF}}$  (low field) of the permanent magnet was oriented horizontally along a transversal axis. This choice of orientations facilitated the use of a standard ENDOR probe head. The high field magnet was equipped with an high-resolution Bruker NMR probe for  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^2\text{D}$  detection and pulsing. The high-resolution probe was manufactured with an end-to-end axial bore. The low-field magnet was used with a standard Bruker ENDOR probe head comprising a cylindrical dielectric resonator (fig. 12.3, p. 206). Both probes were connected to each other with a sample transfer tube from German silver through which the cylindrical sample container (o.d. of 4.1

mm) was shuttled pneumatically. Shuttling was actuated via piezoelectric proportional pressure valves controlled via a programmable, feedback-less time step controller. Irradiation of the EPR transitions as well as shuttling was triggered from the high-field NMR console, as determined in the pulse program. After DNP the sample was shot to the high-field magnet. To protect the high-resolution NMR probe from the shock of sample impact a translatable carbon fire shock transfer rod was inserted in the longitudinal bore of the NMR probe (fig. 12.4, p. 207). This rod was almost as light as the sample container. It was used to efficiently transfer the momentum and energy of the sample container to a viscoelastic dampening foam, placed below the high-resolution probe on a heavy support table. The position of the sample container in the sample transfer track could be measured using an array of 16 light barriers. One complete sample passage yielded 58 measurement of time-position coordinates, triggered by the interruption and clearance of the light beams by the opaque sample caps.

Although initial DNP experiments at low field gave reasonable DNP enhancements of about  $-60$  for 5 mM and of  $-130$  for 25 mM TEMPONE-D,<sup>15</sup>N solutions, these enhancements could not be translated to reproducible, experimental shuttle DNP enhancements. This was due to problems with sample sealing, leading to bubbles in the sample volume after shuttling.

The first published shuttle DNP experiment providing reliable and reproducible signal enhancement was performed on a 5 mM solution on TEMPONE-D,<sup>15</sup>N. For unknown reason the observed enhancement factor, directly measured at low field, was only  $-27$ . The sample was shuttled to the high-field magnet within 139 ms. An additional 111 ms stabilization period at high field was necessary before the application of the <sup>1</sup>H 90° pulse for reliable detection of the NMR signal. The high-field enhancement factor  $\epsilon_{\text{HF}} = \langle I_z \rangle / I_0 = -0.38$  was small but negative, indicating residual DNP. Compared to the initial magnetization at high field, 40% of the magnetization was lost during the sample transfer and the stabilization period.

A series of these experiments has been performed where the period between shuttle-down trigger and the start of proton detection was incremented from 150 ms to 6.5 s. During this time the residual magnetization decayed back to equilibrium, mainly in the high-field position, finally showing a positive signal with the 14.09 T Boltzmann amplitude [63, Fig. 5]. This recovery experiment was repeated using the identical pulse sequence, including shuttling, but without DNP. The resulting relaxation curves could be nicely fitted to an mono-exponential decay, gaining the relaxation rate of 0.875 /s. This compares well to the independently measured relaxation rate of 0.87 /s at 14.09 T using an equivalent sample [63, fig. 5].

The very same relaxation dispersion experiment, in conjunction with the

known sample trajectory, the experimentally determined magnetic field profile along the sample track, and the known initial enhancement factor of  $\epsilon_{\text{LF}}(t = 0 \text{ ms}) = -27$ , was used to simulate the relaxation losses during sample transfer [63, fig. 6] and the following stabilization period. The resulting post-shuttle enhancement factor of  $\epsilon_{\text{HF}}^{\text{calc}}(t = 250 \text{ ms}) = -0.36$  is in excellent agreement with the experimental value  $\epsilon_{\text{HF}}(t = 250 \text{ ms}) = -0.38$ .

Based on this validated model, the residual post-shuttle enhancement factor (with and without stabilization period) was simulated as a function of the initial DNP enhancement at low field. Taking the previously obtained pre-shuttle DNP enhancement factor of  $-100^{62}$  and by eliminating the stabilization period, a post-shuttle enhancement of  $\epsilon_{\text{LF}}^{\text{calc}} = -1.92$  was simulated [63, fig. 6].

In conclusion, it was demonstrated that reliable and reproducible DNP experiments could be performed with the existing setup. A valid simulation of the transfer characteristics including the trajectory details, post-shuttle delays and initial magnetization was established. It was shown that a post-shuttle DNP enhancement factor of approximately  $-2$  should be possible by solving technical difficulties. Nevertheless, no signal enhancement with  $|\epsilon_{\text{HF}}| \geq 1$  was achieved, yet.

## 7.4 $^1\text{H}$ and $^{13}\text{C}$ Dynamic Nuclear Polarization in Aqueous Solution with a Two-Field (0.34 T/14 T) Shuttle DNP Spectrometer, Marcel Reese et al., 2009

The previously described shuttle DNP setup was improved in several ways to allow for more reliable detection and for significant post shuttle signal enhancements. Experiments on  $^1\text{H}$  and  $^{13}\text{C}$  were performed.

The sample was restricted in length in the sample capillary using plugs prepared in-situ from UV glue (fig. 12.5, p. 207). This had two advantages in respect to the previous setup, not restricting the sample: The restricted length (3 mm for proton DNP, 6 mm for  $^{13}\text{C}$  DNP) assured that the sample could be placed entirely in the homogeneous  $B_1$  field region of the microwave cavity, which led to homogeneous DNP, homogeneous heating and thus homogeneous line broadening. Furthermore, the leakage problem was reliably solved. Now a 40 W Varian microwave amplifier was used, effectively providing 20 W of power at 9.7 GHz to the ENDOR probe head. For experiments on  $^{13}\text{C}$ , using MW irradiation times of up to 20 s, a reduced sample diameter of 0.46 mm was used to reduce the heating to no more than

12 °C. To reduce the post-shuttle stabilization period the lower shuttle valve (used to shuttle the sample to the low-field position via over pressure) was replaced against a valve providing over or under pressure. Before and after sample arrival the sample was sucked against the shock transfer rod, reducing rebound and vibrations. An adjustable copper spring system was implemented on the bottom of the high-resolution probe head to assure a quick and reproducible return of the shock transfer rod into its pre-impact position before NMR-detection.

For water, doped with 5mM TEMPONE-D,<sup>15</sup>N, a low-field DNP enhancement of -110 was measured in-situ, using the ENDOR probe and the Bruker Minispec. The microwave irradiation time was 3 s and the microwave power was 5 W. The sample was restricted to 3 mm length in a 0.9 mm ID tube. Using this sample a maximum high-field enhancement factor  $\epsilon_{\text{HF}} = -2.6$  was measured on the integral over the water line (fig. 12.7, p. 209). The amplitude enhancement factor ranged up to  $\epsilon_{\text{HF,A}} = -2.05$ . The integral enhancement factor of  $-2.6$  indicated that, in the error bounds of the experiment, no water magnetization was lost during sample transfer.

The water chemical shift is very sensitive to heating effects. Its chemical shift temperature coefficient is 0.01 ppm/°C. This effect still led to a inhomogeneous broadening of the post-DNP NMR line. It was seen that the line shape is narrowing down during post-shuttle cooling, initially leading to *higher* amplitude enhancements, despite of reduced integral enhancements [63, Fig. 1]. The water chemical shift was used to measure the maximum heating of the sample to 12 °C.

<sup>13</sup>C DNP enhancements on chlorophorm-<sup>13</sup>C were performed using 25 mM TEMPONE-D,<sup>15</sup>N. The sample was restricted to 6 mm length in a 0.9 mm ID tube. The sample was irradiated for 15 s with a microwave power of 20 W. High-field enhancement factors of up to +15 were observed, indicating predominant scalar cross-relaxation. This translates to a low-field enhancement factor of at least  $\epsilon_{\text{LF}} = +15 \cdot 14.09 \text{ T}/0.34 \text{ T} = +630$ . No <sup>13</sup>C-line broadening due to heating was observed.

A similar experiment was performed with an aqueous solution of 4 M urea-<sup>13</sup>C with 25 mM TEMPONE-D,<sup>15</sup>N (see also fig. 12.6, p. 208). The sample was restricted to 6 mm length in a 0.46 mm ID tube. A maximum high-field enhancement factor of  $\epsilon_{\text{HF}} = -4 \pm 1$  was observed, translating to a low-field enhancement factor of  $\epsilon_{\text{LF}} = -168 \pm 42$ . The sign of the enhancement factor was unambiguously measured to be negative, indicating a predominantly dipolar polarization mechanism. To probe for indirect "three-spin" polarization of <sup>13</sup>C by surrounding water protons, which get directly polarized as well and gain a negative polarization, a series of solutions containing 25%, 50%, 75% or 100% D<sub>2</sub>O was performed. No systematics influence on the

size of the  $^{13}\text{C}$  enhancement was found, in agreement with previous data for large radical concentration.<sup>70</sup>

Based on  $^{13}\text{C}$  relaxation dispersion measurements on an aqueous 4M urea- $^{13}\text{C}$  solution containing 25 mM TEMPONE-D,  $^{15}\text{N}$  [64, fig. S4] a simulation of the relaxation losses of  $^{13}\text{C}$  polarization during sample transfer was performed, analogously to the previous simulation for water protons (ch. 7.3). Based on this simulation only 7% of the  $^{13}\text{C}$  magnetization of urea gets lost during sample transfer. Considering this result the pre-shuttle enhancement factor was  $-181 \pm 42$ .

Initial, unpublished simulations had already indicated that the presented setup, using two dedicated magnets for low-field polarization and high-field detection, could not be used for high-field DNP of larger biomolecules. These molecules suffer from severe relaxation during sample transfer due to rates of  $\geq 10^4/\text{s}$  in the lowest field in the transfer track [64, fig. S1]. The simulations were based on a collective proton relaxation model<sup>71</sup> and the known field in the sample transfer track as well as the known path-time diagram of sample transfer.

Similar simulations, based on the previously mentioned collective relaxation model, on shuttle field cycling relaxometry data of ubiquitin and of glucose (ch. 11, p. 198) indicated that a new setup, using a stray field position of the high-field magnet for DNP, could be used for high-field DNP of biomolecules. Here, the sample transfer distance was reduced from 1.5 m to 0.47 m and the magnetic field in the transfer track was always bigger than 0.34 T (previously 5 mT) [64, Fig. S1].

## 7.5 Shuttle DNP Spectrometer with a Two-Center Magnet, Alexander Krahn et al., 2010, submitted

A new liquid state DNP setup was built for polarizing a liquid sample at 0.34 T and detection of NMR at 14.09 T. Instead of using two dedicated magnets, as in a previous setup, a ferro-shim system was inserted into the bore of the 14.09 T magnet (fig. 12.8, p. 209). It created a homogeneous region in the stray field, suitable for DNP at 0.34 T. The homogeneous low-field region was 470 mm above the center of the NMR magnet. The ferro shim system was equipped with a sweep coil and a gradient coil for adjusting the value of  $B_0^{\text{LF}}$  and to compensate for residual field gradients. The magnetic field strength in between both homogeneous regions was monotonously increasing from 0.34 T to 14.09 T, and thus much higher than for the previous setup.

There, the field dropped to approximately 5 mT.

The newly developed low-field probe head (fig. 12.10, p. 211) was comprising a new microwave resonating structure with a  $TM_{110}$  mode. This development was necessary as the magnetic field direction at the low-field and high-field position were identical, in contrast to the previous setup where the fields were perpendicular to each other. Conventional EPR probes require the  $B_0$  field to be perpendicular to the sample tube and EPR probe head, and thus a new EPR resonator had to be developed.

Due to the reduced shuttle distance of only 470 mm (previously 1500 mm), the sample transfer time using nitrogen gas could be reduced to 40 ms, one third of the previous time. The indicated pressure for sample transfer was 4 bar.

The sample tubes had an inner diameter of either 0.76 mm or 1 mm. The sample was restricted in length using vespel inserts (for initial low-field experiments) or using UV glue (for the shuttle DNP experiments).

For initial DNP experiments at low field an additional NMR coil was integrated into the EPR cavity. These experiments were performed to test and optimize the resonator structure and for optimal placing of the ferro-shim system. For later shuttle DNP experiments this NMR coil was removed to increase the Q-factor of the microwave resonator and to reduce microwave field distortions, responsible for heating of the aqueous sample.

The first results with shuttle DNP show significant post shuttle enhancement factors for protons. The contained 5 mM TEMPONE-D, $^{15}N$  in a solution of 80%  $H_2O$  and 20%  $D_2O$  and 10 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) and 100 mM ethanol. It was restricted to 11 mm length. The sample transfer time was 37 ms. An additional post-shuttle delay, used as stabilization period, was necessary before reliable detection. The applied power was 15 W. After approximately 8 s the maximum enhancement for water, ethanol and DSS was reached. All enhancements were negative, implying a dominant dipolar relaxation mechanism. The observed maximum enhancement factor  $\epsilon_{HF}$  of ethanol was  $-1.7$ , for DSS  $-1.2$  and for water  $-3.2$  [65, fig. 7, tab. 1].

The relaxation rates of all resonances of water, ethanol and DSS have been measured at low field and high field. These relaxation rates provided upper and lower relaxation rate bounds for calculating the relaxation losses during sample transfer. Using the measured transfer time of 37 ms and considering the post-shuttle delay of 70 ms, the losses for water were calculated to 7-10%, for ethanol to 7-9% and for DSS to 13-15%.

For the first time, proton signal enhancement for glucose with  $|\epsilon_{HF}| \geq 1$  was observed. A 10 mM TEMPONE-D, $^{15}N$  solution in  $D_2O$  with 5 mM

DSS and 500 mM D-glucose was used. All water, DSS and glucose signals show negative enhancements [65, Fig. 8], indicating a dominant dipolar enhancement mechanism. For glucose resonances the enhancements were in a range of  $\epsilon_{\text{HF}} = -1.4$  to  $-2.8$  [65, fig. 9, tab. 3]. For water the highest enhancement was  $-3.7$  and for DSS  $-2.0$ . For water this is the highest post-shuttle enhancement factor observed so far and corresponds to a low-field enhancement factor of  $-158$ . Using experimental upper and lower relaxation rate bounds as before, the losses for glucose were quantified to 10-30% of the pre-shuttle magnetization.

In conclusion, it was shown that shuttle DNP for signal enhancement of liquid state NMR is applicable to small molecules. For glucose signal enhancement factors of up to  $\epsilon_{\text{HF}} = -2.8$  were observed. If repeated signals are to be summed for increasing the signal to noise ratio, as conventionally done in NMR, the reduction in measurement time using this enhancement would be a factor of  $7.8 = 2.8^2$ . It remains to be seen if the signals of biological macromolecules can be effectively enhanced.

## 7.6 Optimization of Dynamic Nuclear Polarization Experiments in Aqueous Solution at 15 MHz/9.7 GHz: A Comparative Study with DNP at 140 MHz/94 GHz, Maria-Teresa Türke et al., 2010, submitted

Further, detailed characterizations of low-field DNP at 0.34 T and 3.4 T were performed. Parameters having influence on the observed signals were analyzed and, if required, were taken under control. This is including the sample temperature, the resonator temperature, magnetic field drift, the sample diameter, the sample length, linearity of the detection system and the analysis of quantification methods for the measured DNP enhancements. The highest ever reported proton DNP enhancement of  $-170$  for an aqueous sample under ambient conditions was observed. The highest enhancement factor observed at W-band was  $-43$ .

We had been criticized that the previously observed high coupling factors and DNP enhancements of  $-100$  and  $-140$  were originating from excessive sample heating under DNP conditions, increasing the coupling factor.<sup>72</sup> As

the coupling factors had been determined independently from NMR relaxation dispersion experiments under temperature control, this seemed unlikely to us.

One possible problem in DNP, using up to 34 W of microwave power, is the heating of the sample or resonator. To check for the latter problem a novel setup was realized to allow for high power wobbling and matching of the EPR resonator. In previous experiments the EPR probe was matched at low power. Then the high-power amplifier was inserted between bridge and probe. This prohibited the detection of the reflection coefficient of the probe as the amplifier is a one-way device. To allow the reflected power to reach the EPR bridge a back-route was implemented, using two circulators and one attenuator. Now, a signal from the bridge was passing through the a first low-power circulator, was amplified by the amplifier, passed the second, high-power circulator and then was partially reflected at the probe. The reflected signal was passing the high-power circulator, routed though the attenuator and, after passing the low power circulator, was detected in the EPR bridge [66, fig. 1].

Using this setup under DNP conditions, irradiating with 34 W, showed that the matching got worse on a time scale of a few seconds. The absorption dip was shifting in frequency and lost the matching condition. This behavior was anticipated to diminish the maximum DNP enhancement factor, as less power would enter the resonator. The problem could be solved by cooling the cavity using 10 l/min nitrogen gas, blown into the cavity. This gas flow was used in all experiments of the paper summarized here.

One possible source of error for the measurement of enhancement factors is a slow field drift of the  $B_0$  field. This can affect the experiment in two ways. Firstly, the EPR line might move too much to be irradiated effectively. This is possible as the resonances of the radicals used here are narrow (0.5-2 G). A field drift of this size might actually occur on EPR electromagnets, as they are normally not equipped with a field-frequency lock system. Complete saturation might become impossible under off-resonance irradiation, leading to too low DNP enhancements. More severely, too high enhancement factors might be created artificially by the field drift affecting the acquisition of the Boltzmann signal. Normally, all DNP experiments can be done in a very short time scale. 8 experiments, needed for a full NMR phase cycle, are enough to gain a very satisfactory signal to noise. In this time no significant field drift will occur. The detection of the Boltzmann signal, in the contrary, requires about 1000 scans to be available with a good signal to noise. During the required time of up to one hour a field drift can diminish the NMR signal, as the different free induction decays do not add up coherently. When a signal with this influence is used as a (too small)

Boltzmann reference, the resulting DNP enhancement factor would appear too large. If a coupling factor is deduced from it, it would appear too large as well. To avoid this effect, a field-frequency lock (FF-lock) was implemented in the used spectrometer.

The effect of sample heating was evaluated using two methods. For the larger samples (with an inner diameter of 0.9 mm) it was possible to measure the temperature directly under DNP conditions. For this, a fiber optical temperature sensor was used. The measurements indicated that the temperature raise in the sample could not be reduced by using the nitrogen gas flow. For the smaller samples this temperature sensor could not be used. Here the temperature increase was measured from the applied power, the heat capacity of the sample and the change of Q-factor under sample insertion.<sup>73</sup> Applying this method to the 0.9 mm i.d. samples gave very similar results to direct temperature measurements. The indirect experiments gave evidence that samples of an inner diameter of 0.45 mm do not show detectable heating under DNP conditions [66, fig. 5]. The sample showing the enhancement factor of  $-170$  had a inner diameter of 0.45 mm. Thus, heating as the origin for the large enhancement was excluded.

Often DNP enhancements get published as a function of the applied microwave power. This is relevant as more power will lead to a higher  $B_1$  field, leading to higher saturation and by this to higher DNP enhancement. Nevertheless, different setup will have different conversion factors  $c_c$  creating different  $B_1$  fields from the same microwave power  $P_{MW}$ :  $(B_1)^2 = (c_c)^2 P_{MW}$ . Using the trityl radical TAM, showing very long relaxation times, it was possible to measure the  $B_1$  field strength in the resonator under critically coupled DNP conditions, providing an instrument-independent measure for the microwave amplitude.

For all used samples, with radical concentration in the range of 5 mM to 50 mM, power-dependent enhancement curves were acquired. All of these enhancement curves reached a plateau, meaning that the enhancement did not increase any further with increasing power. This is due to the fact that the irradiated transition was completely saturated and by this the saturation factor had reached its maximum. Further gain of enhancement would only be possible by irradiating the other transition of the radical independently.

This conclusion was supported by independent pulsed ELDOR (electron-electron double resonance) experiments. Here, in general, the spectrum will first be irradiated (saturated) in a given frequency, then a detection pulse and detection at a different frequency will follow.

In the experiment performed here, the low-field transition of TEMPONE-D,<sup>15</sup>N was chosen as detection frequency. The saturation frequency pulse of 100 ns duration and  $B_1 = 2.5$  G was successively swept over the com-

plete spectrum. If both frequencies fell together no signal was detected, indicating full saturation of the directly irradiated transition. If the other resonance was irradiated the recorded resonance was decreased to 62% or 55%, for radical concentration of 5 mM or 10 mM, respectively. This means that under complete saturation of one line, the other line is only partially saturated, mediated by cross relaxation effects like inter-radical Heisenberg spin exchange. The calculated total saturation factors were 0.81 or 0.77, for the 5 mM or 10 mM concentration, respectively. Higher enhancement can only be achieved by irradiating and saturating both lines simultaneously.

The previously observed DNP enhancement factors of  $-100$  or  $-140$  could be reproduced (using a 0.9 mm ID tube) and rationalized with only a  $10$  °C increase of sample temperature compared to the 0.45 mm ID tubes. The highest enhancement factor ever reported using radicals in solution at X-band was observed, without heating. Systematic studies on experimental parameters provided a reliable protocol for accurately measuring DNP enhancement factors. The coupling parameters and saturation factors for the given systems and experimental conditions were determined, in agreement with our previously reported coupling factors.<sup>61,62</sup> Double irradiation was unambiguously identified as option for higher DNP enhancements. For 5 mM or 10 mM TEMPONE-D,<sup>15</sup>N solutions the observed maximum enhancement factors of approximately  $\epsilon_{\text{LF}} = -150$  and  $\epsilon_{\text{LF}} = -160$  can still be increased by 23% or 30% to  $\epsilon_{\text{LF}}^{\text{max}} = -185$  or  $\epsilon_{\text{LF}}^{\text{max}} = -208$ , respectively. This would translate to post-shuttle DNP enhancement factors of  $\epsilon_{\text{HF}} = -4.4$  or  $\epsilon_{\text{HF}} = -4.9$ .



# Chapter 8

## Conclusion and Outlook

In the presented projects, the applicability of dynamic nuclear polarization (DNP) to liquid, aqueous samples for signal enhancement of high-resolution NMR in structural biology was investigated. Here, DNP at low temperatures, requiring repeated freezing and thawing of the sample<sup>56</sup> or sample dissolution using hot solvents,<sup>53</sup> should be avoided. Instead, direct polarization in the liquid state at ambient temperature was intended. The repeated use of one and the same sample for repeated measurements, as required for most high-dimensional NMR spectra, was a prerequisite.

High-resolution NMR for structural biology requires high magnetic fields ( $\geq 9$  T) to achieve an acceptable chemical shift resolution and sensitivity. The latter is due to the higher Boltzmann magnetization and the higher Larmor frequency at higher fields. Nevertheless, the highest DNP enhancements, especially for larger target molecules (solute), were expected to be found at much lower magnetic fields.<sup>60</sup> Furthermore, to avoid microwave induced heating, DNP with aqueous samples at high magnetic fields would have required extremely small sample diameters and thus very small sample volumes of only a few nanoliters.<sup>59</sup> Thus, polarization at low field, allowing for big sample volume and high enhancement factors, and NMR detection at high field, giving a good chemical shift resolution and a high Larmor frequency for sensitive NMR detection, was intended.

In the summarized studies initial low-field DNP experiments at the fields of 0.34 T (X-band) and 3.4 T (W-band) were performed.<sup>61,62</sup> These experiments led to the highest so far observed enhancement factors for the protons of water of  $\epsilon_{\text{LF}} = -100$ , using non-isotope-labeled trityl radicals, and of  $\epsilon_{\text{LF}} = -140$ , using labeled radicals (p. 125). Considering polarization at 0.34 T and detection at 14.09 T, these low-field enhancement factors would be translated to high-field enhancement factors of  $\epsilon_{\text{HF}} = -2.4$  and  $\epsilon_{\text{HF}} = -3.3$ , respectively. Although higher magnetization was observed in

W-band DNP<sup>61</sup> the larger sample diameter in X-band and practicability reasons led us choose the X-band field for initial shuttle DNP experiments.

Motivated by the described low-field DNP results, a shuttle DNP setup was built comprising a 14.09 T superconducting high-field magnet for high resolution NMR detection and a 0.34 T low-field permanent magnet for DNP polarization (p. 126).<sup>63</sup> The low-field magnet was placed on top of the high-field magnet. The magnetic centers were 1.5 m separated from each other. Initially, only small high-field DNP enhancement factors for water protons of  $\epsilon_{\text{HF}} = -0.38$  could be observed. Improvements for this dual magnet setup achieved high-field enhancement factors for water protons of up to  $\epsilon_{\text{HF}} = -2.6$  (p. 128).<sup>64</sup> The high-field enhancement factors of  $^{13}\text{C}$  in chloroform- $^{13}\text{C}$  and for 4M urea- $^{13}\text{C}$  in water were  $\epsilon_{\text{HF}} = +15$  or  $\epsilon_{\text{HF}} = -4 \pm 1$ , respectively. Due to relaxational losses during sample transfer, larger molecules could not be effectively polarized ( $|\epsilon_{\text{HF}}| \geq 1$ ) for high-resolution detection. These losses were due to the low magnetic fields ( $\geq 5$  mT) and the resulting high relaxation rates experienced by the sample during sample transfer. Simulations indicated that avoiding the low-field regime during sample transfer would significantly reduce relaxation losses for effective signal enhancement of larger solutes like glucose or ubiquitin.

A new setup was built where a ferro shim system was inserted into the high-field NMR magnet (p. 130).<sup>65</sup> Now a stray field position in the high-resolution magnet was used for DNP. The sample transfer distance was reduced to 0.47 m and the transfer time to only 40 ms, one third of the time of the previous setup. The magnetic field on the transfer path was higher than 0.34 T at all sample positions during sample transfer. Now, small solutes were effectively polarized and the observed maximum enhancement factors were  $\epsilon_{\text{HF}} = -1.7$  for ethanol,  $\epsilon_{\text{HF}} = -1.2$  for DSS and  $\epsilon_{\text{HF}} = -3.2$  for water. For the first time with shuttle DNP, high-field proton signal enhancements for glucose with  $|\epsilon_{\text{HF}}| \geq 1$  were observed. For different resonances they were in a range of  $\epsilon_{\text{HF}} = -1.4$  to  $-2.8$ .

The conditions influencing the observed low-field DNP enhancement factors were analyzed (p. 132).<sup>66</sup> This is including the sample temperature, the resonator temperature, magnetic field drift, the sample diameter, the sample length, linearity of the detection system and the analysis of quantification methods for the measured DNP enhancements. If required, these parameters were taken under control to avoid misinterpretation of the data. The highest ever reported proton DNP enhancement factor of  $\epsilon_{\text{LF}} = -170$  for the protons of water under ambient conditions was observed. The highest enhancement factor observed for water protons at W-band was  $\epsilon_{\text{LF}} = -43$ . The coupling parameters and saturation factors for the used systems were determined, in agreement with our previously reported coupling factors.<sup>61,62</sup>

Double irradiation was unambiguously identified as option for higher DNP enhancements. Using double irradiation on the characterized 5 mM or 10 mM TEMPONE-D,<sup>15</sup>N solutions the observed maximum single irradiation enhancement factors of approximately  $\epsilon_{\text{LF}} = -150$  and  $\epsilon_{\text{LF}} = -160$  can still be increased by 23% or 30% to  $\epsilon_{\text{LF}}^{\text{max}} = -185$  or  $\epsilon_{\text{LF}}^{\text{max}} = -208$ , respectively. Neglecting relaxation losses during sample transfer, this would translate to post-shuttle DNP enhancement factors of  $\epsilon_{\text{HF}} = -4.4$  or  $\epsilon_{\text{HF}} = -4.9$ .

Further experiments using the present shuttle DNP setup (comprising the ferro-shim system, high field on the sample track and short transfer times) will involve samples containing small globular proteins and unfolded proteins like  $\alpha$ -synuclein. The applicability of isotope-labeling schemes involving deuteration with specific protonation of methyl groups will be performed. Generally, due to the high dynamics of methyl groups their protons show long relaxation times and will experience low relaxation losses during sample transfer. Deuterating non-methyl protons will increase the methyl proton life times further. It is still an open question if large molecules can be polarized by low-field DNP at all. No experimental evidence for this is reported so far. It remains to be seen if the achievable enhancement factors and the relaxation losses during sample transfer will allow for effective post-shuttle enhancement of protein signals.

In general, the detection of the small amount of protein in the restricted sample volume of no more than 8  $\mu\text{L}$  might be problematic using the present high-resolution NMR probe. Currently, the NMR-coil diameter of  $d = 5$  mm is in a suboptimal ratio to the sample diameter of no more than 1 mm. The coil diameter can only be reduced significantly by avoiding the use of a sample container (o.d.  $\approx 3.2$  mm). As smaller samples can probably not be shuttled pneumatically, a electric-motor-driven mechanical shuttle device is currently under construction at the MPIBPC, based on the Redfield design in ref. [74] and on personal communication and support by Prof. Redfield. Here, the sample capillary will be mounted to the end of a carbon fiber rod. This setup should allow the use of an NMR coil diameter of about  $d = 2$  mm. The  $B_1$  field generated by a coil (and reciprocally the voltage induced in the coil by a sample) scales as  $B_1 \propto 1/d$ . Compared to the present setup ( $d = 5$  mm), roughly estimated, the induction signal should be 2.5 times higher with a  $d = 2$  mm coil. This corresponds to a reduction of measurement time by a factor of  $6.25 = 2.5^2$  if signal accumulation is performed. The smallest used sample diameter was  $d_{\text{min}} = 0.45$  mm. If the diameter of the coil and the sample capillary could be adapted to this value, allowing to use a  $d = 1$  mm coil, the induction signal could be increased by the factor 5, resulting in a reduction of measurement time by a factor of 25.

If considerable high-field enhancement factors can be observed for proteins

and if the NMR coil diameter can be reduced to match the diameter of the sample, effective sensitivity gain for high-resolution NMR in structural biology could be accomplished using shuttle DNP.

Appendix I  
INPHARMA



# Chapter 9

## Theory

### 9.1 Simulation of INPHARMA Peak Volumes

#### 9.1.1 Setup of the Kinetic Matrix

The entries of the kinetic matrix (2.8, p. 17) are derived from the chemical rate constants of the equation:



Writing all the rates with their dependence of the direction of reaction yield the following kinetic equations:

$$\begin{aligned} \left(\frac{d[\text{B}]}{dt}\right)_{12} &= \left(\frac{d[\text{TA}]}{dt}\right)_{12} = -k_{12}[\text{TA}][\text{B}] = -\left(\frac{d[\text{TB}]}{dt}\right)_{12} = -\left(\frac{d[\text{A}]}{dt}\right)_{12} \\ \left(\frac{d[\text{A}]}{dt}\right)_{21} &= \left(\frac{d[\text{TB}]}{dt}\right)_{21} = -k_{21}[\text{TB}][\text{A}] = -\left(\frac{d[\text{B}]}{dt}\right)_{21} = -\left(\frac{d[\text{TA}]}{dt}\right)_{21} \end{aligned} \quad (9.2)$$

Here, I chose single of them:

$$\left(\frac{d[\text{B}]}{dt}\right)_{12} = -k_{12}[\text{AT}][\text{B}] \quad (9.3)$$

The rate constant of this reaction is  $-k_{12}[\text{AT}]$ . This values could almost be used as an diagonal entry of the second submatrix  $\mathbf{K}^{\text{B}}$  of the kinetic matrix. As the differential equation (2.1) possesses the minus sign, the value  $k_{12}[\text{AT}]$

will be the final entry. For all diagonal elements of the kinetic matrix this procedure is applied using the respective, relevant kinetic equation (the ones on the left hand side of (9.2)).

As chemical exchange will not destroy magnetization (which would be the case by only using diagonal entries in the kinetic matrix  $\mathbf{K}$ ) the magnetization has to be transferred to the respective binding destination of the species. This is done by using the off-diagonal elements of the kinetic matrix. These rates are derived from the right hand side of formula (9.2). Now, the sum of the rate constants in each column of  $\mathbf{K}$  is zero, indicating that the magnetization is preserved under chemical exchange.

## 9.2 Derivation: INPHARMA Energy and Gradient

### 9.2.1 Lemma 1

The following lemma is used in the main text (2.23) for determining a single summand of the power series (2.22):

$$\begin{aligned}
\frac{\partial(\mathbf{R}^N)_{kl}}{\partial\mathbf{R}_{mn}} &= \frac{\partial}{\partial\mathbf{R}_{mn}} \left( \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{k\alpha_1} \cdot \mathbf{R}_{\alpha_1\alpha_2} \cdot \mathbf{R}_{\alpha_2\alpha_3} \cdot \dots \cdot \mathbf{R}_{\alpha_{N-1}l} \right) \\
&= \sum_{\alpha_1 \dots \alpha_{N-1}} \frac{\partial\mathbf{R}_{k\alpha_1}}{\partial\mathbf{R}_{mn}} \cdot \mathbf{R}_{\alpha_1\alpha_2} \cdot \dots \cdot \mathbf{R}_{\alpha_{N-1}l} \\
&\quad + \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{k\alpha_1} \cdot \frac{\partial\mathbf{R}_{\alpha_1\alpha_2}}{\partial\mathbf{R}_{mn}} \cdot \mathbf{R}_{\alpha_2\alpha_3} \cdot \dots \cdot \mathbf{R}_{\alpha_{N-1}l} \\
&\quad + \quad \vdots \\
&\quad + \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{k\alpha_1} \cdot \mathbf{R}_{\alpha_1\alpha_2} \cdot \mathbf{R}_{\alpha_2\alpha_3} \cdot \dots \cdot \frac{\partial\mathbf{R}_{\alpha_{N-1}l}}{\partial\mathbf{R}_{mn}}
\end{aligned} \tag{9.4}$$

The final steps yield

$$\begin{aligned}
\frac{\partial(\mathbf{R}^N)_{kl}}{\partial \mathbf{R}_{mn}} &= \sum_{\alpha_1 \dots \alpha_{N-1}} \frac{\partial \mathbf{R}_{k\alpha_1}}{\partial \mathbf{R}_{mn}} \cdot \mathbf{R}_{\alpha_1\alpha_2} \cdots \mathbf{R}_{\alpha_2\alpha_3} \cdots \mathbf{R}_{\alpha_{N-1}l} \\
&+ \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{k\alpha_1} \cdot \frac{\partial \mathbf{R}_{\alpha_1\alpha_2}}{\partial \mathbf{R}_{mn}} \cdot \mathbf{R}_{\alpha_2\alpha_3} \cdots \mathbf{R}_{\alpha_{N-1}l} \\
&+ \quad \vdots \\
&+ \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{k\alpha_1} \cdot \mathbf{R}_{\alpha_1\alpha_2} \cdot \mathbf{R}_{\alpha_2\alpha_3} \cdots \frac{\partial \mathbf{R}_{\alpha_{N-1}l}}{\partial \mathbf{R}_{mn}} \\
&= \delta_{mk} \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{n\alpha_2} \cdot \mathbf{R}_{\alpha_2\alpha_3} \cdots \mathbf{R}_{\alpha_{N-1}l} \\
&+ \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{km} \mathbf{R}_{n\alpha_3} \cdots \mathbf{R}_{\alpha_{N-1}l} \\
&+ \quad \vdots \\
&+ \delta_{nl} \sum_{\alpha_1 \dots \alpha_{N-1}} \mathbf{R}_{k\alpha_1} \cdot \mathbf{R}_{\alpha_1\alpha_2} \cdot \mathbf{R}_{\alpha_2\alpha_3} \cdots \mathbf{R}_{\alpha_{N-2}m} \\
&= \delta_{mk} (\mathbf{R}^{N-1})_{nl} \\
&+ \sum_{f=1}^{N-2} (\mathbf{R}^f)_{km} (\mathbf{R}^{N-1-f})_{nl} \\
&+ \delta_{nl} (\mathbf{R}^{N-1})_{km}.
\end{aligned} \tag{9.5}$$

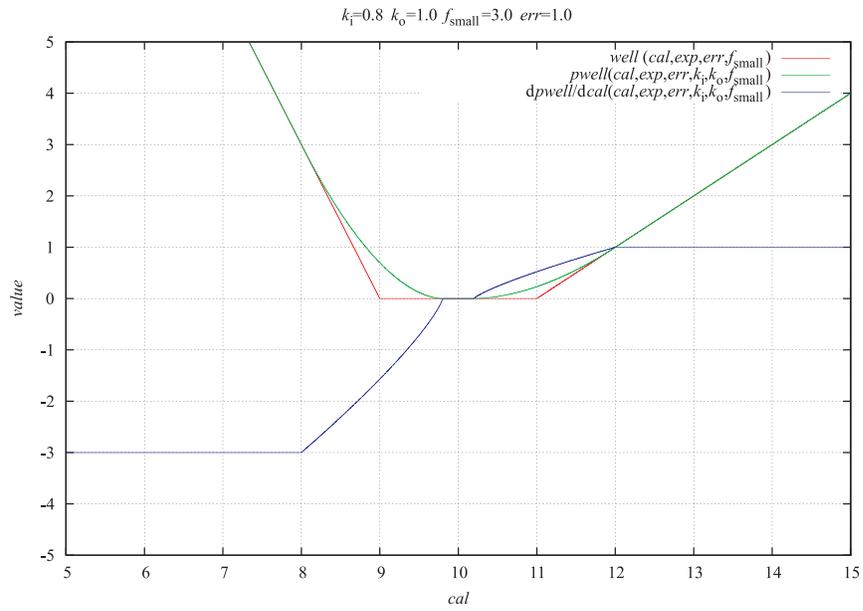
### 9.2.2 Lemma 2

The following lemma is used in the main text (2.24) for determining a single summand of the power series (2.22). Its derivation uses the result of the previous lemma (9.5). For the diagonal matrices  $\mathbf{C}$  and  $\mathbf{A}^0$  the definitions  $\mathbf{C}_k \equiv \mathbf{C}_{kk}$  and  $\mathbf{A}_l^0 \equiv \mathbf{A}_{ll}^0$  are used.

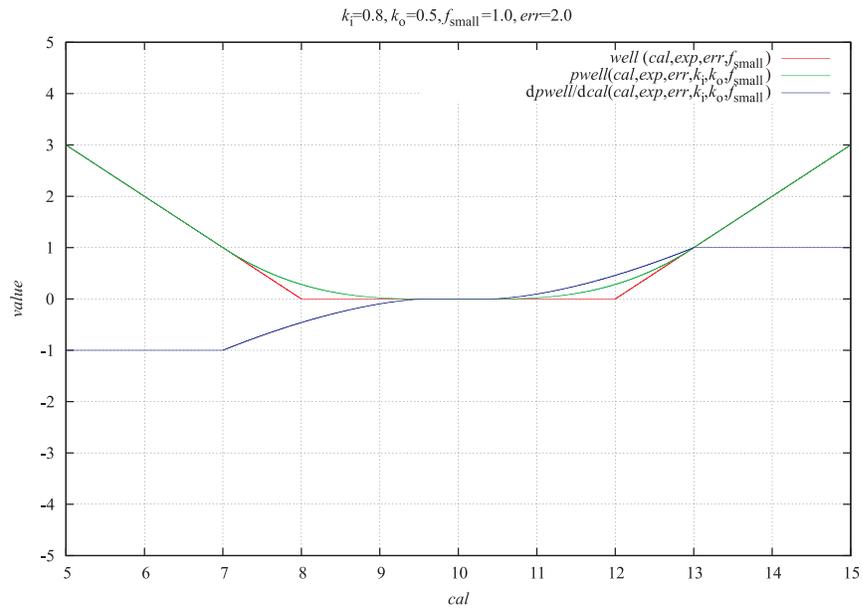
$$\begin{aligned}
\sum_{kl} \mathbf{C}_k \mathbf{F}_{kl} \frac{\partial (\mathbf{R}^N)_{kl}}{\partial \mathbf{R}_{mn}} \mathbf{C}_l^{-1} \mathbf{A}_l^0 &= \sum_{kl} \mathbf{C}_k \mathbf{F}_{kl} \left\{ \delta_{km} (\mathbf{R}^{N-1})_{nl} \right. \\
&\quad + \sum_{f=1}^{N-2} (\mathbf{R}^f)_{km} (\mathbf{R}^{N-1-f})_{nl} \\
&\quad \left. + \delta_{nl} (\mathbf{R}^{N-1})_{km} \right\} \mathbf{C}_l^{-1} \mathbf{A}_l^0 \\
&= \sum_l \mathbf{C}_m \mathbf{F}_{ml} \mathbf{A}_l^0 \mathbf{C}_l^{-1} (\mathbf{R}^{N-1})_{ln} \\
&\quad + \sum_{kl} \sum_{f=1}^{N-2} (\mathbf{R}^f)_{mk} \mathbf{C}_k \mathbf{F}_{kl} \mathbf{A}_l^0 \mathbf{C}_l^{-1} (\mathbf{R}^{N-1-f})_{ln} \\
&\quad + \sum_k (\mathbf{R}^{N-1})_{mk} \mathbf{C}_k \mathbf{F}_{kn} \mathbf{A}_n^0 \mathbf{C}_n^{-1} \\
&= (\mathbf{CFA}^0 \mathbf{C}^{-1} \mathbf{R}^{N-1})_{mn} \\
&\quad + \left( \sum_{f=1}^{N-2} \mathbf{R}^f \mathbf{CFA}^0 \mathbf{C}^{-1} \mathbf{R}^{N-1-f} \right)_{mn} \\
&\quad + (\mathbf{R}^{N-1} \mathbf{CFA}^0 \mathbf{C}^{-1})_{mn} \\
&= \left( \sum_{f=1}^N \mathbf{R}^{f-1} \mathbf{CFA}^0 \mathbf{C}^{-1} \mathbf{R}^{N-f} \right)_{mn} \\
&= \left( \sum_{f=1}^N \mathbf{R}^{f-1} \mathbf{F}^A \mathbf{R}^{N-f} \right)_{mn}
\end{aligned} \tag{9.6}$$

## 9.3 Alternative Penalty Functions

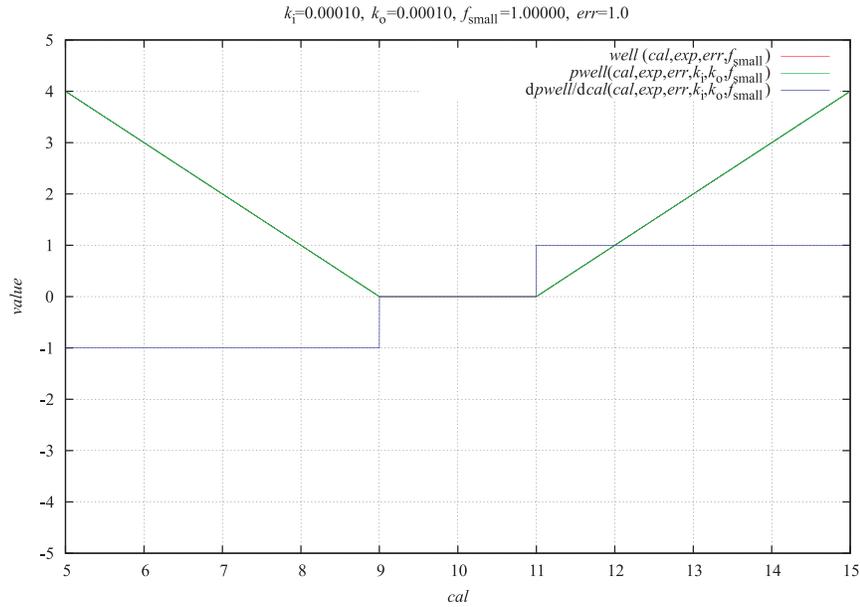
Further examples for the *pwell* function with different parameters are given. This refers to chapter 2.2.1 in the main text (p. 25 ff).



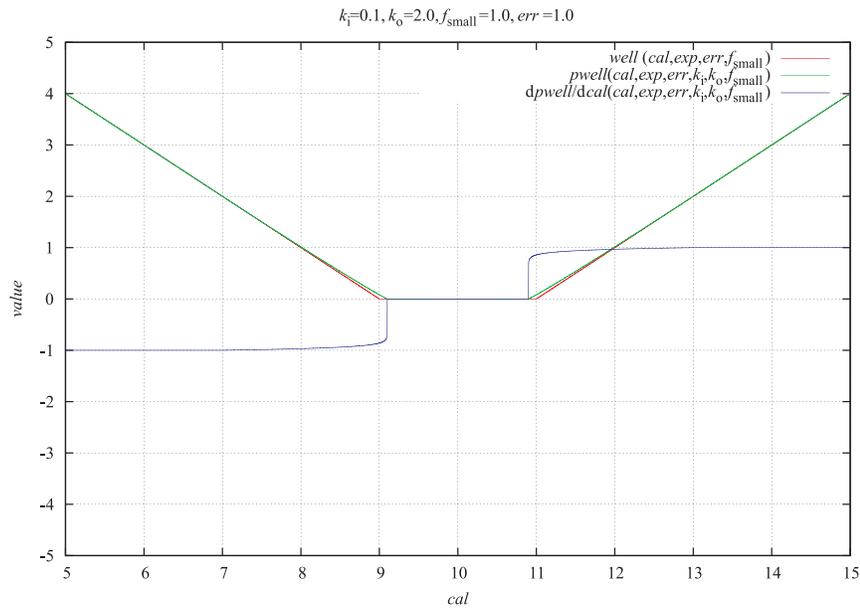
**Figure 9.1:** In this example of the  $well$  and  $pwell$  functions the error bound has been reduced from 2 to 1 compared to fig. 2.1



**Figure 9.2:** Here, a symmetrical case of the functions with  $f_{small} = 1.0$  is given.



**Figure 9.3:** In this example of the *pwell* functions the transition area has been chosen to be vanishingly small. The *well* and *pwell* functions become indistinguishable. The almost discrete jump in the derivative of the *pwell* function can be clearly seen.



**Figure 9.4:** Here the inner transition area of *pwell* was chosen to be much smaller than the outer one and so the slope of *pwell* drops rapidly to zero.

# Chapter 10

## Results

### 10.1 6-Spin System

#### 10.1.1 One-dimensional Energy Landscape

The following script was used to calculate the results for figures 4.2 to 4.3 on page 40 and the following pages.

```
set seed 584931 end                                {*Seed for random velocity assignment.*}

set echo on message on end

topology
residue REC
atom HR1 MASS=1.0 end
end
residue LIG
atom HL1 MASS=1.0 end
END
end

segment
name=""
chain
  coor
  @simple_free_5.pdb

  end
end
end
```

```
coor
@simple_free_5.pdb
```

```
end
```

```
coor
disp=reffe
@simple_free_5.pdb
end
```

```
set echo on message on end
```

```
INPharma
```

```
help
```

```
tauf=100e-12
```

```
taub=10e-09
```

```
k12=1000000
```

```
k21=3000000
```

```
    CL10=10
```

```
    CL20=5
```

```
    CTO=2
```

```
    NUE=800000000
```

```
CUTOff=1000
```

```
    FUNC WELL 1.0
```

```
!2    FUNC PWEL 1.0 0.1 0.05
```

```
!3    FUNC PWEL 1.0 0.5 0.25
```

```
!4    FUNC PWEL 1.0 0.9 0.9
```

```
!5    FUNC PWEL 1.0 1. 0.9
```

```
!6    FUNC WELL 0.2
```

```
!7    FUNC PWEL 0.2 0.1 0.05
```

```
!8    FUNC PWEL 0.2 0.5 0.25
```

```
!9    FUNC PWEL 0.2 0.9 0.9
```

```
!10   FUNC PWEL 0.2 1. 0.9
```

```
FORCeconstant=1
```

```
IEXP=2
```

```
    MIXIngTimes
```

```
    ADD 0.3
```

```
    END
```

```
! Selection of active protons --> Setup hlist1, hlist11, hlist2, hlist12
```

```
ACTIve
```

```
(resid 1) ! LIG1 free
```

```
(resid 2) ! LIG2 free
```

```

(resid 3) ! TAR1 bound
(resid 4) ! LIG1 bound
(resid 5) ! TAR2 bound
(resid 6) ! LIG2 bound

! Definition of detectable groups --> Setup of group list (one for both)
GROUP
ADD "A1"      (name HL1 and resid 1)
ADD "B1"      (name HL1 and resid 2)
END
!
! ! Definition of peakvolumes --> Setup of group list (one for both)
!
VOLUMes
  ADD "A1" "A1"      0.57002E+01  0.57002E+00
END ! VOLUMES
END ! INPHARMA

flags exclude * include inpharma end ! {*Include the relaxation energy term.*}
vector do (store1=0) (all)
vector do (store1=1) (name HL1 and resid 4)

write coor output=refine_inpharma_0mod.pdb end

flags exclude * include inph end
energy end

evaluate ($step = 1 )
evaluate ($nstep = 1 )
evaluate ($nstepmax = 1000 )

evaluate ($mstep = 1 )
evaluate ($mstepmax = 1 )
evaluate ($shift = 10 )

flags exclude * include inpn end

coor copy end ! Copy the original coordinates the the reference set

DISPLAY #M N mstep nstep e0 x10 x20 r10 r20 e1 x11 x21 r11 r21
DISPLAY #M N 3      4  5  6  7  8  9  10  11  12 13  14
!set DISPLAY="display.txt" END
set echo off message off end
while ( $mstep <= $mstepmax ) loop mstep
  while ( $nstep <= $nstepmax ) loop nstep

```

```
vector do (x = x + ($nstep-$nstepmax/2)*0.01 ) ( resid 4 )
vector show ave ( X ) ( resid 4 )
eval ($x10 = $result)
vector show ave ( X ) ( resid 6 )
eval ($x20 = $result)
pick bond (resid 3) (resid 4) geometry
eval ($r10 = $result)
pick bond (resid 5) (resid 6) geometry
eval ($r20 = $result)
energy end

eval ($e0 = $ener)

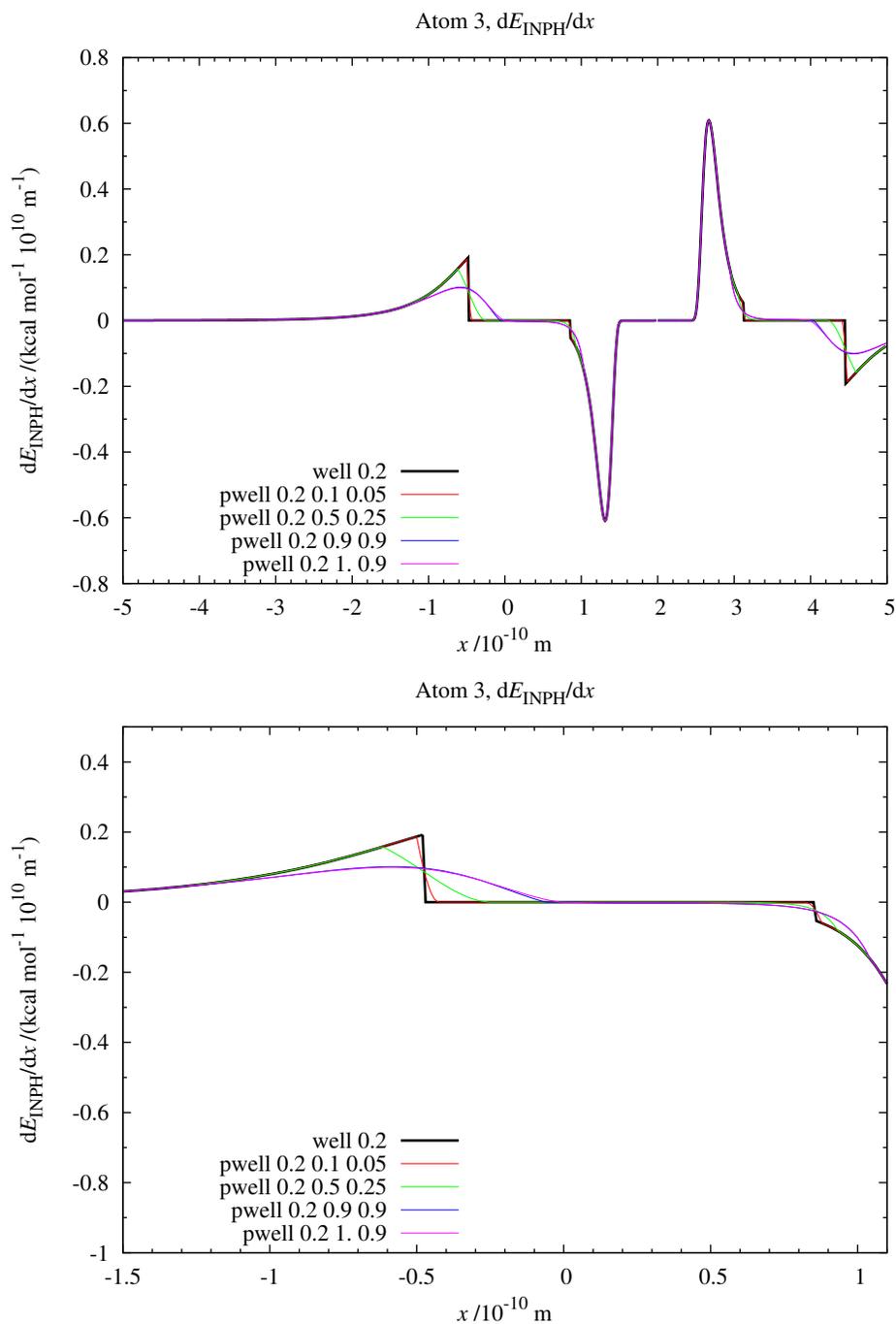
vector show ave ( X ) ( resid 4 )
eval ($x11 = $result)
vector show ave ( X ) ( resid 6 )
eval ($x21 = $result)
pick bond (resid 3) (resid 4) geometry
eval ($r11 = $result)
pick bond (resid 5) (resid 6) geometry
eval ($r21 = $result)
eval ($e1 = $ener)

DISPLAY M N $mstep $nstep $e0 $x10 $x20 $r10 $r20 $e1 $x11 $x21 $r11 $r21

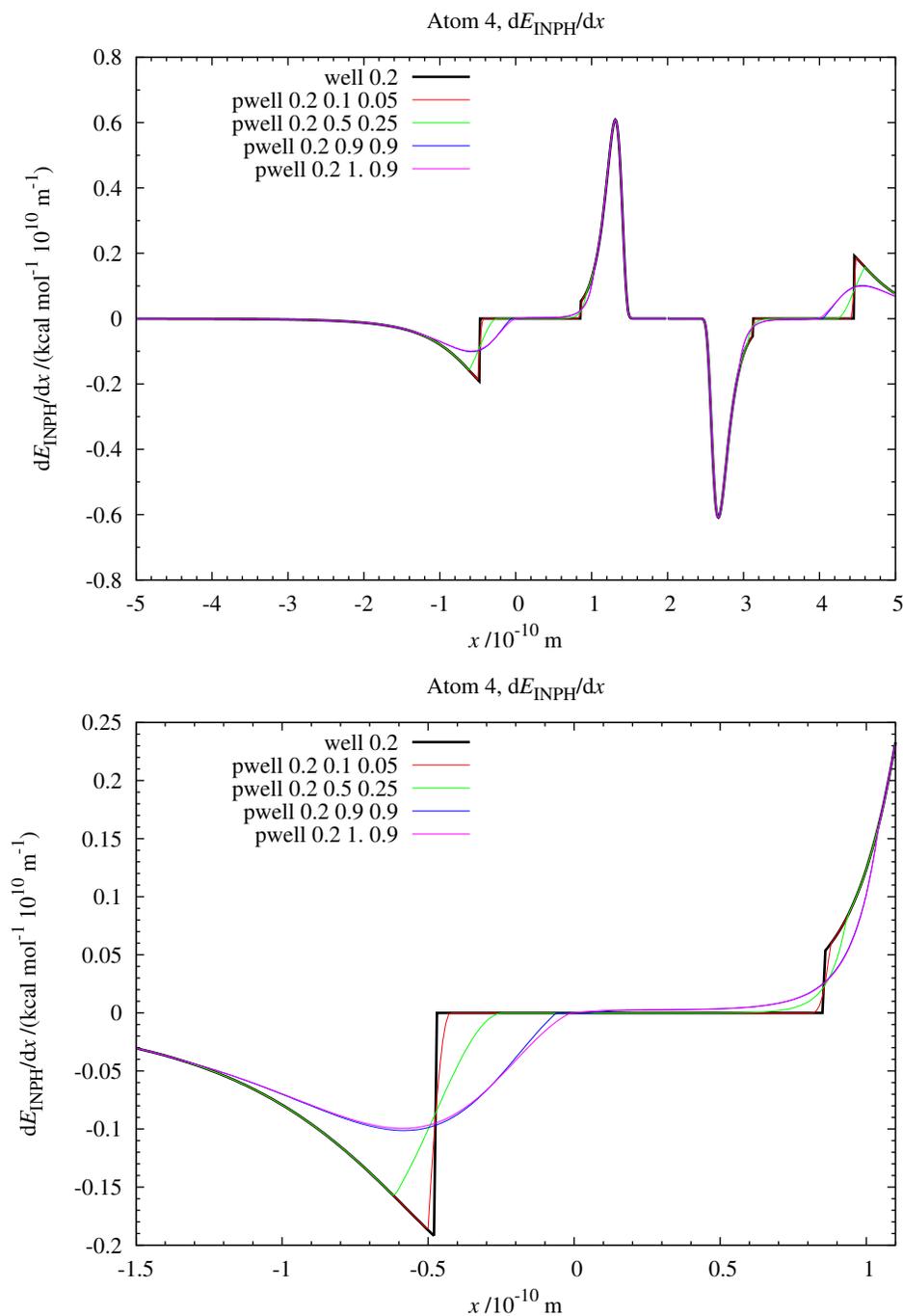
coor swap end !Now bring back the original coordinates from the reference
!set to the main set
coor copy end !And refresh the reference backup

evaluate ($nstep = $nstep + 1 )
end loop nstep
evaluate ($mstep = $mstep + 1 )
evaluate ($nstep = 1 )
evaluate ($message = "INCREMENT")
DISPLAY $message
end loop mstep
stop
```

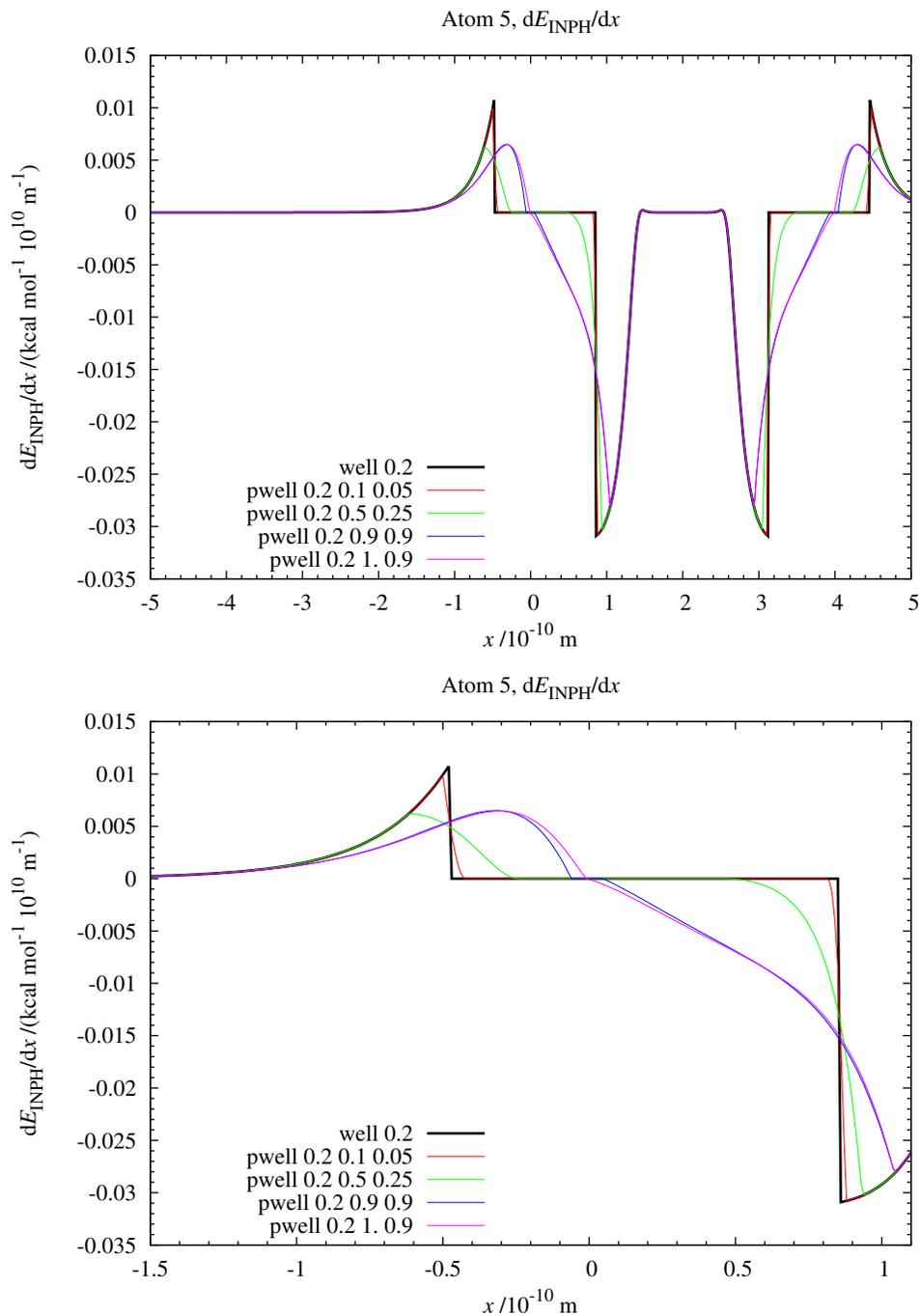
Here the analytical derivatives for moving atom 4 along the  $x$ -axis in chapter 4.1.1 using  $f_{\text{small}} = 0.2$  are given. The corresponding plots using  $f_{\text{small}} = 1.0$  (4.4-4.7) can be found on page 43.



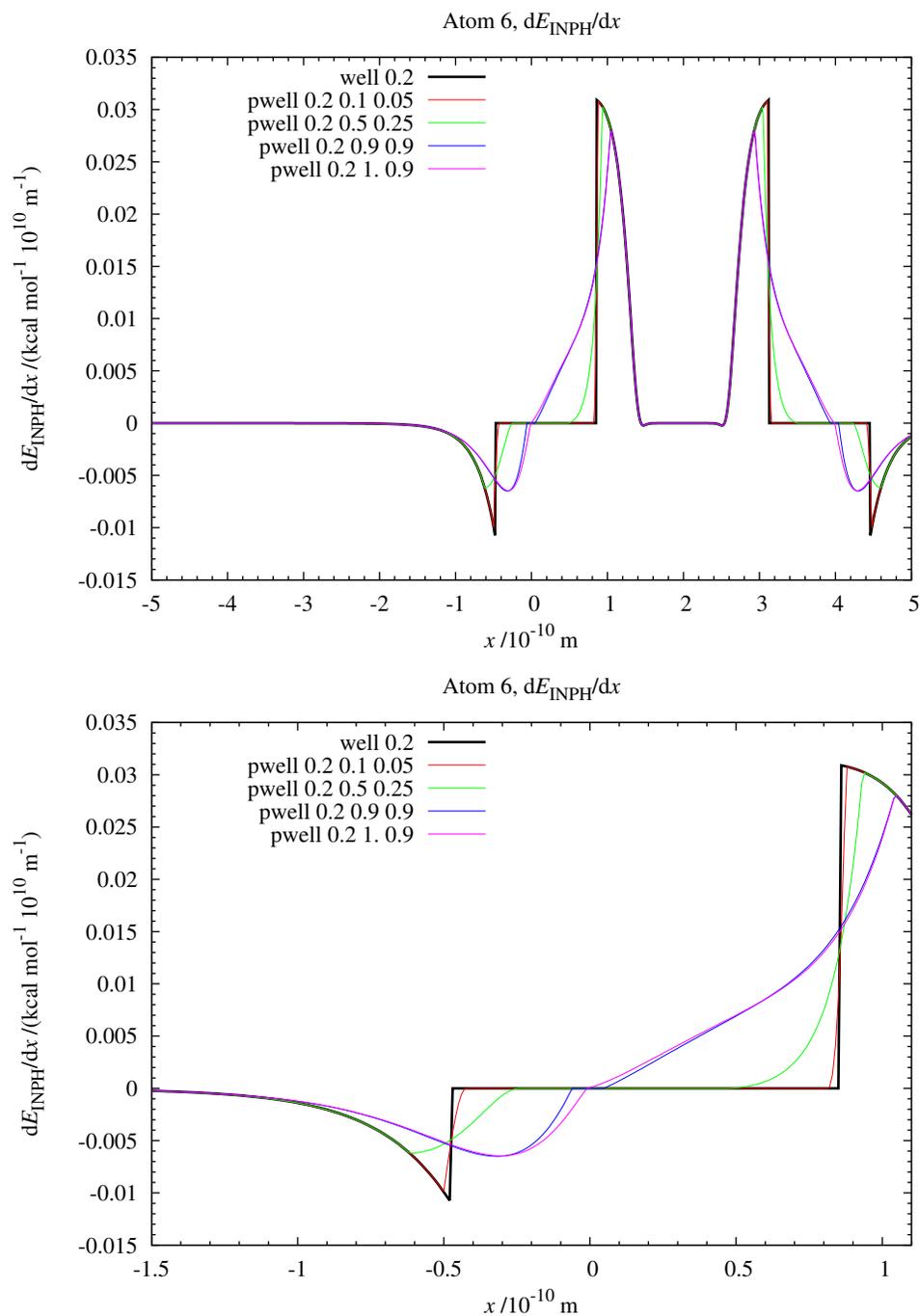
**Figure 10.1:** Top: The derivative of the INPHARMA energies for atom 3 (target 1, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of atom 4 (ligand 1, bound) is plotted considering only a single experimental diagonal peak. The functions *well* and *pwell* have been used with  $f_{\text{small}} = 0.2$ . All their parameters are given in the key in the sequence  $f_{\text{small}}, k_i, k_o$  (see fig. 2.1). Bottom: A close-up of the low-penalty area around  $x = 0 \text{ \AA}$  is shown.



**Figure 10.2:** Top: The derivative of the INPHARMA energies for atom 4 (ligand 1, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of the same atom (atom 4; ligand 1, bound) is plotted considering only a single experimental diagonal peak. Bottom: Close-up of the low-penalty area around  $x = 0$ .



**Figure 10.3:** Top: The derivative of the INPHARMA energies for atom 5 (target 2, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of atom 4 (ligand 1, bound) is plotted considering only a single experimental diagonal peak. Bottom: Close-up of the low-penalty area around  $x = 0 \text{ \AA}$ .



**Figure 10.4:** Top: The derivative of the INPHARMA energies for atom 6 (ligand 2, bound), for the system conformations as described in the text and fig. 4.1, under a linear move of atom 4 (ligand 1, bound) is plotted considering only a single experimental diagonal peak. Bottom: Close-up of the low-penalty area around  $x = 0$  Å.

### 10.1.2 Minimization

This is the script used for minimization in chapter 4.1.3 (figs. 4.13 to 4.14).

```
set seed 584931 end      {*Seed for random velocity assignment.*}

set echo on message on end

topology
residue REC
atom HR1 MASS=1.0 end
end
residue LIG
atom HL1 MASS=1.0 end
END
end

segment
name=""
chain
  coor
    @simple_free_5.pdb
  end
end
end

coor
@simple_free_5.pdb
end

coor
disp=reffe
@simple_free_5.pdb
end

set echo on message on end
INPHarma
help
tauf=100e-12
taub=10e-09
k12=1000000
k21=3000000
  CL10=10
```

```
CL20=5
CTO=2
NUE=800000000
CUTOff=1000
FUNC PWEL 0.2 1. 0.9
FORCeconstant=1
IEXP=2
    MIXIngtimes
        !ADD 0.0
        ADD 0.3
        !ADD 0.6
    END

! Selection of active protons --> Setup hlist1, hlist11, hlist2, hlist12
ACTIve
(resid 1)    ! LIG1 free
(resid 2) ! LIG2 free
(resid 3) ! TAR1 bound
(resid 4)    ! LIG1 bound
(resid 5) ! TAR2 bound
(resid 6) ! LIG2 bound

! Definition of detectable groups --> Setup of group list (one for both)
GROUP
ADD "A1"      (name HL1 and resid 1)

ADD "B1"      (name HL1 and resid 2)
END
!
!! Definition of peakvolumes --> Setup of group list (one for both)
!
VOLUmes
    ADD "A1" "A1"          0.57002E+01  0.57002E+00
    ADD "A1" "B1"          0.30432E+00  0.30432E-01
    ADD "B1" "B1"          0.37764E+01  0.37764E+00
end ! VOLUMES
end ! INPHARMA

flags exclude *
    include inpharma
end ! {*Include the relaxation energy term.*}

write coor output=refine_inpharma_0.pdb end
```

```

write coor from=reffe output=refine_inpharma_0_comp.pdb end

vector do (store1=0) (all)

vector do (store1=1) (name HL1 and resid 4)

write coor output=refine_inpharma_0mod.pdb end

flags exclude * include inph end
energy end

evaluate ($step = 1 )
evaluate ($nstep = 1 )
evaluate ($nstepmax = 100 )

evaluate ($mstep = 1 )
evaluate ($mstepmax = 100 )
evaluate ($shift = 10 )

flags exclude * include inpn end

coor copy end ! Copy the original coordinates the the reference set

DISPLAY #M N mstep nstep e0 x10 x20 r10 r20 e1 x11 x21 r11 r21
DISPLAY #M N 3 4 5 6 7 8 9 10 11 12 13 14
set echo off message off end
while ( $mstep <= $mstepmax ) loop mstep
  while ( $nstep <= $nstepmax ) loop nstep

vector do (x = x + ($nstep-$nstepmax/2)*0.1 -5 + 1.6) ( resid 4 )
  vector show ave ( X ) ( resid 4 )
  eval ($x10 = $result)
vector do (x = x + ($mstep-$mstepmax/2)*0.1 -5 + 1.6) ( resid 6 )
  vector show ave ( X ) ( resid 6 )
  eval ($x20 = $result)
  pick bond (resid 3) (resid 4) geometry
  eval ($r10 = $result)
  pick bond (resid 5) (resid 6) geometry
  eval ($r20 = $result)
  energy end

eval ($e0 = $ener)

```

```

minimize powell
drop=0.001
nstep 50000
debug=false
nprint 50000
TOLGradient=0.00000000001
end

    vector show ave ( X ) ( resid 4 )
    eval ($x11 = $result)
    vector show ave ( X ) ( resid 6 )
    eval ($x21 = $result)
    pick bond (resid 3) (resid 4) geometry
    eval ($r11 = $result)
    pick bond (resid 5) (resid 6) geometry
    eval ($r21 = $result)
    energy end
    eval ($e1 = $ener)

DISPLAY M N $mstep $nstep $e0 $x10 $x20 $r10 $r20 $e1 $x11 $x21 $r11 $r21

coor swap end ! Now bring back the original coordinates
                ! from the reference set to the main set
coor copy end ! And refresh the reference backup

evaluate ($nstep = $nstep + 1 )
    end loop nstep
evaluate ($mstep = $mstep + 1 )
evaluate ($nstep = 1 )
evaluate ($message = "INCREMENT")
DISPLAY $message
end loop mstep
stop

```

### 10.1.3 Numerical and Analytical Derivatives

This was the protocol used to compare the analytical with the numerical derivative in (ch. 4.1.2)

```

set seed 584931 end          {*Seed for random velocity assignment.*}

```

```
set echo on message on end
```

```
topology
```

```
    residue REC
        atom HR1 MASS=1.0 end
    end
    residue LIG
        atom HL1 MASS=1.0 end
    END
```

```
end
```

```
segment
```

```
    name=""
```

```
    chain
```

```
        coor
```

```
            @simple_free_5.pdb
```

```
        end
```

```
    end
```

```
end
```

```
coor
```

```
    @simple_free_5.pdb
```

```
end
```

```
coor
```

```
    disp=refe
```

```
    @simple_free_5.pdb
```

```
end
```

```
set echo on message on end
```

```
INPHarma
```

```
    help
```

```
    tauf=100e-12
```

```
    taub=10e-09
```

```
    k12=1000000
```

```
    k21=3000000
```

```
    CL10=10
```

```
    CL20=5
```

```
    CT0=2
```

```
    NUE=800000000
```

```
    CUTOff=1000
```

```
    EXPS=0.99999
```

```

        FUNC PWEL 1 0.2 0.3
        FORCEconstant=1
        IEXP=2
MIXIngtimes
        !ADD 0.0
        ADD 0.3
        ADD 1.2
END

! Selection of active protons --> Setup hlist1, hlist11, hlist2, hlist12
ACTIVE
    (resid 1)          ! LIG1 free
    (resid 2)          ! LIG2 free
    (resid 3)          ! TAR1 bound
    (resid 4)          ! LIG1 bound
    (resid 5)          ! TAR2 bound
    (resid 6)          ! LIG2 bound

! Definition of detectable groups --> Setup of group list (one for both)
GROUP
    ADD "A1"          (name HL1 and resid 1)
    ADD "B1"          (name HL1 and resid 2)
END

!
!   ! Definition of peakvolumes --> Setup of group list (one for both)
!
VOLUMes
    ADD "A1" "A1"    0.57002E+01  0.57002E+00  0.48891E+01  0.48891E+00
    ADD "A1" "B1"    0.30432E+00  0.30432E-01  0.12202E+01  0.12202E+00
    ADD "B1" "B1"    0.37764E+01  0.37764E+00  0.26114E+01  0.26114E+00
end ! VOLUMES

end ! INPHARMA
! {*Include the relaxation energy term.*}
flags exclude * include inpharma end

write coor output=refine_inpharma_0.pdb end
write coor from=reffe output=refine_inpharma_0_comp.pdb end

vector do (store1=0) (all)

vector do (store1=1) (name HL1 and resid 4)

write coor output=refine_inpharma_0mod.pdb end

```

```

flags exclude * include inph end
energy end

evaluate ($step = 1 )
evaluate ($nstep = 1 )
evaluate ($nstepmax = 1000 )

evaluate ($mstep = 1 )
evaluate ($mstepmax = 1 )
evaluate ($shift = 10 )

flags exclude * include inpn end

coor copy end ! Copy the original coordinates the the reference set

DISPLAY #M N mstep nstep e0 x10 x20 r10 r20 e1 x11 x21 r11 r21
DISPLAY #M N 3      4   5   6   7   8   9  10  11  12 13  14
set echo off message off end
while ( $mstep <= $mstepmax ) loop mstep
  while ( $nstep <= $nstepmax ) loop nstep
    vector do (x = x + ($nstep-$nstepmax/2)*0.01 ) ( resid 4 )
    vector show ave ( X ) ( resid 4 )
    eval ($x10 = $result)
    vector show ave ( X ) ( resid 6 )
    eval ($x20 = $result)
    pick bond (resid 3) (resid 4) geometry
    eval ($r10 = $result)
    pick bond (resid 5) (resid 6) geometry
    eval ($r20 = $result)
    energy end

    eval ($e0 = $ener)

    vector show ave ( X ) ( resid 4 )
    eval ($x11 = $result)
    vector show ave ( X ) ( resid 6 )
    eval ($x21 = $result)
    pick bond (resid 3) (resid 4) geometry
    eval ($r11 = $result)
    pick bond (resid 5) (resid 6) geometry
    eval ($r21 = $result)
    eval ($e1 = $ener)
  end
end

```

```
DISPLAY M N $mstep $nstep $e0 $x10 $x20 $r10 $r20 $e1 $x11 $x21 $r11 $r21

coor swap end ! Now bring back the original coordinates
              ! from the reference
              ! set to the main set
coor copy end ! And refresh the reference backup

evaluate ($nstep = $nstep + 1 )
end loop nstep
evaluate ($mstep = $mstep + 1 )
evaluate ($nstep = 1 )
evaluate ($message = "INCREMENT")
DISPLAY $message
end loop mstep
stop
```

## 10.2 12-Spin System

### 10.2.1 Minimization

This is the skript for generating the 10,000 random models and to perform the minimization in chapter 4.2 (figs. 4.18, 4.19):

```
set seed 584931 end    {*Seed for random velocity assignment.*}
```

```
set echo on message on end
```

```
topology
```

```
residue REC
```

```
atom HR1 MASS=1.0 end
```

```
atom HR2 MASS=1.0 end
```

```
end
```

```
residue LIG
```

```
atom HL1 MASS=1.0 end
```

```
atom HL2 MASS=1.0 end
```

```
END
```

```
end
```

```
segment
```

```
name=""
```

```
chain
```

```
  coor
```

```
    @simple_free_6.pdb
```

```
  end
```

```
end
```

```
end
```

```
coor
```

```
@simple_free_6.pdb
```

```
end
```

```
coor
```

```
disp=refe
```

```
@simple_free_6.pdb
```

```
end
```

```
coor
```

```
disp=comp
@simple_free_6.pdb
end

set echo on message on end
INPHarma
help
tauf=100e-12
taub=10e-09
k12=1000000
k21=3000000
      CL10=10
      CL20=5
      CTO=2
      NUE=800000000
      FUNC PWEL 0.2 1 1
CUTOff=1000
      EXPS=0.99999
FORCeconstant=1000000
IEXP=2
      MIXIngtimes
      !ADD 0.0
      ADD 0.3
      ADD 0.6
      END

! Selection of active protons --> Setup hlist1, hlist11, hlist2, hlist12
ACTIVE
(resid 1) ! LIG1 free
(resid 2) ! LIG2 free
(resid 3) ! TAR1 bound
(resid 4) ! LIG1 bound
(resid 5) ! TAR2 bound
(resid 6) ! LIG2 bound

! Definition of detectable groups --> Setup of group list (one for both)
GROUP
ADD "A1"      (name HL1 and resid 1)
ADD "A2"      (name HL2 and resid 1)

ADD "B1"      (name HL1 and resid 2)
ADD "B2"      (name HL2 and resid 2)
END
```

```

!
!! Definition of peakvolumes --> Setup of group list (one for both)
!
VOLUmes
  ADD "A1" "A1" 0.33163E+01 0.33163E+00 0.22230E+01 0.22230E+00
  ADD "A1" "A2" 0.13943E+01 0.13943E+00 0.15814E+01 0.15814E+00
  ADD "A1" "B1" 0.16873E+00 0.16873E-01 0.28685E+00 0.28685E-01
  ADD "A1" "B2" 0.11304E-01 0.11304E-02 0.35619E-01 0.35619E-02
  ADD "A2" "A2" 0.40867E+01 0.40867E+00 0.27190E+01 0.27190E+00
  ADD "A2" "B1" 0.29554E-01 0.29554E-02 0.94046E-01 0.94046E-02
  ADD "A2" "B2" 0.15754E-02 0.15754E-03 0.88486E-02 0.88486E-03
  ADD "B1" "B1" 0.27229E+01 0.27229E+00 0.17475E+01 0.17475E+00
  ADD "B1" "B2" 0.34871E+00 0.34871E-01 0.47096E+00 0.47096E-01
  ADD "B2" "B2" 0.31708E+01 0.31708E+00 0.23194E+01 0.23194E+00

end ! VOLUMES
end ! INPHARMA

! {*Include the relaxation energy term.*}
flags exclude * include inpharma end

write coor output=refine_inpharma_0.pdb end
write coor from=refe output=refine_inpharma_0_comp.pdb end
write coor output=refine_inpharma_0mod.pdb end

flags exclude * include inph end

evaluate ($step = 1 )
evaluate ($nstep = 1 )
evaluate ($nstepmax = 100 )

evaluate ($mstep = 1 )
evaluate ($mstepmax = 100 )
evaluate ($shift = 10 )

flags exclude * include inph end

write traj asci=false outp=trajectory.dcd sele=(all) end
coor copy end ! Copy the original coordinates the the reference set

DISPLAY #M N mstep nstep e0 x10 x20 r10 r20 e1 x11 x21 r11 r21
DISPLAY #M N 3 4 5 6 7 8 9 10 11 12 13 14

```

```

set echo off message off end
while ( $mstep <= $mstepmax ) loop mstep
  while ( $nstep <= $nstepmax ) loop nstep

    evaluate ($noise = $mstep*0.01 )
vector do (x = x + 2*(0.5-rand())*$noise ) (all)
vector do (y = y + 2*(0.5-rand())*$noise ) (all)
vector do (z = z + 2*(0.5-rand())*$noise ) (all)

    vector show ave ( X ) ( resid 4 and name "HL1")
    eval ($x10 = $result)
    vector show ave ( X ) ( resid 6 and name "HL1")
    eval ($x20 = $result)
    pick bond (resid 3) (resid 4) geometry
    eval ($r10 = $result)
    pick bond (resid 5) (resid 6) geometry
    eval ($r20 = $result)
    DISPLAY NUMERICAL DERIVATIVE BEFORE MINIMIZATION
    flags exclude * include inpn end
    energy end
    eval ($e0 = $ener)

    flags exclude * include inph end
minimize powell
drop=1
nstep 50000
debug=false
nprint 50000
TOLGradient=0.00000000001
end

write traj next end

vector show ave ( X ) ( resid 4 and name "HL1")
eval ($x11 = $result)
vector show ave ( X ) ( resid 6 and name "HL2")
eval ($x21 = $result)
pick bond (resid 3) (resid 4) geometry
eval ($r11 = $result)
pick bond (resid 5) (resid 6) geometry
eval ($r21 = $result)

```

```

DISPLAY NUMERICAL DERIVATIVE AFTER MINIMIZATION
flags exclude * include inpn end
energy end
eval ($e1 = $ener)

evaluate ($filename =
  ENCODE($mstep)+"/minimized_"+ENCODE($mstep)+"_"+ENCODE($nstep)+".pdb")
write coor output=$filename end

DISPLAY M N $mstep $nstep $e0 $x10 $x20 $r10 $r20 $e1 $x11 $x21 $r11 $r21

coor swap end ! Now bring back the orginal coordinates
              ! from the reference set to the main set
coor copy end ! And refresh the reference backup

  evaluate ($nstep = $nstep + 1 )
end loop nstep
evaluate ($mstep = $mstep + 1 )
evaluate ($nstep = 1 )
evaluate ($message = "INCREMENT")
DISPLAY $message
end loop mstep
stop

```

And this the pdb simple\_free\_6.pdb with the initial coordinates :

ATOM	1	HL1	LIG	1	0.000	8.000	0.000
ATOM	2	HL2	LIG	1	-2.000	8.000	0.000
ATOM	3	HL1	LIG	2	0.000	2.000	0.000
ATOM	4	HL2	LIG	2	-2.000	2.000	0.000
ATOM	5	HR1	REC	3	2.000	6.000	0.000
ATOM	6	HR2	REC	3	4.000	6.000	0.000
ATOM	7	HL1	LIG	4	0.000	6.000	0.000
ATOM	8	HL2	LIG	4	-2.000	6.000	0.000
ATOM	9	HR1	REC	5	2.000	0.000	0.000
ATOM	10	HR2	REC	5	4.000	0.000	0.000
ATOM	11	HL1	LIG	6	0.000	0.000	0.000
ATOM	12	HL2	LIG	6	-2.000	0.000	0.000

## 10.3 Tubulin - Epothilone A - Baccatin III

### 10.3.1 Minimization

The following script was used for the INPHARMA restrain minimization of the EPO-BAC-TUBB complex (ch. 4.3.1). Only the main script is given and not the detailed coordinates of the complex or the details of all topology and parameter files.

```

set seed 584931 end          {*Seed for random velocity assignment.*}

set echo on message on end

evaluate ($file1= "HDtrial3_2108w.pdb" )
evaluate ($file2= "HDtrial3_935w.pdb" )
evaluate
  ($filedir1="/usr/users/mreese/run19/structures/it1/water/sequential/")
evaluate
  ($filedir2="/usr/users/mreese/run20/structures/it1/water/sequential/")
evaluate ($id = "test.44.1.inp.")

evaluate ($file1= "HDtrial3_2108w.pdb" )
evaluate ($file2= "HDtrial3_935w.pdb" )

evaluate ($infile1= $filedir1+"/"+$file1 )
evaluate ($infile2= $filedir2+"/"+$file2 )

evaluate ($par_nonbonded="PARAM19")
evaluate ($OT1_becomes="O")
evaluate ($OT2_becomes="OXT")

topology
  @../toppar/epo/topallhdg5.3.pro
  @../toppar/epo/epoa_20040915.top
  @../toppar/bac/BAC_learned.top
end

!##### Complex 1
segment
  name="TUBB"
  chain
    @TOPPAR:toph19.pep          {*Read peptide bond file      *}

```

```

        coordinates @$infile1      {*interpret coordinate file to *}
    end                               {*obtain the sequence      *}
end

segment
    name="EPOA"
    chain
        coordinates @$infile1      {*interpret coordinate file to *}
    end                               {*obtain the sequence      *}
end

vector do (name=$OT1_becomes) (name OT1)
vector do (name=$OT2_becomes) (name OT2)

coor @$infile1
##### free ligand 1
vector do (segid="T1 ") (segid "TUBB")
vector do (segid="L1 ") (segid "EPOA")

segment
    name="EPOA"
    chain
        coordinates @$infile1      {*interpret coordinate file to *}
    end                               {*obtain the sequence      *}
end

coor @$infile1

vector do (segid="L1F ") (segid "EPOA")

##### complex 2
segment
    name="TUBB"
    chain
        @TOPPAR:toph19.pep          {*Read peptide bond file      *}
        coordinates @$infile2      {*interpret coordinate file to *}
    end                               {*obtain the sequence      *}
end

segment
    name="BAC "
    chain
        coordinates @$infile2      {*interpret coordinate file to *}

```

```

        end                                { *obtain the sequence      * }
    end

vector do (name=$OT1_becomes) (name OT1)
vector do (name=$OT2_becomes) (name OT2)
coor      @$infile2

vector do (segid="T2  ") (segid "TUBB")
vector do (segid="L2  ") (segid "BAC ")
!##### free ligand 2
segment
    name="BAC "
    chain
        coordinates @$infile2          { *interpret coordinate file to * }
    end                                { *obtain the sequence      * }
end

coor @$infile2
vector do (segid="L2F ") (segid "BAC ")

!#####

parameter
    @../toppar/epo/parallhdg5.3_PARAM19.pro
    @../toppar/epo/epoa_20040915.pro
    @../toppar/bac/BAC_learned.pro
    nbonds
    atom nbxmod -3 tolerance 0.5
    repel 0.75 rcon 4.0 rexp 2 cutnb 4.5
end
end

parameter { *Set energy parameters.* }
NBONds
    CUTNB=12 WMIN=1.5
    REPEL = 0.0 cdie shift
    TOLERance = 0.5 NBXMOD = 5
    ctonnb=5.5 ctofnb=11.0
END
end

{ Seperate complexes and ligands }
vector do (x=x+100) (segid "T2  " or segid "L2  " or segid "L2F ")

```

```
vector do (y=y+50) (segid "L2F " or segid "L1F ")
```

```
!##### END GENERATE
```

#### CONStraints

```
  INTERaction ( segid T1 or segid L1 )
                ( segid T1 or segid L1 ) WEIGhts * 1. END
  INTERaction ( segid T2 or segid L2 )
                ( segid T2 or segid L2 ) WEIGhts * 1. END
  INTERaction ( segid L1F )
                ( segid L1F ) WEIGhts * 1. END
  INTERaction ( segid L2F )
                ( segid L2F ) WEIGhts * 1. END
```

```
END
```

```
!energy end
```

```
!#####
!
```

```
! Define exchangable=exchanged protons of Aminoacids
```

```
vector identify (store9) (not all)
vector identify (store9) (store9 or name "HN")
vector identify (store9) (store9 or (name HZ# and resname "LYS"))
vector identify (store9) (store9 or (name "HE" and resname ARG))
vector identify (store9) (store9 or (name HH## and resname ARG))
vector identify (store9) (store9 or (name HD1 and resname HIS))
vector identify (store9) (store9 or ( name HE2 and resname HIS ))
vector identify (store9) (store9 or (name HE1 and resname TRP))
vector identify (store9) (store9 or (name HG and resname CYS))
vector identify (store9) (store9 or (name HD2# and resname ASN))
vector identify (store9) (store9 or (name HG and resname SER))
vector identify (store9) (store9 or (name HG1 and resname THR))
vector identify (store9) (store9 or (name HH and resname TYR))
vector identify (store9) (store9 and hydrogen)
```

```
!Define INPHARMA active selections
```

```
vector identify (store1)
  (not store9 and segid L1F and hydrogen and
  (not name H003) and (not name H007))          ! LIG1 free
vector identify (store2)
  (not store9 and segid L2F and hydrogen and not name HA%)
                                                    ! LIG2 free
vector identify (store3)
  (not store9 and segid T1 and hydrogen )       ! TAR1 bound
```

```

vector identify (store4)
  (not store9 and segid L1 and hydrogen and
  (not name H003) and (not name H007))           ! LIG1 bound
vector identify (store5)
  (not store9 and segid T2 and hydrogen )       ! TAR2 bound
vector identify (store6)
  (not store9 and segid L2 and hydrogen and not name HA%)
                                                    ! LIG2 bound

#####

vector do (store7=0) (all)
vector do (store7=1)
  (not byres((resname EPOA or resname BAC) around 14.))
CONStraints fix (store7) end
vector do ( fbeta = 0. ) ( all )
vector do ( fbeta = 500. ) ( not hydrogen and not store7 )
vector do ( fbeta = 50000. ) ( all )

vector do ( resid="1001" ) ( segid "L1F " )
vector do ( resid="1002" ) ( segid "L2F " )
vector do ( resid="2001" ) ( segid "L1 " )
vector do ( resid="2001" ) ( segid "L2 " )
noe
  reset
  nrestraints = 100000      ! allocate space for NOEs
  ceiling 1000
end

noe class ambi @ambig_BAC.tbl @ambig_EPO.tbl end
noe
  averaging * R-6
  potential * soft
  scale * 1.0
  sqconstant * 1.0
  sqexponent * 2
  soexponent * 1
  rswitch * 1.0
  sqoffset * 0.0
  asymptote * 2.0
  msoexponent * 1
  masymptote * -0.1

```

```
    mrswitch      * 1.0
    avexpo hbond 20
end

noe class unambi @unambig_EPO.tbl end
noe
    averaging * sum
    potential * soft
    scale      * 100.0
    sqconstant * 1.0
    sqexponent * 2
    soexponent * 1
    rswitch    * 1.0
    sqoffset   * 0.0
    asymptote  * 2.0
    msoexponent * 1
    masymptote * -0.1
    mrswitch   * 1.0
    avexpo hbond 20
end

restraints
    dihedral
        nassign=10000
        scale 100.0
@dihe_EPO_bound.tbl
@dihe_EPO_free.tbl
end

set echo on message on end
INPHarma
    help
    tauf=160e-12
    taub=1270e-9 !1270e-9

    k12=1000000
    k21=2850000
    CL10=100
    CL20=100
    CT0=1.79
    nue=900000000
    cutoff=10
    EXPS=0.99999
```

```
FUNC PWEL 0.2 1 1
FORCeconstant=1000000.0
MIXIngtimes
  ADD 0.0215
  ADD 0.0415
  ADD 0.0715
  ADD 0.1015
  ADD 0.2015
END
```

```
! Selection of active protons --> Setup hlist1, hlist11, hlist2, hlist12
ACTIVE
```

```
(store1)! LIG1 free
(store2)! LIG2 free
(store3)! TAR1 bound
(store4)! LIG1 bound
(store5)! TAR2 bound
(store6)! LIG2 bound
```

```
! Definition of detectable groups --> Setup of group list (one for both)
GROUP
```

```
  ADD "E21" (segid L1F and name H21# and resname EPOA)
  ADD "E19" (segid L1F and name H19 and resname EPOA)
  ADD "E27" (segid L1F and name H27# and resname EPOA)
  ADD "E23" (segid L1F and name H23# and resname EPOA)
  ADD "E22" (segid L1F and name H22# and resname EPOA)
  ADD "E25" (segid L1F and name H25# and resname EPOA)
  ADD "E3" (segid L1F and name H03 and resname EPOA)
  ADD "E7" (segid L1F and name H07 and resname EPOA)
  ADD "E12" (segid L1F and name H12 and resname EPOA)
  ADD "E13" (segid L1F and name H13 and resname EPOA)
  ADD "E15" (segid L1F and name H15 and resname EPOA)
  ADD "E17" (segid L1F and name H17 and resname EPOA)
  ADD "E24" (segid L1F and name H24# and resname EPOA)
  ADD "E25E22" (segid L1F and (name H25# or name H22#)
    and resname EPOA)
  ADD "B16" (segid L2F and name #H47 and resname BAC)
  ADD "B17" (segid L2F and name #H46 and resname BAC)
  ADD "B0" (segid L2F and (name 1H09 or name 1H05)
    and resname BAC)
  ADD "BM" (segid L2F and (name 1H08 or name 1H06)
    and resname BAC)
  ADD "BP" (segid L2F and name 1H07 and resname BAC)
```

```

ADD "B2" (segid L2F and name 1H02 and resname BAC)
ADD "B3" (segid L2F and name 1H10 and resname BAC)
ADD "B5" (segid L2F and name 1H15 and resname BAC)
ADD "B6" (segid L2F and name #H16 and resname BAC)
ADD "B7" (segid L2F and name 1H17 and resname BAC)
ADD "B10" (segid L2F and name 1H21 and resname BAC)
ADD "B13" (segid L2F and name 1H26 and resname BAC)
ADD "B18" (segid L2F and name #H44 and resname BAC)
ADD "B19" (segid L2F and name #H19 and resname BAC)
ADD "B20A" (segid L2F and name 1H14 and resname BAC)
ADD "B20B" (segid L2F and name 2H14 and resname BAC)
ADD "B17B16" (segid L2F and (name #H47 or name #H46)
              and resname BAC)

end
@absolute_peaks_2004-07-13_B_XPLOR.inp
end ! VOLUMES
end ! INPHARMA

flags exclude inpharma include elec end

evaluate
( $filename = $id+"inph_0.calcforces_"+"$file1"+"_"+"$file2"+"1000000.pdb")

coor copy end ! write coordindates to comparision set
! for later RMSD calculation
write coor output=$filename end
flags exclude * include inph end

energy end

DISPLAY # CORRELATION 0 0_comp
INPH WRITe END

evaluate ($step=1)
evaluate ($lstep=1)
evaluate ($GRAD = 10)
while ($lstep <= 6) loop mini
flags exclude * end
flags include elec vdw bond angle dihe impr inph noe cdihe end
  minimize powell
    drop=10
    nstep 100
    debug=true

```

```
        nprint=1
        TOLGradient=1
    end
    vector do (mass = 100) (all)
    vector do ( fbeta = 500. ) ( all )

        evaluate ($step=$step + 1)
        evaluate ($lstep=$lstep + 1)
end loop mini

energy end

DISPLAY # CORRELATION 1 mini.init
INPH WRITE END

evaluate
    ( $filename = $id+"inph_SAmulti.init"+$file1+"_"+$file2+".1000000.pdb")
write coor output=$filename end

coor rms sele=(not store7) end
DISPLAY $filename RMSD $result
stop
```

The cross peaks used for minimization where the following  
(format ("from"-peak "to"-peak)):

(BM E21), (BP E21), (B0 E21), (E21 B16), (E21 B17),  
 (BM E27), (BP E27), (B0 E27), (BM E22), (BP E22),  
 (B0 E22), (BM E23), (BP E23), (B0 E23), (BM E25),  
 (BP E25), (B0 E25), (E19 B16), (E19 B17), (E21 B2 ),  
 (E15 B17), (E15 B16), (E15 B0 ), (E17 B17), (B3 E22),  
 (B13 E27), (B5 E22), (B5 E24), (B5 E23), (B5 E27),  
 (B5 E21), (B5 E17), (B2 E23), (B2 E27), (B2 E17),  
 (B10 E25), (B10 E23), (B10 E21), (BM E24), (BM E7 ),  
 (BM E15), (BM E17), (BM E19), (BP E24), (BP E17),  
 (B0 E24), (B0 E12), (B0 E13), (B0 E7 ), (B0 E17),  
 (B0 E19)

with the following peakvolumes:<sup>12</sup>

ADD BM E21 0	18	71	18	130	20	241	21	625	19
ADD BP E21 0	12	41	12	48	11	115	19	372	20
ADD B0 E21 0	18	36	18	111	20	180	22	558	17
ADD E21 B16 0	14	79	14	92	14	156	26	518	26
ADD E21 B17 0	13	48	13	119	15	180	29	675	26
ADD BM E27 0	12	50	12	150	20	222	21	647	19
ADD BP E27 0	11	19	11	78	11	105	20	345	20
ADD B0 E27 0	19	30	19	94	20	203	23	606	17
ADD BM E22 0	12	43	12	115	19	181	21	520	19
ADD BP E22 0	7	13	7	46	11	66	18	277	19
ADD B0 E22 0	17	43	17	94	19	151	22	464	17
ADD BM E23 0	14	34	14	86	19	177	20	457	19
ADD BP E23 0	2	0	2	34	11	83	18	240	20
ADD B0 E23 0	15	33	15	75	20	136	23	407	17
ADD BM E25 0	13	35	13	133	19	171	19	610	19
ADD BP E25 0	10	33	10	64	11	98	20	332	19
ADD B0 E25 0	18	28	18	113	20	193	23	530	17
ADD E19 B16 0	14	28	14	46	17	83	20	211	25
ADD E19 B17 0	16	34	16	53	17	79	20	289	26
ADD E21 B2 0	11	3	11	14	14	77	31	199	26
ADD E15 B17 0	16	25	16	44	13	69	18	235	21
ADD E15 B16 0	12	16	12	44	12	86	17	192	20
ADD E15 B0 0	12	18	12	44	14	79	20	196	21
ADD E17 B17 0	11	36	11	66	12	76	18	242	30
ADD B3 E22 0	2	0	2	31	12	68	17	234	17
ADD B13 E27 0	16	1	16	26	16	87	21	286	21
ADD B5 E22 0	12	24	12	46	13	69	21	207	24
ADD B5 E24 0	1	0	1	1	15	46	23	153	23

```

ADD B5 E23 0 2 0 2 20 13 43 21 201 22
ADD B5 E27 0 13 12 13 37 15 40 22 215 24
ADD B5 E21 0 13 19 13 53 14 62 22 170 23
ADD B5 E17 0 13 1 13 0 3 26 22 64 19
ADD B2 E23 0 17 1 17 21 11 42 23 83 20
ADD B2 E27 0 24 1 24 34 13 32 23 168 20
ADD B2 E17 0 2 0 2 0 3 1 21 85 22
ADD B10 E25 0 2 0 2 31 15 41 22 211 15
ADD B10 E23 0 2 0 2 22 12 38 21 115 13
ADD B10 E21 0 2 0 2 31 17 41 27 135 14
ADD BM E24 0 15 30 15 115 19 175 22 454 19
ADD BM E7 0 2 0 2 31 17 52 20 151 19
ADD BM E15 0 2 0 2 0 3 51 17 205 19
ADD BM E17 0 2 0 2 48 21 58 21 198 19
ADD BM E19 0 2 0 2 60 20 108 21 266 19
ADD BP E24 0 2 0 2 42 12 52 20 255 18
ADD BP E17 0 2 0 2 0 3 28 19 103 20
ADD B0 E24 0 18 24 18 66 20 131 24 408 17
ADD B0 E12 0 19 33 19 31 19 88 23 194 17
ADD B0 E13 0 2 0 2 33 21 39 22 223 18
ADD B0 E7 0 2 0 2 32 19 37 23 141 17
ADD B0 E17 0 2 0 2 44 21 53 26 161 17
ADD B0 E19 0 2 0 2 49 21 81 25 183 17

```

### 10.3.2 Simulated Annealing

The following script was used for the INPHARMA restrain driven simulated annealing of the EPO-BAC-TUBB complex (ch. 4.3.2). Only the main script is given.

```

set seed 584931 end                                     {*Seed for random velocity assignment.*}

set echo on message on end

evaluate ($file1= "HDtrial3_2108w.pdb" )
evaluate ($file2= "HDtrial3_935w.pdb" )
evaluate ($filedir1=
           "/usr/users/mreese/run19/structures/it1/water/sequential/" )
evaluate ($filedir2=
           "/usr/users/mreese/run20/structures/it1/water/sequential/" )
evaluate ($id = "test.39.1.inp.")

evaluate ($file1= "HDtrial3_3462w.pdb" )
evaluate ($file2= "HDtrial3_3999w.pdb" )

```

```

evaluate ($infile1= $filedir1+"/"+$file1 )
evaluate ($infile2= $filedir2+"/"+$file2 )

evaluate ($par_nonbonded="PARAM19")
evaluate ($OT1_becomes="O")
evaluate ($OT2_becomes="OXT")

topology
    @../toppar/epo/topallhdg5.3.pro
    @../toppar/epo/epoa_20040915.top
    @../toppar/bac/BAC_learned.top
end

!##### Complex 1
segment
    name="TUBB"
    chain
        @TOPPAR:toph19.pep                {*Read peptide bond file *}
        coordinates @$infile1             {*interpret coordinate file to *}
    end                                    {*obtain the sequence *}
end

segment
    name="EPOA"
    chain
        coordinates @$infile1             {*interpret coordinate file to *}
    end                                    {*obtain the sequence *}
end

vector do (name=$OT1_becomes) (name OT1)
vector do (name=$OT2_becomes) (name OT2)

coor @$infile1
!##### free ligand 1
vector do (segid="T1 ") (segid "TUBB")
vector do (segid="L1 ") (segid "EPOA")

segment
    name="EPOA"
    chain
        coordinates @$infile1             {*interpret coordinate file to *}

```

```

        end                                {*obtain the sequence    *}
    end

    coor @$infile1

    vector do (segid="L1F ") (segid "EPOA")

    !##### complex 2
    segment
        name="TUBB"
        chain
            @TOPPAR:toph19.pep                {*Read peptide bond file    *}
            coordinates @$infile2            {*interpret coordinate file to *}
        end                                    {*obtain the sequence        *}
    end

    segment
        name="BAC "
        chain
            coordinates @$infile2            {*interpret coordinate file to *}
        end                                    {*obtain the sequence        *}
    end

    vector do (name=$OT1_becomes) (name OT1)
    vector do (name=$OT2_becomes) (name OT2)
    coor     @$infile2

    vector do (segid="T2 ") (segid "TUBB")
    vector do (segid="L2 ") (segid "BAC ")
    !##### free ligand 2
    segment
        name="BAC "
        chain
            coordinates @$infile2            {*interpret coordinate file to *}
        end                                    {*obtain the sequence        *}
    end

    coor @$infile2
    vector do (segid="L2F ") (segid "BAC ")

    !#####

    parameter

```

```

@../toppar/epo/parallhdg5.3_PARAM19.pro
@../toppar/epo/epoa_20040915.pro
@../toppar/bac/BAC_learned.pro
nbonds
atom nbxmod -3 tolerance 0.5
repe1 0.75 rcon 4.0 rexp 2 cutnb 4.5
end
end

parameter {*Set energy parameters.*}
NBONds
  CUTNB=12 WMIN=1.5
  REPE1 = 0.0 cdie shift
  TOLerance = 0.5 NBXMOD = 5
  ctonnb=5.5 ctofnb=11.0
END
end

{ Seperate complexes and ligands }
vector do (x=x+100) (segid "T2 " or segid "L2 " or segid "L2F ")
vector do (y=y+50) (segid "L2F " or segid "L1F ")

!##### END GENERATE

CONStraints
  INTERaction ( segid T1 or segid L1 ) ( segid T1 or segid L1 )
    WEIGhts * 1. END
  INTERaction ( segid T2 or segid L2 ) ( segid T2 or segid L2 )
    WEIGhts * 1. END
  INTERaction ( segid L1F ) ( segid L1F ) WEIGhts * 1. END
  INTERaction ( segid L2F ) ( segid L2F ) WEIGhts * 1. END
END

!#####

! Define exchangeable=exchanged protons of Aminoacids

vector identify (store9) (not all)
vector identify (store9) (store9 or name "HN")
vector identify (store9) (store9 or (name HZ# and resname "LYS"))
vector identify (store9) (store9 or (name "HE" and resname ARG))
vector identify (store9) (store9 or (name HH## and resname ARG))
vector identify (store9) (store9 or (name HD1 and resname HIS))

```

```

vector identify (store9) (store9 or ( name HE2 and resname HIS ))
vector identify (store9) (store9 or (name HE1 and resname TRP))
vector identify (store9) (store9 or (name HG and resname CYS))
vector identify (store9) (store9 or (name HD2# and resname ASN))
vector identify (store9) (store9 or (name HG and resname SER))
vector identify (store9) (store9 or (name HG1 and resname THR))
vector identify (store9) (store9 or (name HH and resname TYR))
vector identify (store9) (store9 and hydrogen)

!Define INPHARMA active selections
vector identify (store1)
      (not store9 and segid L1F and hydrogen
       and (not name H003) and (not name H007))    ! LIG1 free
vector identify (store2)
      (not store9 and segid L2F and hydrogen
       and not name HA%)                          ! LIG2 free
vector identify (store3)
      (not store9 and segid T1 and hydrogen )      ! TAR1 bound
vector identify (store4) (not store9 and segid L1 and hydrogen
      and (not name H003) and (not name H007))    ! LIG1 bound
vector identify (store5)
      (not store9 and segid T2 and hydrogen )      ! TAR2 bound
vector identify (store6)
      (not store9 and segid L2 and hydrogen
       and not name HA%)                          ! LIG2 bound

#####

vector do (store7=0) (all)
vector do (store7=1)
      (not byres((resname EPOA or resname BAC) around 14.))
CONStraints fix (store7) end
vector do ( fbeta = 0. ) ( all )
vector do ( fbeta = 500. ) ( not hydrogen and not store7 )
vector do ( fbeta = 500. ) ( all )

vector do ( resid="1001" ) ( segid "L1F " )
vector do ( resid="1002" ) ( segid "L2F " )
vector do ( resid="2001" ) ( segid "L1 " )
vector do ( resid="2001" ) ( segid "L2 " )
noe
  reset
  nrestraints = 100000    ! allocate space for NOEs

```

```
ceiling 1000
end
```

```
noe class ambi @ambig_BAC.tbl @ambig_EPO.tbl end
noe
```

```
averaging * R-6
potential * soft
scale      * 1.0
sqconstant * 1.0
sqexponent * 2
soexponent * 1
rswitch    * 1.0
sqoffset   * 0.0
asymptote  * 2.0
msoexponent * 1
masymptote * -0.1
mrswitch   * 1.0
avexpo hbond 20
end
```

```
noe class unambi @unambig_EPO.tbl end
noe
```

```
averaging * sum
potential * soft
scale      * 100.0
sqconstant * 1.0
sqexponent * 2
soexponent * 1
rswitch    * 1.0
sqoffset   * 0.0
asymptote  * 2.0
msoexponent * 1
masymptote * -0.1
mrswitch   * 1.0
avexpo hbond 20
end
```

```
restraints
```

```
  dihedral
    nassign=10000
    scale 100.0
@dihe_EPO_bound.tbl
```

```
@dihe_EPO_free.tbl
end

set echo on message on end
INPHarma
    help
    tauf=160e-12
    taub=1270e-9

    k12=1000000
    k21=2850000
    CL10=100
    CL20=100
    CTO=1.79
    nue=900000000
    cutoff=10
    EXPS=0.99999
    FUNC PWEL 0.2 1 1
    FORCeconstant=1000000.0
MIXIngtimes
    ADD 0.0215
    ADD 0.0415
    ADD 0.0715
    ADD 0.1015
    ADD 0.2015
END

! Selection of active protons --> Setup hlist1, hlistl1, hlist2, hlistl2
ACTIVE
    (store1)! LIG1 free
    (store2)! LIG2 free
    (store3)! TAR1 bound
    (store4)! LIG1 bound
    (store5)! TAR2 bound
    (store6)! LIG2 bound

! Definition of detectable groups --> Setup of group list (one for both)
GROUP
    ADD "E21"    (segid L1F and name H21# and resname EPOA)
    ADD "E19"    (segid L1F and name H19  and resname EPOA)
    ADD "E27"    (segid L1F and name H27# and resname EPOA)
    ADD "E23"    (segid L1F and name H23# and resname EPOA)
    ADD "E22"    (segid L1F and name H22# and resname EPOA)
```

```

ADD "E25"    (segid L1F and name H25# and resname EPOA)
ADD "E3"     (segid L1F and name H03  and resname EPOA)
ADD "E7"     (segid L1F and name H07  and resname EPOA)
ADD "E12"    (segid L1F and name H12  and resname EPOA)
ADD "E13"    (segid L1F and name H13  and resname EPOA)
ADD "E15"    (segid L1F and name H15  and resname EPOA)
ADD "E17"    (segid L1F and name H17  and resname EPOA)
ADD "E24"    (segid L1F and name H24# and resname EPOA)
ADD "E25E22" (segid L1F and (name H25# or name H22#)
              and resname EPOA)
ADD "B16"    (segid L2F and name #H47  and resname BAC)
ADD "B17"    (segid L2F and name #H46  and resname BAC)
ADD "B0"     (segid L2F and (name 1H09 or name 1H05)
              and resname BAC)
ADD "BM"     (segid L2F and (name 1H08 or name 1H06)
              and resname BAC)
ADD "BP"     (segid L2F and name 1H07  and resname BAC)
ADD "B2"     (segid L2F and name 1H02  and resname BAC)
ADD "B3"     (segid L2F and name 1H10  and resname BAC)
ADD "B5"     (segid L2F and name 1H15  and resname BAC)
ADD "B6"     (segid L2F and name #H16  and resname BAC)
ADD "B7"     (segid L2F and name 1H17  and resname BAC)
ADD "B10"    (segid L2F and name 1H21  and resname BAC)
ADD "B13"    (segid L2F and name 1H26  and resname BAC)
ADD "B18"    (segid L2F and name #H44  and resname BAC)
ADD "B19"    (segid L2F and name #H19  and resname BAC)
ADD "B20A"   (segid L2F and name 1H14  and resname BAC)
ADD "B20B"   (segid L2F and name 2H14  and resname BAC)
ADD "B17B16" (segid L2F and (name #H47 or name #H46)
              and resname BAC)

end

!
! Definition of peakvolumes --> Setup of group list (one for both)
!
      VOLUmes
          @absolute_peaks_2004-07-13_B_XPLOR.inp
      end ! VOLUMES
end ! INPHARMA

flags  exclude inpharma include elec end

evaluate ( $filename = $id+"inph_0.calcforces_"
           +$file1+"_"+"$file2+" ".pdb")

```

```
write coor output=$filename end
flags exclude * include inph end

energy end

DISPLAY # CORRELATION 0 0_comp
INPH WRITe END

evaluate ($step=1)
evaluate ($lstep=1)

while ($lstep <= 1) loop mini
flags exclude * end
flags include elec vdw bond angle dihe impr inph noe cdihe end

    minimize powell
        drop=10
        nstep 50000
        debug=true
        nprint=1
        TOLGradient=0.000000001
    end

    evaluate ($step=$step + 1)
    evaluate ($lstep=$lstep + 1)
end loop mini

energy end

DISPLAY # CORRELATION 1 mini.init
INPH WRITe END

evaluate ( $filename = $id+"inph_mini.init"
          +$file1+"_"+"$file2+" ".pdb")
write coor output=$filename end

vector do (vx = maxw(150.)) (all)
vector do (vy = maxw(150.)) (all)
vector do (vz = maxw(150.)) (all)

vector do (mass = 100) (all)
evaluate ($bath = 150-2.5 )
```

```

while ($bath < 250) loop heat !350
  evaluate ($bath = $bath + 2.5 )

  dynamics verlet
  nstep=50

  timest=0.0001 {ps}
  iasvel=current
  tcoupling=true
  tbath=$bath
  nprint=10 iprfrq=10
  NTRFrq=100 !Frequency to remove center of mass motion
end
evaluate ($filename = "inph_" + $file1 + "_" + $file2 + "_"
          + encode ($step) + ".pdb")
vector do ( fbeta = 500. ) ( all )
evaluate ($step = $step + 1)
end loop heat

evaluate ($filename = $id + "inph_heated" + $file1 + "_" + $file2 + "_"
          + encode ($step) + ".pdb")
write coor output=$filename end

DISPLAY # CORRELATION 2 inph_heated
INPH WRITe END

evaluate ($timest=0.0001)
evaluate ($lstep=1)

while ($lstep <= 5) loop cool !40
  evaluate ($timest=0.0001*$lstep)
  evaluate ($bath = $bath - 1 )
  dynamics verlet
  nstep=20

  timest=$timest {ps}
  iasvel=current
  tcoupling=true
  tbath=$bath
  nprint=10 iprfrq=10
  NTRFrq=10 !Frequency to remove center of mass motion
end
evaluate ($filename = "inph_" + $file1 + "_" + $file2 + "_"

```

```

        + encode ($step) + ".pdb")
    evaluate ($step = $step + 1)
    evaluate ($lstep = $lstep + 1)
end loop cool

evaluate ($lstep=1)
while ($bath >= 35) loop cool !40
    evaluate ($bath = $bath - 1 )
    dynamics verlet
    nstep=20

    timest=$timest {ps}
    iasvel=current
    tcoupling=true
    tbath=$bath
    nprint=10 iprfrq=10
    NTRFrq=10 !Frequency to remove center of mass motion
end
    evaluate ($filename = "inph_" + $file1 + "_" + $file2 + "_"
        + encode ($step) + ".pdb")
    evaluate ($step = $step + 1)
    evaluate ($lstep = $lstep + 1)
end loop cool
evaluate ($lstep = 1)

evaluate ($filename = $id + "inph_cooled" + $file1 + "_" + $file2 + "_"
    + encode ($step) + ".pdb")
write coor output=$filename end

DISPLAY # CORRELATION 3 inph_cooled
INPH WRITe END
evaluate ($lstep = 1)
while ($lstep <= 50) loop mini !50
    minimize powell
        drop=0.0001
        nstep 5
        debug=false
        nprint=1
        TOLGradient=0.000001
    end
    energy end
    evaluate ($filename = "inph_mini_steps_inph_on_"
        + $file1 + "_" + $file2 + "_"

```

```

        + encode ($step) + ".pdb")
    evaluate ($drop = $drop )
    evaluate ($step=$step + 1)
    evaluate ($lstep=$lstep + 1)
end loop mini

evaluate ($filename = $id+"inph_final_mini_inph_on"
        +$file1+"_" +$file2+".pdb" )
write coor output=$filename end

DISPLAY # CORRELATION 4 final_mini_inph_on
INPH WRITe END

flags exclude inph end

evaluate ($lstep = 1)
while ($lstep <= 1) loop mini  !50
    minimize powell
        drop=0.0001
        nstep 10000
        debug=true
        nprint=1
        TOLGradient=1
    end
    energy end
    evaluate ($filename = "inph_mini_steps_inph_off_"
            +$file1+"_" +$file2+"_"
            + encode ($step) + ".pdb")
    evaluate ($drop = $drop )
    evaluate ($step=$step + 1)
    evaluate ($lstep=$lstep + 1)
end loop mini

evaluate ($filename = $id+"inph_final_mini_inph_off"
        +$file1+"_" +$file2+".pdb" )
write coor output=$filename end

flags include inph end

energy end
DISPLAY # CORRELATION 5 final_mini_inph_off
INPH WRITe END

```

stop

- INPHarma library - TRIO program, link - TRIO program tutorial



# Appendix II

## DNP



# Chapter 11

## Contributions

In the following my personal contributions to the published papers on DNP are listed.

- P. Höfer, G. Parigi, C. Luchinat, P. Carl, G. Guthausen, Reese M., T. Carlomagno, C. Griesinger, and M. Bennati. Field Dependent Dynamic Nuclear Polarization with Radicals in Aqueous Solution. *Journal of the American Chemical Society*, 130:3254-3255, 02 2008. (ref. [61])

I was involved in the nitroxide sample preparation and measured  $T_1$  relaxation of water at high field. These measurements could not be used due to the influence of excessive radiation damping. I was involved in correcting the manuscript of the paper.

- P. Höfer, P. Carl, G. Guthausen, T. Prisner, M. Reese, T. Carlomagno, C. Griesinger, and M. Bennati. Studies of Dynamic Nuclear Polarization with Nitroxides in Aqueous Solution. *Applied Magnetic Resonance*, 34:393-398, 2008. (ref. [62])

I was involved in selecting and purchasing the used nitroxide radicals. I prepared the nitroxide sample solutions and assisted the X-band DNP experiments with single and double irradiation of the EPR resonances. I was involved in writing and correcting the manuscript of the paper.

- M. Reese, D. Lennartz, T. Marquardsen, P. Höfer, A. Tavernier, P. Carl, T. Schippmann, M. Bennati, T. Carlomagno, F. Engelke, and C. Griesinger. Construction of a Liquid-State NMR DNP Shuttle DNP Spectrometer: First Experimental Results and Evaluation of Optimal Performance Characteristics. *Applied Magnetic Resonance*, 34:301-311, 2008. (ref. [63])

A preliminary sample shuttling setup was existing in Göttingen. This was meant for initial feasibility tests and field cycling relaxometry measurements using a 600 MHz Bruker NMR spectrometer. In collaboration with Thorsten Marquardsen I tested and developed sample containers withstanding the shuttle process. The containers were partially manufactured by Bruker and at the MPIBPC workshop. I select suitable glass tubing and adapted existing cap constructions. I tested the existing shuttle DNP probe head and existing shuttle pneumatics system. The observed tuning and matching instabilities during shuttling were solved by using more stable capacitors and a single NMR coil. This modifications were performed by Thorsten Marquardsen. Simple repairs of the shuttle probe head were done by myself. An extended version of the shuttle transfer system was necessary to allow for shuttling over the distance of 1.5 m. This version was built in the MPIBPC workshop according to my specifications. To reduce the influence of sample impact to the NMR probe I designed the shock transfer system, comprising a carbon fiber rod, a shock transfer cone, a dampening foam cushion an a heavy support table. This system was manufactured in the MPIBPC workshop. I selected, ordered and evaluated the function of different dampening materials like viscoelastic gels and viscoelastic gel-foams. I measured the respective shuttle times and rebound characteristics, using data from the two light barriers implemented in the shuttle probe head and the transfer system. I used this setup for stability and reproducibility measurements with 1D and 2D spectroscopy. I used the setup to measure stray field relaxometry of protons and carbons of various samples like glucose,  $^{13}\text{C}$  labeled glucose, ubiquitin and water.

I used these measurements, the collective relaxation model of Luchinat and Parigi,<sup>71</sup> the stray field maps of the 0.34 T permanent magnet and the 14.09 T superconducting magnet, to predict the possible enhancement factors for various possible experimental shuttle DNP setups. Different sample transfer characteristics, based on the shuttle pressure and sample geometry were considered. I drafted alternative setups, using an axillary transfer field, provided by an cylindrical, tubular array of neodymium Hallbach magnets. I simulated the possible transfer fields of such arrays, based on the material characteristics of the magnets, for different geometries. Depending on the outer diameter of either 54 mm or 27 mm this assembly was found to provide up to 2.72 T or 1.92 T, respectively. The simulations indicated that even these elevated fields could not significantly reduce relaxation losses during sample transfer, as most of the losses still occurred in the vicinity of the 0.34 T low-field magnet where the sample was still slow. To solve

this problem I simulated the benefit of an additional pulsed coil in the vicinity of the permanent magnet to provide a field of 1 T in about 4.3 ms. This field strength was found to be insufficient to reduce the relaxation losses sufficiently. Furthermore, the effects of rotating the main magnetic field during sample transfer were simulated. This was necessary as the the low-field and high-field magnet field directions were perpendicular to each other. It was found that the magnetization of the sample would rotate adiabatically during sample transfer, always staying parallel to the effective field.

The simulations led to the conclusion that the planned system, using two dedicated magnet of 0.34 T and 14.09 T (not using any transfer fields) was sufficient for initial tests of the concept of shuttle DNP. It became clear that the effective shuttle DNP enhancement of larger molecules would require a shorter shuttle distance and a higher field on the sample track, as later demonstrated by the ferro-shim shuttle DNP system of [65].

I was involved in initial tests of the shuttle-DNP spectrometer, based at the Bruker site in Karlsruhe. For this setup a second version of the shock transfer system was built at the MPIBPC. The experiments in Karlsruhe demonstrated the need for better control of the sample transfer. The sample had to be slowed down before impact to reduce the residual rebound behavior after impact. I motivated the construction of an alternative pneumatic shuttle system, using two dedicated valves at the top and the bottom of the transfer system, for sample deceleration prior to impact. This design was implemented by Bruker under direction of Thorsten Marquardsen and Frank Engelke. I motivated the construction of a system to measure the sample position during the sample transfer. Thorsten Marquardsen suggested the use of the light barrier array, which he implemented.

I measured the published low-field DNP and shuttle DNP experiments on water with 5 mM TEMPONE-D, $^{15}\text{N}$ , also involving the detection of post-shuttle magnetization recovery. I performed the data analysis of these experiments. I performed the respective simulations which could explain the observed post-shuttle enhancements. These simulations were based on the known initial magnetization, the measured sample transfer characteristics and the field-cycling relaxometry data by Luchinat and Parigi. I used this validated simulation model to determine the experimental limits for shuttle DNP in the present setup. Furthermore, I calculated the theoretical limit for any shuttle DNP spectrometer, based on the conservation of magnetization during sample transfer, assuming a pure dipolar DNP mechanism.

I prepared the first version of the manuscript, prepared the figures and was involved in and coordinated the correction of the manuscript.

- M. Reese, M.-T. Türke, I. Tkach, G. Parigi, C. Luchinat, T. Marquardsen, A. Tavernier, P. Höfer, F. Engelke, C. Griesinger, and M. Bennati.  $^1\text{H}$  and  $^{13}\text{C}$  Dynamics Nuclear Polarization in Aqueous Solution with a Two-Field (0.34 T/14 T) Shuttle DNP Spectrometer. *Journal of the American Chemical Society*, 131:15086-15087, 2009. (ref. [64])

The previous DNP setup did not provide reproducible shuttle DNP enhancement with enhancement factors  $|\epsilon_{\text{HF}}| \geq 1$ . It suffered from unreliable sample sealing. I suggested sealing of the samples in capillaries, using UV glue. I selected and purchased glues of different quality and from different vendors. Initial experiments were promising but the handling and reliable sealing of the samples was still difficult due to incompatibility of the epoxy-based glues with water. Thorsten Marquardsen purchased further glues and a industry scale UV lamp for further tests. Finally, we found the used acryl-based glue to be very suitable.

I motivated further optimization of the shuttle hardware. This includes the use of a vacuum-capable lower pneumatics valve which allowed to suck the sample against the shock transfer rod after sample impact. This resulted in reduced or avoided rebound. I constructed a copper spring system for quick post-impact retraction of the shock transfer rod, allowing for a shorter post-shuttle stabilization period and higher NMR reproducibility. This system was built at the MPIBPC. I performed the published shuttle DNP experiment on water. And performed initial DNP experiments using  $^{13}\text{C}$  urea, glucose- $^{13}\text{C}$  and glycine- $^{13}\text{C}$ .

I measured the sample transfer speed using helium as a driving gas, reducing the shuttling time to 50%, compared to nitrogen.

I was involved in sample selection and initial experiments for further  $^{13}\text{C}$  DNP experiments, also including organic solvents. The published, final experiments were performed by Maria-Teresa Türke and Igor Tkach.

I performed the simulation of relaxation losses for urea- $^{13}\text{C}$  shuttle DNP experiments (based on the the relaxometry data by Luchinat and Parigi) and for future experiments using proteins (based on the collective relaxation model by Luchinat and Parigi). I calculated the maps of relaxation rates in the sample transfer system depending on molecular size and sample position.

I contributed significantly to writing the manuscript and to later corrections.

- A. Krahn, P. Lottmann, T. Marquardsen, A. Tavernier, Türke M.-T., M. Reese, A. Leonov, M. Bennati, P. Höfer, F. Engelke, and C. Griesinger. Shuttle DNP Spectrometer with a Two-Center Magnet. *Physical Chemistry Chemical Physics*, 2010, submitted. (ref. [65])

I was involved in defining the design requirements for the new shuttle DNP setup, comprising the ferro-shim system and a newly developed low-field DNP probe head. I put special attention to requirement of temperature control and to allowing for the later implementation of a mechanical shuttle device. This device should allow for shuttling sample capillaries and should avoid the need for a large sample container. This will allow the use of smaller high-resolution NMR coils, providing a higher mass sensitivity.

Together with Philip Lottmann, I performed initial shuttle DNP experiments and introduced the new diploma student (Philip Lottmann) to DNP.

I was involved in correction and modifications of the manuscript.

- M.-T. Türke, I. Tkach, M. Reese, P. Höfer, and M. Bennati. Optimization of Dynamic Nuclear Polarization Experiments in Aqueous Solution at 15 MHz/9.7 GHz: A Comparative Study with DNP at 140 MHz/94 GHz. *Physical Chemistry Chemical Physics*, 2010, submitted. (ref. [66])

I designed the circulator-based back-route to allow for high-power measurements of the reflection coefficient of the EPR resonator under DNP conditions (X-band). I discussed alternative possibilities with Andreas Tavernier (Bruker). I selected and purchased the required components. I participated in the initial hardware assembly.

I tested or participated in testing of several options to validate the linearity of the minispec receiver. This includes test with DNP polarized samples, external reference sources and also using reference signals provided by the output of the minispec itself.

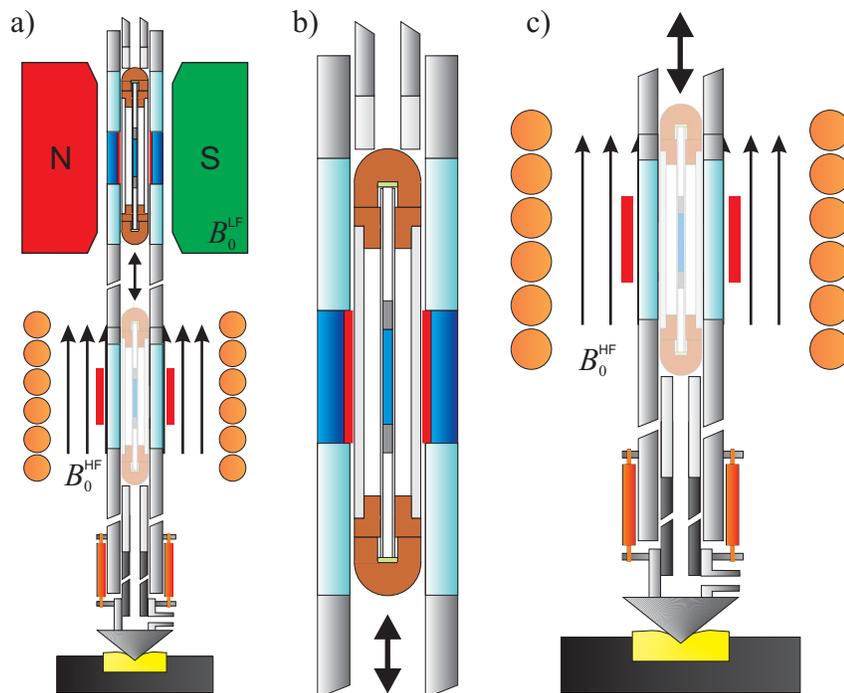
I was involved in method development for the  $B_1$  field calibration in a critically coupled X-band resonator. Initial experiments were done with N@C60 buckminsterfullerenes, comprising a single nitrogen atom in their center. I obtained the N@C60 from Dr. Wolfgang

Harneit, Hahn-Meitner-Institut, Berlin, Germany. I was involved in initial pulse EPR Hahn-echo experiments showing still detectable signal, but allover too low sensitivity. This was due to the low nitrogen content ( $1.3 \cdot 10^{-4}$ ) of the substance.

I was involved in initial, power-depended EPR nutation experiments using the trityl radical TAM and a critically coupled resonator.

# Chapter 12

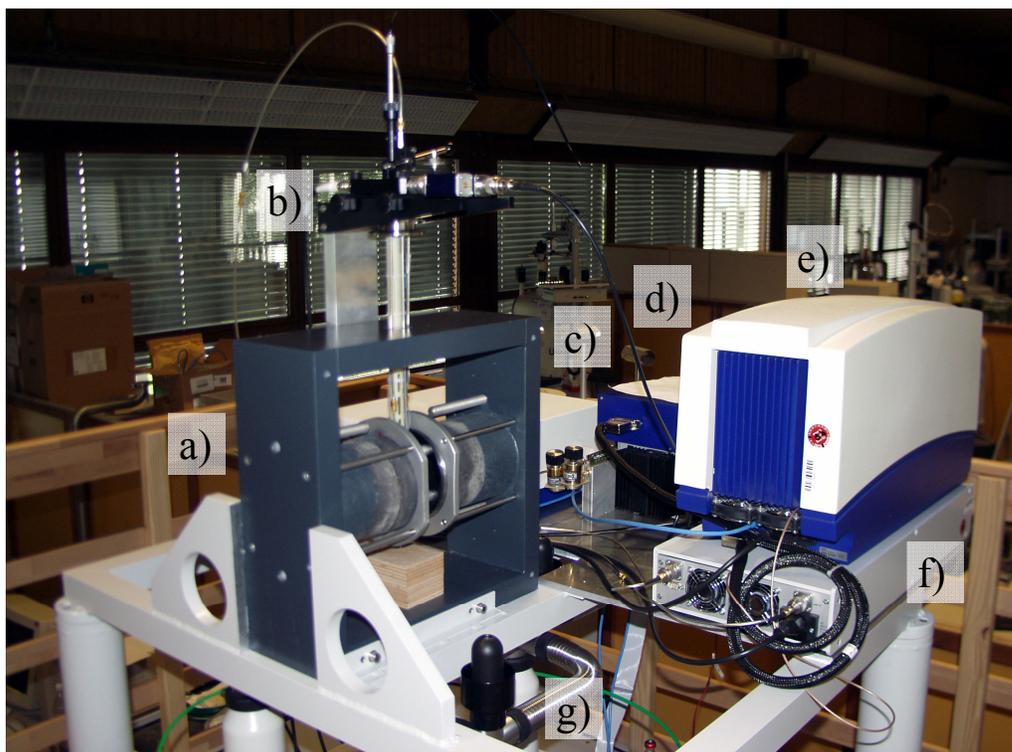
## Supplementary Figures



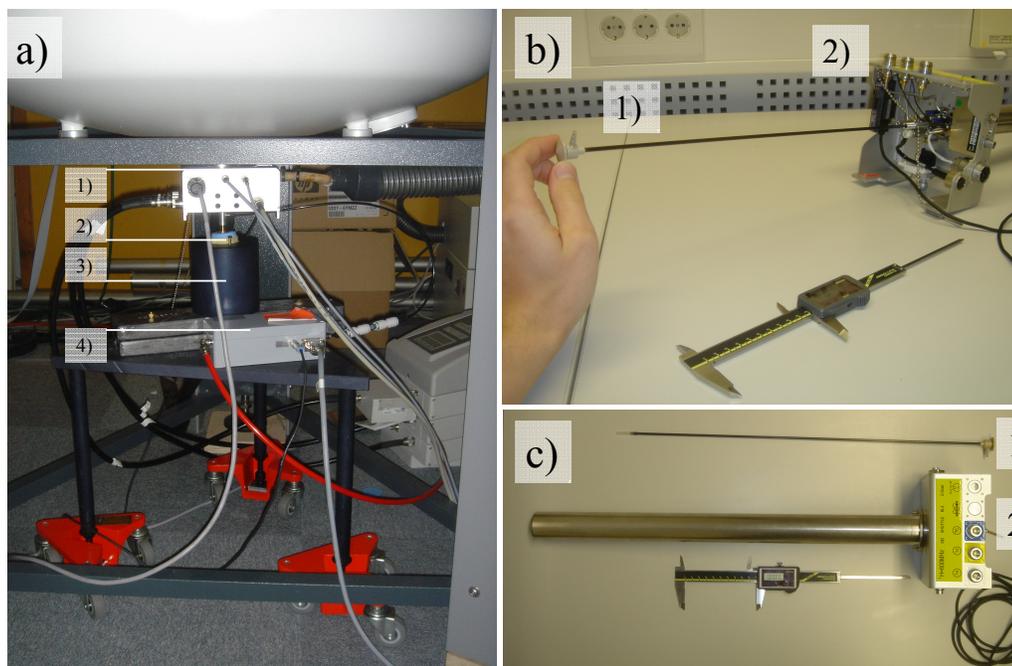
**Figure 12.1:** Sketch of the shuttle-DNP setup using two dedicated magnets.<sup>63,64</sup>  
 a) Overview with the low-field part on the top and the high-field part at the bottom of the figure. The superconducting magnet is depicted by cross sections through the wires (orange circles), the permanent magnet is depicted with its red and green pole shoes. b) Low-field ENDOR resonator, with the NMR coils (red), sapphire dielectric (dark blue) and glass tubing (light blue). The sample container is made from Vespel (brown), glass (light gray), silicon (yellowish green) and acrylic UV glue (dark gray). c) High-resolution part of the shuttle system with the NMR coil (red), the carbon-fiber shock transfer rod (black), comprising an Kel-F impact tube (white) and a PEEK shock-transfer cone (gray). The copper springs (orange) are used for retracing the shock-transfer rod after energy transfer to the gel-foam cushion (yellow) which is mounted on a heavy support table (black).



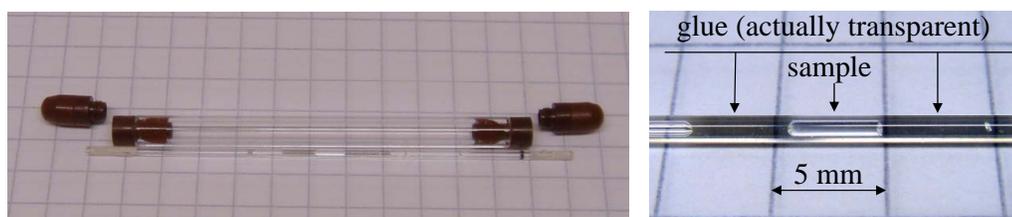
**Figure 12.2:** Photograph of the shuttle-DNP setup using two dedicated magnets.<sup>63,64</sup>



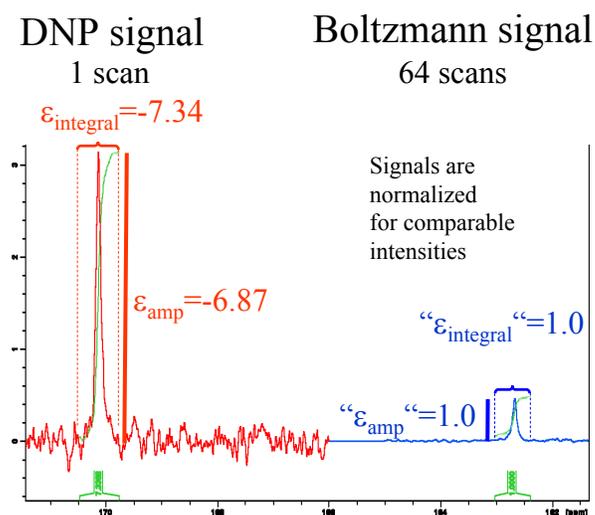
**Figure 12.3:** Photograph of the low-field assembly of the shuttle-DNP setup using two dedicated magnets.<sup>63,64</sup> a) 0.34 T permanent magnet, b) ENDOR probe head, c) e-scan Bruker EPR bridge, d) NMR duplexer and preamplifier, e) Bruker minispec NMR spectrometer, f) 5 W X-band amplifier, g) Bruker ultrashield 600 MHz high-resolution NMR magnet



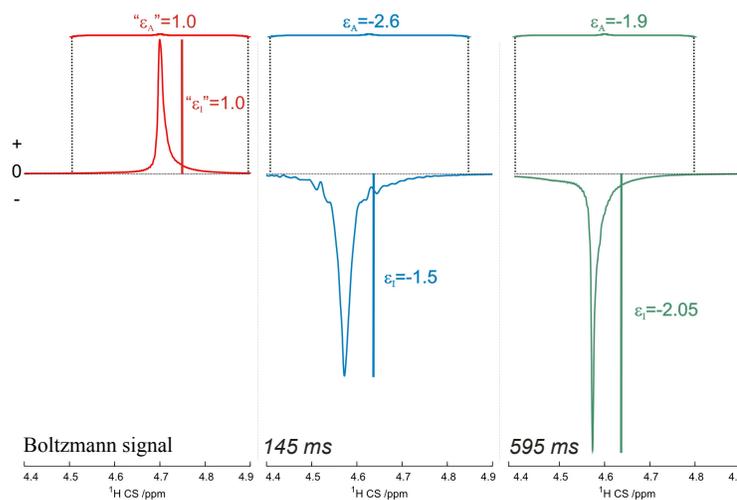
**Figure 12.4:** High resolution probe with shock transfer system. a) Assembled in the high-resolution NMR magnet: 1) NMR probe head, 2) gel foam cushion with shock-transfer cone 3) heavy support table, 4) lower pneumatics valve satellite. b,c) NMR probe (2) with shock transfer rod (1).



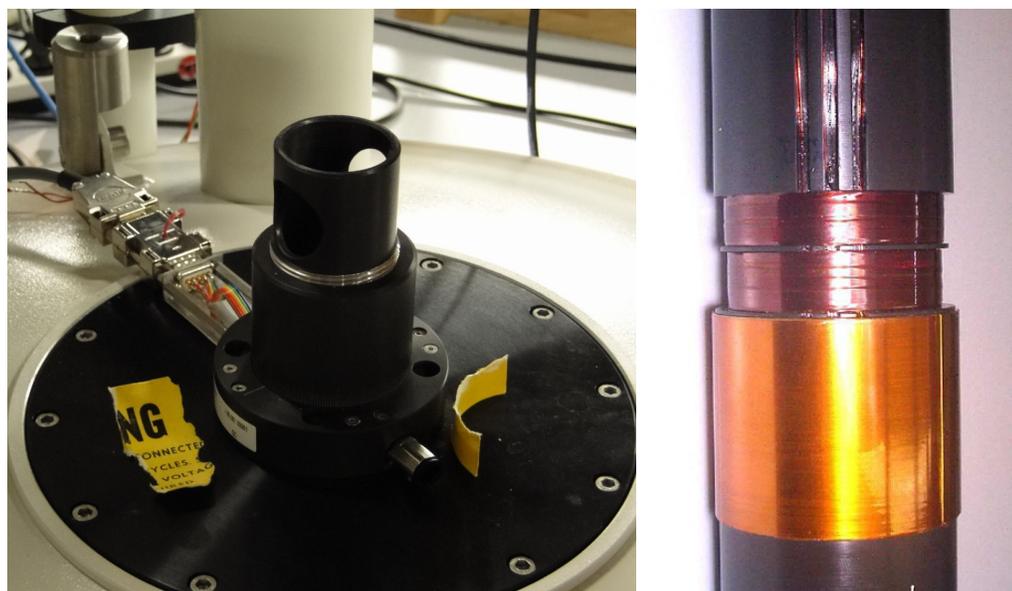
**Figure 12.5:** Photograph of the sample container and sample capillary as used in [64]. The sample is sealed in the capillary by UV glue. A similar container is used in [65].



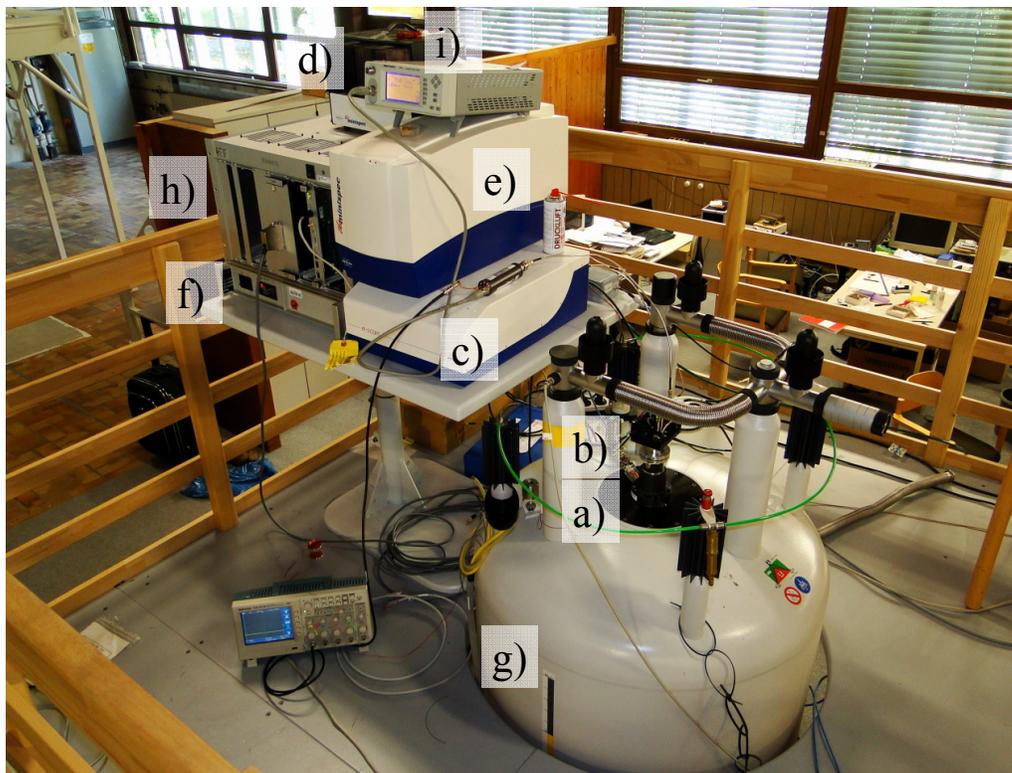
**Figure 12.6:**  $^{13}\text{C}$  subtle-DNP enhanced signal of 4M  $^{13}\text{C}$  UREA in  $\text{D}_2\text{O}$  in comparison to the Boltzmann signal. The radical concentration was 100 mM of TEMPONE-D, $^{15}\text{N}$ . The sample length was 10 mm and the diameter 0.46 mm. The microwave power of 13 W was applied for 10 s. Detection was performed 250 ms after DNP and the sample transfer time was  $\approx 140$  ms. The DNP signal was recorded with a single scan while the Boltzmann signal was recorded in 64 scans. The signals are scaled for comparable intensities (the DNP signal is actually inverted by sign). The integral and amplitude high-field enhancement factors are shown. The solvent was  $\text{D}_2\text{O}$ . This measurement was not included in the previous publication concerning  $^{13}\text{C}$ -DNP.<sup>64</sup>



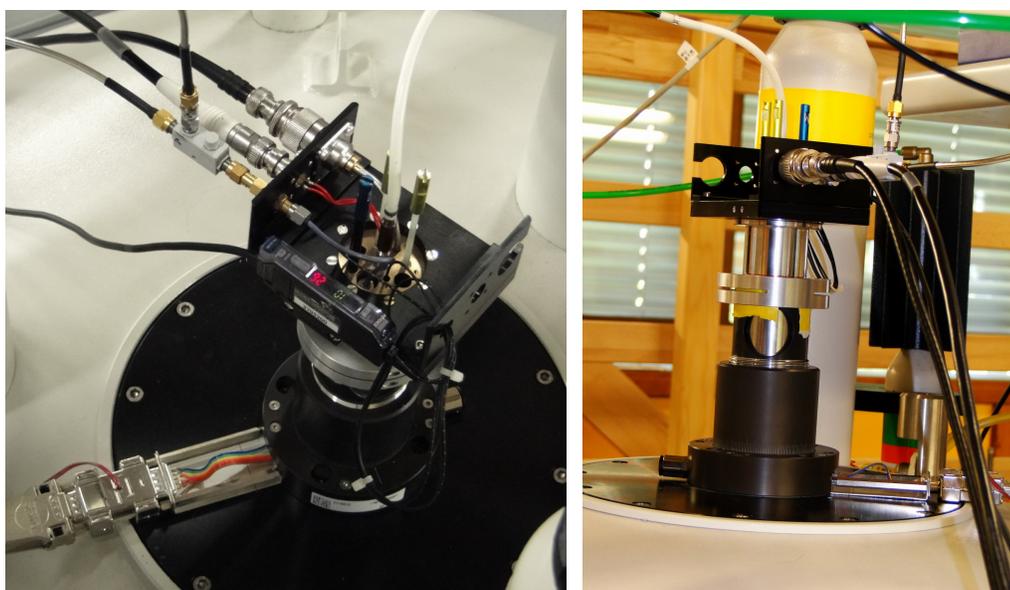
**Figure 12.7:** Shuttle DNP enhancement of water using the shuttle DNP setup with two dedicated magnets [64, fig. 1c]. The sample was water doped with 5 mM TEMPONE-D,<sup>15</sup>N and the experimental parameters were: 3 s microwave irradiation at 5 W, 0.9 mm sample diameter, single scan acquisition. The indicated times of 145 ms and 595 ms are the delays between leaving of the sample at the low-field position and the start of high-resolution NMR acquisition.



**Figure 12.8:** Photograph of the inserted ferro-shim shim system<sup>65</sup> (left) and the a close up view of the system in the vicinity of the sample area and the homogeneous field position. The upper gradient and field sweep coils are visible. The other pair of coils is placed below the lower part of the copper-colored sheet. The sample position is just below the upper edge of the sheet.



**Figure 12.9:** Photograph of the low-field part of the ferro-shim shuttle DNP setup.<sup>65</sup> Compared to the previous setup the permanent magnet is missing. a) ferro-shim system, b) DNP probe, c) e-scan EPR bridge, d) NMR duplexed and preamplifier, e) Bruker minispec NMR spectrometer, f) 40 W X-band amplifier, g) high-resolution NMR magnet, h) Avance-III Bruker BSMS shim system for controlling the low-field gradient coils of the ferro-shim system, i) power meter for monitoring the microwave power. Later, the minispec was removed as NMR at the low-field position was performed using the high-resolution NMR console.



**Figure 12.10:** Photograph of the DNP probe head.<sup>65</sup> Here it is still equipped with an NMR coil. Later the coil was removed. Here, matching of the EPR cavity was still performed with a screw (blue). In later setups a linear motor was used for this purpose.



# Bibliography

- [1] Schwalbe H., Stilz H.U., and Kessler H. NMR Spectroscopy of Biomacromolecules in Drug Discovery and Beyond. *ChemBioChem*, 6:1475–1478, 2005. 3, 107
- [2] Pellecchia, M. and Bertini, I. and Cowburn, D. and Dalvit, C. and Giralt, E. and Jahnke, W. and James, T.L. and Homans, S.W. and Kessler, J. and Luchinat, C. and Meyer, B. and Oschkinat, H. and Peng, J. and Schwalbe, H. and Siegal, G. Perspectives on NMR in drug discovery: a technique comes of age. *Nature*, 7:738–745, 2008. 3, 107
- [3] C. Ludwig and U.L. Guenter. Ligand based NMR methods for drug discovery. *Frontiers in Bioscience*, 14:4565–4574, 2009. 3, 107
- [4] Jeener J., B.H. Meier, P. Bachmann, and R.R. Ernst. Investigation of exchange processes by 2-dimensional NMR-spectroscopy. *Journal of Chemical Physics*, 71(11):4546–4553, 1979. 3, 12, 107
- [5] J.P. Albrand, B. Birdsall, J. Feeney, G.C.K. Roberts, and A.S.V. Burgen. The use of transferred nuclear overhauser effects in the study of the conformations of small molecules bound to proteins. *International Journal of Biological Macromolecules*, 1(1):37 – 41, 1979. 5, 107
- [6] P.J. Cayley, J.P. Albrand, J Feeney, G.C.K. Robert, E.A. Piper, and A.S.V. Burgen. Nuclear magnetic-resonance studies of the binding of trimethoprim to dihydrofolate-reductase. *Biochemistry*, 18(18):3886–3895, 1979. 5, 107
- [7] G.M. Clore and A.M. Gronenborn. Theory and applications of the transferred nuclear Overhauser effect to the study of the conformations of small ligands bound to proteins. *Journal of Magnetic Resonance*, 48(3):402–417, 1982. 5, 107, 109
- [8] D.W. Li, E.F. DeRose, and R.E. London. The inter-ligand Overhauser effect: A powerful new NMR approach for mapping structural rela-

- tionships of macromolecular ligands. *Journal of Biomolecular NMR*, 15(1):71–76, 1999. 5, 8, 107
- [9] R.E. London. Theoretical analysis of the Inter-Ligand Overhauser Effect: A new approach for mapping structural relationships of macromolecular ligands. *Journal of Magnetic Resonance*, 141(2):301–311, 1999. 5, 16, 107
- [10] V.M. Sanchez-Pedregal, M. Reese, J. Meiler, J.J. Blommers Marcel, C. Griesinger, and T. Carlomagno. The INPHARMA Method: Protein-Mediated Interligand NOEs for Pharmacophore Mapping. *Angewandte Chemie International Edition*, 44(27):4172–4175, 2005. 5, 8, 12, 63, 107
- [11] M. Reese. Entwicklung der INPHARMA Methode und Anwendung auf den Tubulin - Epothilon A Komplex, Diplomathesis, University of Göttingen, Germany, 2005. 5, 7, 8, 15, 16, 18, 30, 64, 65, 91, 107
- [12] M. Reese, V.M. Sanchez-Pedregal, K. Kubicek, J. Meiler, M.J.J. Blommers, C. Griesinger, and T. Carlomagno. Structural basis of the activity of the microtubule-stabilizing agent epothilone A studied by NMR spectroscopy in solution. *Angewandte Chemie - International Edition*, 46(11):1864–1868, 2007. 5, 8, 15, 16, 18, 30, 63, 64, 65, 74, 76, 77, 91, 107, 180
- [13] J. Orts, J. Tuma, M. Reese, S.K. Grimm, P. Monecke, S. Bartoschek, A. Schiffer, K.U. Wendt, C. Griesinger, and T. Carlomagno. Crystallography-independent determination of ligand binding modes. *Angewandte Chemie-International Edition*, 47(40):7736–7740, 2008. 5, 8, 16, 107, 108
- [14] J. Orts, C. Griesinger, and T. Carlomagno. The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method. *Journal of Magnetic Resonance*, 200(1):64–73, 2009. 5, 7, 8, 9, 107
- [15] S. Macura and R.R. Ernst. Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy. *Molecular Physics*, 41(1):95–117, 1980. 5, 6
- [16] N.R. Krishna, H.N.B. Moseley, and E.V. Curto. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of NOESY Spectra - Application to transferred NOESY and Protein-Folding Studies. *Journal of Cellular Biochemistry*, 21(21):74, 1995. 5

- [17] M. Mayer and B. Meyer. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie - International Edition*, 38(12):1784–1788, 1999. 5, 109
- [18] C. Dalvit, P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti, and M. Sundstrom. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *Journal of Biomolecular NMR*, 18(1):65–68, 2000. 5, 109
- [19] M.E. Perlman, D.G. Davis, G.W. Koszalka, J.V. Tuttle, and R.E. London. Studies of Inhibitor Binding to Escherichia-Coli Purine Nucleoside Phosphorylase using the Transferred Nuclear Overhauser Effect and Rotating-Frame Nuclear Overhauser Enhancement. *Biochemistry*, 33(24):7547–7559, 1994. 5, 7
- [20] H. N.B Moseley, W. Lee, C.H. Arrowsmith, and N.R. Krishna. Quantitative determination of conformational, dynamic, and kinetic parameters of a ligand-protein/DNA complex from a complete relaxation and conformational exchange matrix analysis of intermolecular transferred NOESY. *Biochemistry*, 36(18):5293–5299, 1997. 5, 7
- [21] W. Braun, C. Bosch, L.R. Brown, N. Go, and K. Wüthrich. Combined use of Proton-Proton Overhauser Enhancements and a Distance Geometry Algorithm for Determination of Polypeptide Conformations - Application to Micelle-bound Glucagon. *Biochimica et Biophysica ACTA*, 667(2):377–396, 1981. 6
- [22] M. Nilges, J. Habazettl, A.T. Brünger, and T.A. Holak. Relaxation matrix Refinement of the Solution Structure of Squash Trypsin-Inhibitor. *Journal of Molecular Biology*, 219(3):499–510, 1991. 6, 19, 20, 25, 109
- [23] J. Mertz, P. Güntert, K. Wüthrich, and W. Braun. Complete relaxation matrix refinement of nmr structures of proteins using analytically calculated dihedral angle derivatives of noe intensities. *Journal of Biomolecular NMR*,, 1:257–269, 1991. 6
- [24] A.T. Brünger. X-PLOR Version 2.1, 1990. 6
- [25] B. Gidas. Global optimization via the Langevin equation. *Proceedings of the 24th IEEE Conference on Decision and Control (Cat. No.85CH2245-9)*, 2-3:774–8, 1985. 6, 63
- [26] B. Hajek. Cooling schedules for optimal annealing. *Mathematics of Operations Research*, 13(2):311–329, 1988. 6, 63, 86

- [27] P.F. Yip. Calculating NOESY intensities by perturbation expansion. *Chemical Physics Letters*, 161(1):50–54, 1989. 6
- [28] P.F. Yip and D. Case. A new method for refinement of macromolecular structures based on nuclear overhauser effect spectra. *Journal of Magnetic Resonance*, 83:643–648, 1989. 6, 7, 23
- [29] P.F. Yip. A computationally efficient method for evaluating the gradient of 2D NOESY intensities. *Journal of Biomolecular NMR*, 2:361–365, 1993. 6, 109
- [30] E.N. Nesterova and Chuprina V.P. An efficient method for calculating analytical derivatives for direct noe refinement of macromolecular structures. *Journal of Magnetic Resonance. Series B*, 101:94–96, 1993. 7, 19, 20, 24, 25
- [31] V. Jayalakshmi and N.R. Krishna. Determination of the conformation of trimethoprim in the binding pocket of bovine dihydrofolate reductase from a std-nmr intensity-restrained corcema-st optimization. *Journal of the American Chemical Society*, 127:14080–14084, 2005. 7
- [32] M.A. Macnaughtan, M. Kamar, G. Alvarez-Manilla, A. Venot, J. Glushka, J. Michael Pierce, and Prestegard J.H. NMR Structural Characterization of Substrates Bound to N-Acetylglucosaminyltransferase V. *Journal of Molecular Biology*, 366(4):1266–1281, 2007. 7
- [33] A. Okada, K. Wakamatsu, T. Miyazawa, and T. Higashijima. Vesicle-bound conformation of melittin: transferred nuclear Overhauser enhancement analysis in the presence of perdeuterated phosphatidylcholine vesicles. *Biochem.*, 33(32):9438–9446, 1994. 7
- [34] G.K. Jarori, N. Murali, and B.D.N. Rao. Two-dimensional transferred nuclear Overhauser effect spectroscopy study of the confirmation of MgATP bound at the active and ancillary sites of rabbit muscle pyruvate kinase. *Biochemistry*, 33(22):6784–6791, 1994. 7
- [35] V.N.S. Sanchez-Pedregal, K. Kubicek, J. Meiler, I. Lyothier, I. Paterson, and T. Carlomagno. The Tubulin-Bound Conformation of Discodermolide Derived by NMR Studies in Solution Supports a Common Pharmacophore Model for Epothilone and Discodermolide. *Angewandte Chemie - International Edition*, 45:7388–7394, 2006. 7
- [36] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, and G.M. Clore. The Xplor-NIH NMR Molecular Structure Determination Package. *Journal of Magnetic Resonance*, 160:66–74, 2003. 12, 108

- [37] C.D. Schwieters, J.J. Kuszewski, and G.M. Clore. Using Xplor-NIH for NMR molecular structure determination. *Progress in NMR Spectroscopy*, 48:47–62, 2006. 12, 108
- [38] M.J.D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal*, 7:152–162, 1964. 25
- [39] O. Ermer, S.A. Mason, F.A.L. Anet, and Miura S.S. Ultrashort non-bonded hydrogen...hydrogen distance in a half-cage pentacyclododecane. *Journal of the American Chemical Society*, 107(8):2330–2334, 1985. 52
- [40] C. Dominguez, R. Boelens, and A.M.J.J. Bovin. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *JACS*, 125:1731–1737, 2003. 64
- [41] J. Löwe, H. Li, K.H. Downing, and E. Nogales. Refined structure of Alpha-Beta Tubulin at 3.5Å resolution. *Journal of Molecular Biology*, 313(5):1045–1057, 2001. 64
- [42] T. Carlomagno, V.M. Sanchez, M.J.J. Blommers, and C. Griesinger. Derivation of dihedral angles from CH-CH dipolar-dipolar cross-correlated relaxation rates: A C-C torsion involving a quaternary carbon atom in epothilone A bound to tubulin. *Angewandte Chemie - International Edition*, 42(22):2515–2517, 2003. 64
- [43] H.J.C. Berendsen, J.P.M. Postma, N.F. van Gunsteren, A. DiNola, , and J.R. Haak. Molecular dynamics with coupling to an external bath . *Journal of Chemical Physics*, 81:3684–3690, 1984. 86
- [44] S. Lahabar and P.J. Narayanan. Singular Value Decomposition on GPU using CUDA. *2009 IEEE International Symposium on Parallel & Distributed Processing*, 1-5:840–849, 2009. 109
- [45] NVidia Corporation. Compute unified device architecture (cuda). 109
- [46] M. Feldman. NVIDIA Unveils Teraflop GPU Computing. *HPC wire*, 2008. 109
- [47] A.W. Overhauser. Polarization of nuclei in metals. *Physical Review*, 91(2):476–476, 1953. 113
- [48] T.R. Carver and C.P. Slichter. Polarization of nuclear spins in metals. *Physical Review*, 92(1):212–213, 1953. 113

- [49] L.R. Becerra, G.J. Gerfen, Temkin R.J., Singel D.J., and Griffin R.G. Dynamics nuclear polarization with a cyclotron-resonance maser at 5 T. *Physical Review Letters*, 71(21):3561–3564, 1993. 114
- [50] P.C.A. van der Wel, K.N. Hu, J. Lewandowski, and R.G. Griffin. Dynamic nuclear polarization of amyloidogenic peptide nanocrystals: GN-NQQNY, a core segment of the yeast prion protein Sup35p. *Journal of the American Chemical Society*, 128(33):10840–10846, 2006. 114
- [51] V. Weis, M. Bennati, M. Rosay, and R.G. Griffin. Solid effect in the electron spin dressed state: A new approach for dynamic nuclear polarization. *Journal of Chemical Physics*, 113(16):6795–6802, 2000. 114
- [52] YohMatsuki, T. Maly, O. Ouari, H. Karoui, F. Le Moigne, E. Rizzato, S. Lyubenova, J. Herzfeld, T. Prisner, Tordo P., and Griffin R.G. Dynamic Nuclear Polarization with a Rigid Biradical. *Angewandte Chemie - International Edition*, 48:4996–5000, 2009. 114
- [53] J.H. Ardenkjaer-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M.H. Lerche, R. Servin, M. Thaning, and K. Golman. Increase in signal-to-noise ratio of  $> 10,000$  times in liquid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18):10158–10163, 2003. 114, 137
- [54] S. Bowen and H. Christian. Time-resolved dynamic nuclear polarization enhanced NMR spectroscopy. *Angewandte Chemie - International Edition*, 47(28):1433–7851, 2008. 115
- [55] L. Frydman, A. Lupulescu, and T. Scherf. Principles and Features of Single-Scan Two-Dimensional NMR Spectroscopy. *Journal of the American Chemical Society*, 125:9204–9217, 2003. 115
- [56] C.G. Joo, K.N. Hu, J.A. Bryant, and R.G. Griffin. In situ temperature jump high-frequency dynamic nuclear polarization experiments: Enhanced sensitivity in liquid-state NMR spectroscopy. *Journal of the American Chemical Society*, 128(29):9428–9432, 2006. 115, 137
- [57] C.G. Joo, A. Casey, C.J. Turner, and R.G. Griffin. In Situ Temperature-Jump Dynamic Nuclear Polarization: Enhanced Sensitivity in Two Dimensional C-13-C-13 Correlation Spectroscopy in Solution. *Journal of the American Chemical Society*, 131(1):12–+, 2009. 115
- [58] J. Leggett, R. Hunter, J. Granweh, A.J. Perez Linde, A. van der Drift, J. McMaster, R.W. Bowtell, A. Horsewill, G. Smith, and W. K. A novel

- integrated dissolution DNP spectrometer for liquid state NMR spectroscopy. *Conference Proceedings, EUROMAR 2008, St. Petersburg, Russia*, page 153, 2008. 115
- [59] M.J. Prandolini, V.P. Denysenkov, M. Gafurov, B. Endeward, and T.F. Prisner. High-Field Dynamic Nuclear Polarization in Aqueous Solutions. *Journal of the American Chemical Society*, 131(17):6090–6092, 2009. 115, 137
- [60] K.H. Hausser and D. Stehlik. Dynamic Nuclear Polarization in Liquids. *Advances in Magnetic Resonance*, 3:79–139, 1968. 116, 117, 118, 137
- [61] P. Höfer, G. Parigi, C. Luchinat, P. Carl, G. Guthausen, Reese M., T. Carlomagno, C. Griesinger, and M. Bennati. Field Dependent Dynamic Nuclear Polarization with Radicals in Aqueous Solution. *Journal of the American Chemical Society*, 130:3254–3255, 2008. 119, 121, 123, 125, 126, 135, 137, 138, 197
- [62] P. Höfer, P. Carl, G. Guthausen, T. Prisner, M. Reese, T. Carlomagno, C. Griesinger, and M. Bennati. Studies of Dynamic Nuclear Polarization with Nitroxides in Aqueous Solution. *Applied Magnetic Resonance*, 34:393–398, 2008. 121, 125, 126, 128, 135, 137, 138, 197
- [63] M. Reese, D. Lennartz, T. Marquardsen, P. Höfer, A. Tavernier, P. Carl, T. Schippmann, M. Bennati, T. Carlomagno, F. Engelke, and C. Griesinger. Construction of a Liquid-State NMR DNP Shuttle Spectrometer: First Experimental Results and Evaluation of Optimal Performance Characteristics. *Applied Magnetic Resonance*, 34:301–311, 2008. 121, 126, 127, 128, 129, 138, 197, 204, 205, 206
- [64] M. Reese, M.-T. Türke, I. Tkach, G. Parigi, C. Luchinat, T. Marquardsen, A. Tavernier, P. Höfer, F. Engelke, C. Griesinger, and M. Bennati.  $^1\text{H}$  and  $^{13}\text{C}$  Dynamic Nuclear Polarization in Aqueous Solution with a Two-Field (0.34 T/14 T) Shuttle DNP Spectrometer. *Journal of the American Chemical Society*, 131:15086–15087, 2009. 122, 130, 138, 200, 204, 205, 206, 207, 208, 209
- [65] A. Krahn, P. Lottmann, T. Marquardsen, A. Tavernier, Türke M.-T., M. Reese, A. Leonov, M. Bennati, P. Höfer, F. Engelke, and C. Griesinger. Shuttle DNP Spectrometer with a Two-Center Magnet. *Physical Chemistry Chemical Physics*, 2010, submitted. 122, 131, 132, 138, 199, 201, 207, 209, 210, 211
- [66] M.-T. Türke, I. Tkach, M. Reese, P. Höfer, and M. Bennati. Optimization of Dynamic Nuclear Polarization Experiments in Aqueous Solution

- at 15 MHz/9.7 GHz: A Comparative Study with DNP at 140 MHz/94 GHz. *Physical Chemistry Chemical Physics*, 2010, submitted. 122, 133, 134, 138, 201
- [67] B.D. Armstrong and S. Han. A new model for Overhauser enhanced nuclear magnetic resonance using nitroxide radicals. *Journal of Chemical Physics*, 127(10):104508, 2007. 123, 124
- [68] R.A. Wind and J.H. Ardenkjaer-Larsen. H-1 DNP at 1.4 T of water doped with a triarylmethyl-based radical . *Journal of Magnetic Resonance*, 141(2):347–354, 1999. 123
- [69] D. Marsh and L. Horvath. Influence of Heisenberg spin exchange on conventional and phase-quadrature EPR lineshapes and intensities under saturation. *Journal of Magnetic Resonance*, 97(1):13–26, 1991. 125
- [70] S. Stevenson, T. Glass, and H. C. Dorn. tsc13C Dynamic Nuclear Polarization: an Alternative Detector for Recycled-Flow NMR Experiments. *Analytical Chemistry*, 70:2623–2628, 1998. 130
- [71] C. Luchinat and G. Parigi. Collective Relaxation of Protein Protons at Very Low Magnetic Field: A New Window on Protein Dynamics and Aggregation. *Journal of the American Chemical Society*, 129(5):1055–1064, 2007. 130, 198
- [72] B.D. Armstrong and S. Han. Overhauser Dynamic Nuclear Polarization To Study Local Water Dynamics. *Journal of the American Chemical Society*, 131(31):11270–11270, 2009. 132
- [73] C. P. Poole. *Electron Spin Resonance*. John Wiley & Sons, New York, 1983. 134
- [74] A. G. Redfield. Shuttling device for high-resolution measurements of relaxation and related phenomena in solution at low field, using a shared commercial 500 MHz NMR instrument. *Magnetic Resonance in Chemistry*, 41:753–768, 2003. 139

# Acknowledgments

I want to express my deep gratitude to Prof. Christian Griesinger for giving me the opportunity to work on such interesting and challenging projects. I am especially thankful for being allowed to work in great independence and still getting his full trust and support in any situation. He was always willing to discuss scientific and organizational aspects of my work without any delay. Considering his tight schedule and his manifold obligations this attitude is thoroughly remarkable.

I am very thankful to Prof. Christoph Schmidt from the University of Göttingen for accepting the supervision of my external dissertation work and for very helpful support.

I want to acknowledge the great supervision of Dr. Teresa Carlomagno of my initial work on liquid-state DNP. She has always had great ideas how to overcome scientific and technical difficulties and gave great impulses for further studies.

I deeply appreciate the trustful collaboration with Dr. Marina Bennati. I found our collaboration on the DNP project especially fruitful. I appreciate the great relationship to her group members Dr. Igor Tchak and Maria-Teresa Türke. In the tight collaboration on DNP, concerning NMR and EPR spectroscopy, they have taught me a lot about the theoretical, experimental and instrumental aspects of EPR, for which I want to express my deep gratitude. I want to express my special thanks to Maria-Teresa Türke for the pleasant working atmosphere during our stays at Bruker in Karlsruhe.

The hardware developments on the shuttle DNP project would not have been possible without the support by Dr. Peter Höfer, Dr. Frank Engelke, Dr. Alexander Krahn, Thorsten Marquardsen and Andreas Tavernier from Bruker. I am very grateful for the pleasant and productive working atmosphere with them. Any kind of technical problem could be discussed and resolved. Their genuine ideas and solutions were crucial for the DNP project to succeed. I want to express my special thanks to Thorsten Marquardsen and Andreas Tavernier for all the time they spent after working hours to allow further experiments to be performed. The intensive and detailed dis-

cussions with Andreas Tavernier about technical and experimental aspects of EPR were just great and I have learned a lot from him. I am deeply grateful for that.

I want to thank Rainer Schürkötter, Christian Klaba, Jens Schnelle, Bernd Henkner, Michael Zigan, Mario Lengauer from the precision-engineering workshop and Herbert Nolte from the glass techniques workshop of the MPIBPC for their great technical work concerning the realization of the 1.5 m shuttle relaxometer, the shock transfer system and the shuttle sample containers. All of them and in particular Rainer Schürkötter contributed to finding the best possible technical realization for any given problem.

I want to thank Dr. Wolfgang Harneit for providing the N@C60 Buckminster Fullerenes for initial experiments to determine the  $B_1$  field strength in a critically coupled EPR resonator.

I want to thank Prof. Christian Griesinger, Dr. Teresa Carlomagno, Dr. Ulrich Wendt, Dr. Stefan Bartoschek, Dr. Peter Monecke, Dr. Jennifer Tuma and Julien Orts for the pleasant and effective working atmosphere in INPHARMA relating projects.

# Curriculum Vitae

Marcel Reese

[marcel.reese@gmx.net](mailto:marcel.reese@gmx.net)

---

## Personal information

Date of birth: Oct. 11, 1977

Place of birth: Bielefeld, Germany

Nationality: German

---

## Education

- Since Jul. 2005      Dissertational work, University of Göttingen and Max Planck Institute for Biophysical Chemistry, Germany, in supervision by Prof. Christoph Schmidt and Prof. Christian Griesinger.  
Topics: a) General sensitivity improvement of nuclear magnetic resonance spectroscopy with the emphasis on shuttle DNP. b) Improvement of the INPHARMA method for determining the atomic structure of drug-protein interactions by experimental and computational means.
- Oct. 2001 – Jul. 2005      Diploma studies in physics, University of Göttingen and Max Planck Institute for Biophysical Chemistry, Germany, in supervision by Prof. Theo Geisel and Prof. Christian Griesinger. Acquisition of Diploma in Physics, comparable to a Master of Sciences in Physics.  
Topic: Development of the INPHARMA method for determining the atomic structure of drug-protein interactions by experimental and computational means.
- Oct. 1999 – Sep. 2001      Prediploma studies in physics, University of Paderborn, Germany. Acquisition of Prediploma in Physics.
- Oct. 1998 – Sep. 1999      Studies in biology and natural sciences, University of Bielefeld, Germany.
- Aug. 1988 – Jun. 1997      Secondary school, Hans-Ehrenberg-Schule, Bielefeld, Germany. Acquisition of “Abitur” (general qualification for university entrance), followed by civil service until autumn 1998

---

## Advanced courses

- Nov. 24 – 28, 2008      Triplet state DNP course, Paul Scherrer Institute, Villigen, Switzerland, “Dynamic Nuclear Polarisation using triplet states, a course by W.Th. Wenckebach”
- Jul. 4–5, 2008      NMRCM Summer School, St. Petersburg, Russia
- Summer 2008      Bruker Hardware Course, Bruker Biospin GmbH, MPIbPC, Göttingen, Germany
- Jul. 6, 2007      DNP COST NMR SCHOOL, Tarragona, Spain, “Electron- Nuclear Spin Interactions”
- Mar. 6, 2007      COMSOL Multiphysics Workshop, COMSOL Multiphysics GmbH, Göttingen, Germany
- Dec. 4, 2004      Course on Scientific Communication, MPIbPC, Göttingen, Germany
- Nov. 10–12, 2003      Programming of parallel computers, GWDG, Göttingen, Germany
- Aug. 31 – Sep. 5, 2003      Advanced Computing in NMR Spectroscopy, Magnetic Resonance Center (CERM) of the University of Florence, Florence, Italy

---

*Internships*

Apr. – Jul. 2002	Theoretical and Computational Biophysics group, Dr. Helmut Grubmüller, Max Planck Institute for Biophysical Chemistry, Germany, “Analysis of molecular dynamics simulations of lipid bilayers in electric fields”
Summer 2000	Nine weeks internship, Dr. Werner Platzer, Fraunhofer Institute for Solar Energy Systems, “Evaluation and simulation of an active façade module in interconnection with a building”
Dec. 28 – Jan. 6, 98/99	Project participant, “Norrskén 98.99..2000”, Swedish Institute of Space Physics (IRF), Kiruna, Sweden, “Observation of aurora borealis and recording its audio-frequency radio signals”

---

*Research visits*

Jan. 17 – Feb. 1, 2004	Institute of Inorganic Chemistry, Slovak Academy of Sciences, Bratislava, Slovakia, Prof. Malkin and Dr. Olga Malkina, “Simulation of NMR parameters from atomic coordinates by DFT and first principle methods using deMon and GAUSSIAN”
Aug. 18 – 20, 2009	Brandeis University, Boston, MA, Prof. Redfield, “Understanding the electro-pneumatic shuttle NMR relaxometer”

---

*Technical experience*

NMR, EPR and DNP hardware development, simple RF and MW circuitry, simple pulse EPR, CW EPR, first photo-CIDNP experience, field cycling relaxometry, physical modelling, 1D and 2D NMR, NMR structure calculations, numerical simulations, algorithm design, DFT calculations, computational molecular dynamics studies, force field development, restraint development for XPLOR-NIH

---

*Computer skills*

<i>Programming skills:</i>	C++, FORTRAN parallel programming using MPI, openMP and multithreading compilers intense use of massive parallel computer facilities: GWDG (Göttingen, Germany), RZG (Garching, Germany), NIC (Jülich, Germany) XPLOR-NIH restraint implementation
<i>Operating systems:</i>	MS Windows, Linux, Mac OS X, virtualization using VMware
<i>Software:</i>	Matlab, Maple 12, Autocad/Inventor 2009, Comsol Multiphysics, Pymol, Molmol, VMD, Gromacs, CNS, XPLOR-NIH, HADDOCK, XWin-NMR, Topspin, Felix, Gaussian, deMon, MS Office, Corel Draw X3, LaTeX

---

*Teaching and supervision*

Mar. – Sep. 2005	Supervision and scientific introduction for chemistry diploma student
2007	Lectures on shell scripting and shell programming
Sep. 2007	Training of sanofi-aventis employer for using my parallel INPHARMA program
Summer semester 2003	Supervising student of the introductory course to experimental physics for medical students

---

*Language skills*

German: mother tongue  
English: fluent  
Spanish, French, Chinese: beginner level

---

*Oral presentations*

*Invited talks at conferences*

- Oct. 27 – 29, 2008 Joint NMR-Life and Extend-NMR Meeting: “New approaches to biomolecular structure determination”, Berlin, Germany, “Sample Shuttling Bio-DNP in Liquids”.
- Jun. 12 – 17, 2008 BIT WORLD CANCER CONGRESS 2008, Shanghai, China, “The INPHARMA Method: Drug-Receptor Interaction and Structure Studied by Drug-Only-Observed NMR”
- Oct. 22, 2007 International Symposium on Drug Discovery and Design by NMR, Yokohama, Japan, talk in substitution for Prof. Griesinger, “Drug protein interactions studied by NMR”

*Invited talks at universities or research institutes*

- Aug. 27, 2009 Third Institute of Physics, University of Stuttgart, Stuttgart, Germany, presented to Prof. Wrachtrup’s institute, “INPHARMA: Drug Receptor Interaction by Drug-Only Observed NMR - Sensitivity Enhancement of Liquid State NMR by Shuttle DNP”
- Apr. 7, 2009 IBM Almaden Research Center, San Jose, California, USA, “INPHARMA: Drug Receptor Interaction by Drug-Only Observed NMR - Sensitivity Enhancement of Liquid State NMR by Shuttle DNP and beyond”
- Apr. 6, 2009 University of California, Santa Barbara, California, USA, presented to Prof. Han’s group, “Sensitivity Enhancement of Liquid State NMR by Shuttle DNP”

*Further talks*

- Sep. 2 – 4, 2009 The Second International Symposium “Dynamic Nuclear Polarization: Theory - Hardware-Applications- Radicals”, Königstein, Germany, “A Liquid-State Shuttle DNP Spectrometer for 600 MHz NMR: Construction and Results for 1H and 13C Signal Enhancement”
- Nov. 17, 2008 Triplet state DNP course, Paul Scherrer Institute, Villigen, Switzerland, “Sample Shuttling Bio-DNP in Liquids and Beyond”
- Sep. 26 – 27, 2008 Third annual Bio-DNP progress meeting, Florence, Italy, “Sample Shuttling Bio-DNP in Liquids”

---

*Poster presentations*

- Jul. 5–9, 2009 EUROMAR 2009, Göteborg, Sweden, “A Liquid-State Shuttle DNP Spectrometer for 600 MHz High-Resolution NMR: Construction and Results for 1H and 13C Sensitivity Enhancement”
- Mar. 29 – Apr. 3, 2009 ENC 2009, Asilomar, CA, USA, “A Field Cycling Liquid-State DNP Spectrometer for 600 MHz High-Resolution NMR: Construction and Results for 1H and 13C Sensitivity Enhancement”
- Jul. 6–11, 2008 EUROMAR 2008, St. Petersburg, Russia, “The INPHARMA Method: Drug-Receptor Interaction and Structure Studied by Drug-only-Observed NMR”
- Jun. 12–17, 2008 BIT World Cancer Congress - 2008, Shanghai, China, “The INPHARMA Method: Drug-Receptor Interaction and Structure Studied by Drug-Only-Observed NMR”
- Mar. 9–14, 2008 ENC 2008, Asilomar, CA, USA, “Development of a Field Cycling DNP Spectrometer for 600 MHz NMR, First Experimental Results and Evaluation of Optimal Performance Characteristics”
- Aug. 29–31, 2007 Dynamic Nuclear Polarisation Symposium, Nottingham, Great Britain, “High Resolution Field Cycling DNP and Relaxometry by Pneumatic Sample Shuttling”
- Jul. 1–6, 2007 Euromar Magnetic Resonance Meeting 2007, Tarragona, Spain, “High Resolution Field Cycling NMR and Relaxometry by Pneumatic Sample Shuttling”, Version 2
- Jun. 3–8, 2007 Xi Chianti Meeting, “High Resolution Field Cycling NMR and Relaxometry by Pneumatic Sample Shuttling”, Version 1
- May 31 – Jun. 2, 2007 5th Fast Field Cycling NMR Relaxometry Conference, Turin, Italien, “High Resolution Field Cycling NMR and Relaxometry by Pneumatic Sample Shuttling”, Version 1
- Aug. 20–25, 2006 XXIIInd ICMRBS, Göttingen, Deutschland, “Theory, algorithms and user-friendly and fast parallel program for INPHARMA based lead discovery and optimization”
- Apr. 23–28, 2006 ENC 2006, Pacific Grove CA, USA, “Theory, algorithms and user-friendly and fast parallel program for INPHARMA based lead discovery and optimization”

---

Publications

- M.-T. Türke, I. Tkach, M. Reese, P. Höfer, and M. Bennati. Optimization of Dynamic Nuclear Polarization Experiments in Aqueous Solution at 15 MHz/9.7 GHz: A Comparative Study with DNP at 140 MHz/94 GHz., *Physical Chemistry Chemical Physics*, 2010, submitted.
- A. Krahn, P. Lottmann, T. Marquardsen, A. Tavernier, Türke M.- T., M. Reese, A. Leonov, M. Bennati, P. Höfer, F. Engelke, and C. Griesinger, Shuttle DNP Spectrometer with a Two-Center Magnet, *Physical Chemistry Chemical Physics*, 2010, submitted.
- <sup>1</sup>H and <sup>13</sup>C Dynamic Nuclear Polarization in Aqueous Solution with a Two-Field (0.35 T/14 T) Shuttle DNP Spectrometer, Marcel Reese, Maria-Teresa Türke, Igor Tkach, Giacomo Parigi, Claudio Luchinat, Thorsten Marquardsen, Andreas Tavernier, Peter Höfer, Frank Engelke, Christian Griesinger and Marina Bennati, *Journal of the American Chemical Society* 131(42), pp 15086-15087 (2009)
- Field Dependent Dynamic Nuclear Polarization (DNP) with Radicals in Aqueous Solution, Peter Höfer, Giacomo Parigi, Claudio Luchinat, Patrick Carl, Gisela Guthausen, Marcel Reese, Teresa Carlomagno, Christian Griesinger, Marina Bennati, *Journal of the American Chemical Society* 130(11), pp 3254–3255 (2008)
- Crystallography-Independent Determination of Ligand Binding Modes Julien Orts, Jennifer Tuma, Marcel Reese, Kaspar Grimm, Peter Monecke, Stefan Bartoschek, Alexander Schiffer, K. Ulrich Wendt, Christian Griesinger, Teresa Carlomagno, *Angewandte Chemie Int Ed Engl.* 47(40), pp 7736 – 7740 (2008)
- Studies of Dynamic Nuclear Polarization with Nitroxides in Aqueous Solution, Peter Höfer, Patrick Carl, Gisela Guthausen, Marcel Reese, Teresa Carlomagno, Christian Griesinger, Marina Bennati, *Applied Magnetic Resonance* 34(3-4), pp 393-398 (2008)
- Construction of a liquid state NMR DNP shuttle spectrometer, first experimental results and evaluation of optimal performance characteristics, Marcel Reese, Dirk Lennartz, Thorsten Marquardsen, Peter Höfer, Andreas Tavernier, Patrick Carl, Toralf Schippmann, Marina Bennati, Teresa Carlomagno, Frank Engelke, Christian Griesinger, *Applied Magnetic Resonance* 34(3-4), pp 301-311 (2008)
- Structural Basis of the Activity of the Microtubule-Stabilizing Agent Epopthilone A Studied by NMR Spectroscopy in Solution, Marcel Reese, Victor M Sánchez-Pedregal, Karel Kubicek, Jens Meiler, Marcel J J Blommers, Christian Griesinger, Teresa Carlomagno, *Angewandte Chemie Int Ed Engl.* 46(11), pp 1864-1868 (2007)
- The INPHARMA method: protein-mediated interligand NOEs for pharmacophore mapping, Victor M Sánchez-Pedregal, Marcel Reese, Jens Meiler, Marcel J J Blommers, Christian Griesinger, Teresa Carlomagno, *Angewandte Chemie Int Ed Engl.* 44(27), pp 4172-4175 (2005)

The following link provides easy online access to my publications and to the all-over citation statistics:

<http://www.researcherid.com/rid/C-9434-2009>.

Addendum to the Dissertation:

Sensitivity Enhancement  
of  
Liquid-State NMR  
and  
Improvement of the INPHARMA  
Method

Marcel Reese

from Bielefeld

Bielefeld, 2010-06-09



# Preface

In my dissertation<sup>1</sup> I have described the theory, a method and a program to use NOESY-based spectra under the influence of chemical exchange to directly determine and refine structural molecular models. This description was using a symmetrization matrix to speed-up the calculation of the spectra as desired for the application to INPHARMA and trNOE spectra. Although this procedure is applicable to some ILOE (inter-ligand NOE) spectra this applicability is not of a general nature but depends on the nature of the required kinetic matrix.

Here I additionally describe a general theory and procedure applicable to all kinds of NOESY spectra involving chemical exchange. The derived gradient, following the example of my dissertation, can be used for structural refinement of macromolecular complexes via molecular dynamics simulated annealing and energy minimization.

Furthermore the theory for gradient-based refinement of structural models using experimental data of a multitude of samples of different compositions is presented. In the same context the simplified model of dynamic averaging will be discussed.

Marcel Reese  
Bielefeld, 2010-06-01

minor editorial changes:  
Cambridge, USA, 2011-02-15



# Contents

<b>Preface</b>	<b>iii</b>
<b>Table of Contents</b>	<b>iv</b>
<b>1 Theory, General Case</b>	<b>1</b>
1.1 Simulation of NOESY Peak Volumes, General Case . . . . .	2
1.2 Derivation: NOESY Penalty Energy and Gradient . . . . .	3
1.2.1 Lemma 3 . . . . .	6
1.3 Implementation Procedure, General Case . . . . .	7
<b>2 Multiplexing</b>	<b>9</b>
2.1 Theory . . . . .	11
2.2 Implementation . . . . .	13
<b>3 Dynamical Averaging</b>	<b>15</b>
3.1 Theory . . . . .	15
3.2 Implementation, Dynamical Averaging . . . . .	18
<b>Bibliography</b>	<b>21</b>



# Chapter 1

## Theory, General Case

My disserational work was motivated by the INPHARMA method . This method's theoretical description was the basis for the derivation of the theory for the described gradient-based optimization methods. Different techniques which are based on the NOESY experiment will require different chemical exchange matrices for their accurate simulation. Not for all of them the symmetrization matrix is generally available. This is including ILOE (inter ligand NOE) spectra.<sup>2</sup> For other spectra these matrices can be easily derived (trNOE, INPHARMA).

In the INPHARMA case it is easy to symmetrize the effective relaxation matrix  $\mathbf{R}^*$ , using the diagonal symmetrization matrix  $\mathbf{C}$  with  $\mathbf{C}_{ii} = 1/\sqrt{\mathbf{K}_{ii}}$ , to gain  $\mathbf{R}^{*s} = \mathbf{C}^{-1}\mathbf{R}^*\mathbf{C}$ . The left hand symmetrization matrix  $\mathbf{C}^{-1}$  and the right-hand symmetrization matrix  $\mathbf{C}$  are their respective inverse which facilitates their further use in the derivation of the gradients. Furthermore the eigenvectors of the symmetrized relaxation matrix  $\boldsymbol{\chi}^*$  are orthonormal and symmetric with  $(\boldsymbol{\chi}^*)^{-1} = (\boldsymbol{\chi}^*)^{tr}$ . This is saving computation time; the inverse is easy to obtain. For eigenvectors of non-symmetric matrices the situation is different and the inverse has to be explicitly determined.

In general a symmetrization procedure is conceivable where the symmetrization matrices are not diagonal and the left-hand and the right-hand symmetrization matrices are not their respective inverse. This case describes symmetrizability of the second kind.<sup>3</sup> "A matrix  $A$  is symmetrizable of the second kind if and only if there exist two nonzero matrices  $L$  and  $R$  such that  $\tilde{A} = L^T A R$  is a symmetric matrix." If the matrices  $L$  and  $R$  can be obtained the matrix  $\tilde{A}$  is available. Now it is straight forward to determine the eigenvectors and eigenvalues of the symmetric matrix  $\tilde{A}$  and from them to determine the eigenvectors and eigenvalues of  $A$ . This procedure is not generally faster than diagonalizing  $A$  directly.

Calculating the eigenvectors and eigenvalues of the not symmetric  $\mathbf{R}^* =$

$\mathbf{R} + \mathbf{K}$  is possible also in the cases where symmetrization can not be (easily) applied. Then the diagonalization procedure will require more time and additionally the determination of the inverse of the eigenvector matrix is required.

Following the same sequence as in the derivation during my dissertation I will derive the gradient of a NOESY-based penalty energy for the general case where symmetrization is not considered but the left-hand and right-hand eigenvectors as well as the eigenvalues of the effective relaxation matrix  $\mathbf{R}^*$  are available and were determined by an arbitrary method.

This theory covers gradient-based refinement of all possible systems where NOESY spectra can be measured on. This will include systems with a multitude of ligands, a multitude of receptors, several binding-pockets, etc.

## 1.1 Simulation of NOESY Peak Volumes, General Case

The physical system and the mathematical description don't change in this general case where symmetrization of the effective relaxation matrix is not performed. Formulas [1, (2.1)-(2.9)] keep unchanged.

In the general case the solution for the time evolution of the grouped peak volumes<sup>4,5</sup> is:

$$\begin{aligned} \mathbf{A}_m^g(\tau_m) &= \mathbf{G} \mathbf{A}_m(\tau_m) \mathbf{G}^{\text{tr}} \\ &= \mathbf{G} \exp(-\mathbf{R}^* \tau_m) \mathbf{A}^0 \mathbf{G}^{\text{tr}} \\ &= \mathbf{G} \boldsymbol{\chi}^* \exp(-\boldsymbol{\Lambda}^* \tau_m) (\boldsymbol{\chi}^*)^{-1} \mathbf{A}^0 \mathbf{G}^{\text{tr}}. \end{aligned} \quad (1.1)$$

The following symbols are used:

- $\mathbf{A}_m^g$  Grouped peak volume matrix for the mixing time with index  $m$
- $\mathbf{A}_m$  Individual-atom peak volume matrix for the mixing time with index  $m$
- $\mathbf{A}^0$  Individual-atom initial peak volume matrix
- $\mathbf{G}$  Grouping matrix
- $\boldsymbol{\Lambda}$  Diagonal matrix of the eigenvalues of the effective, generally not symmetrized relaxation matrix  $\mathbf{R}^*$
- $\boldsymbol{\chi}^*$  Orthonormal matrix of eigenvectors of the effective, relaxation matrix  $\mathbf{R}^*$
- $\tau_m$  Mixing time of the NOESY experiment with index  $m$

For clarity I define  $\mathbf{L} \equiv \boldsymbol{\chi}^*$  and  $\boldsymbol{\Lambda} \equiv \boldsymbol{\Lambda}^*$ . Now the equation becomes

$$\begin{aligned} \mathbf{A}_m^g &= \mathbf{G} \mathbf{A}_m \mathbf{G}^{\text{tr}} \\ &= \mathbf{G} \mathbf{L} \exp(-\boldsymbol{\Lambda} \tau_m) \mathbf{L}^{-1} \mathbf{A}^0 \mathbf{G}^{\text{tr}}. \end{aligned} \quad (1.2)$$

## 1.2 Derivation: NOESY Penalty Energy and Gradient

The definition of the penalty energy and the derivation of its gradient in the general case do not differ to the symmetrizable case for formulas [1, (2.12-21.7)].

Formula [1, (2.18)] needs to be modified. Instead of  $\mathbf{R}^{*s}$  (which is symmetric) the more general  $\mathbf{R}^* = \mathbf{R} + \mathbf{K}$  is used.

The derivative of  $E_{\text{INPH}}$  with respect to the Cartesian coordinates of the protons can be calculated as follows, according to the chain rule of differentiation:

$$\frac{\partial E}{\partial \mathbf{r}_i} = \sum_{\alpha\beta klmnj} \frac{\partial E}{\partial \mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \frac{\partial \mathbf{A}_{kl}}{\partial \mathbf{R}_{mn}^*} \frac{\partial \mathbf{R}_{mn}^*}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \mathbf{r}_i}. \quad (1.3)$$

The formulas [1, (2.19-2.21)] stay unchanged.

To calculate  $\partial E / \partial \mathbf{R}_{mn}^*$  I define, for readability,  $\mathbf{R} \equiv \mathbf{R}^* = \mathbf{L}\mathbf{A}\mathbf{L}^{-1}$ .

$$\begin{aligned} \frac{\partial E}{\mathbf{R}_{mn}} &= \sum_{kl} \mathbf{F}_{kl} \frac{\partial \mathbf{A}_{kl}}{\partial \mathbf{R}_{mn}} \\ &= \sum_{kl} \mathbf{F}_{kl} \frac{\partial (\exp(-\tau \mathbf{R}) \mathbf{A}^0)_{kl}}{\partial \mathbf{R}_{mn}} \\ &= \sum_{kl} \mathbf{F}_{kl} \left[ \frac{\partial}{\partial \mathbf{R}_{mn}} \left( 1 - \tau \mathbf{R} + \frac{1}{2!} \tau^2 \mathbf{R}^2 - \frac{1}{3!} \tau^3 \mathbf{R}^3 + \dots \right)_{kl} \right] \mathbf{A}_l^0 \end{aligned} \quad (1.4)$$

The derivation of the lemma [1, (9.5)] (and formula [1, (2.23)]) is not affected as it does neither require nor use a symmetry of  $\mathbf{R}$ .

Now I can gain the result of the derivative of a single summand of the power series (1.4). The factor  $(-\tau)^N / N!$  is skipped here:

$$\sum_{kl} \mathbf{F}_{kl} \frac{\partial (\mathbf{R}^N)_{kl}}{\partial \mathbf{R}_{mn}} \mathbf{A}_l^0 = \left( \sum_{f=1}^N (\mathbf{R}^{\text{tr}})^{f-1} \mathbf{F}^A (\mathbf{R}^{\text{tr}})^{N-f} \right)_{mn} \quad (1.5)$$

Here the definition  $\mathbf{F}^A \equiv \mathbf{F}\mathbf{A}^0$  is used. The details for this derivation can also be found in (1.15, p. 6).

Now, including the previously skipped factor  $(-\tau)^N / N!$ , the derivative of the energy  $E$  with respect to the relaxation matrix  $\mathbf{R} = \mathbf{R}^*$  gets

$$\frac{\partial E}{\partial \mathbf{R}} = \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} (\mathbf{R}^{\text{tr}})^{f-1} \mathbf{F}^A (\mathbf{R}^{\text{tr}})^{N-f}. \quad (1.6)$$

This can be further simplified using  $\mathbf{R} = \mathbf{R}^* = \mathbf{L}\mathbf{\Lambda}\mathbf{L}^{-1}$  with  $\mathbf{L}^{-1}\mathbf{L} = \mathbf{I}$  and so  $\mathbf{R}^N = \mathbf{L}\mathbf{\Lambda}^N\mathbf{L}^{-1}$ . In the following the notation  $\mathbf{R}^{\text{tr},N} \equiv (\mathbf{R}^{\text{tr}})^N$  and  $\mathbf{R}^{N,\text{tr}} \equiv (\mathbf{R}^N)^{\text{tr}}$  is used. In a first step I gain:

$$\frac{\partial E}{\partial \mathbf{R}} = \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} (\mathbf{L}\mathbf{\Lambda}\mathbf{L}^{-1})^{\text{tr},f-1} \mathbf{F}^{\text{A}} (\mathbf{L}\mathbf{\Lambda}\mathbf{L}^{-1})^{\text{tr},N-f} \quad (1.7)$$

I will need the following identity (with  $\mathbf{\Lambda}_k \equiv \mathbf{\Lambda}_{kk}$ )

$$\begin{aligned} (\mathbf{L}\mathbf{\Lambda}\mathbf{L}^{-1})_{ij}^{\text{tr}} &= (\mathbf{L}\mathbf{\Lambda}\mathbf{L}^{-1})_{ji} \\ &= \sum_k \mathbf{L}_{jk} \mathbf{\Lambda}_k \mathbf{L}_{ki}^{-1} \\ &= \sum_k \mathbf{L}_{ik}^{-1,\text{tr}} \mathbf{\Lambda}_k \mathbf{L}_{kj}^{\text{tr}} \\ &= (\mathbf{L}^{-1,\text{tr}} \mathbf{\Lambda} \mathbf{L}^{\text{tr}})_{ij}. \end{aligned} \quad (1.8)$$

Furthermore the following identity is needed

$$\begin{aligned} (\mathbf{L}^{\text{tr}} \mathbf{L}^{-1,\text{tr}})_{ij} &= \sum_k \mathbf{L}_{ik}^{\text{tr}} \mathbf{L}_{kj}^{-1,\text{tr}} \\ &= \sum_k \mathbf{L}_{jk}^{-1} \mathbf{L}_{ki} \\ &= \mathbf{I}_{ij}. \end{aligned} \quad (1.9)$$

Now formula (1.7) can be further simplified:

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{R}} &= \mathbf{L}^{-1,\text{tr}} \left( \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} \mathbf{\Lambda}^{f-1} \cdot \mathbf{L}^{\text{tr}} \mathbf{F}^{\text{A}} \mathbf{L}^{-1,\text{tr}} \cdot \mathbf{\Lambda}^{N-f} \right) \mathbf{L}^{\text{tr}} \\ &= \mathbf{L}^{-1,\text{tr}} \left( \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^N}{N!} \mathbf{\Lambda}^{f-1} \cdot \mathbf{L}^{\text{tr}} \mathbf{F}^{\text{A}} \mathbf{L}^{-1,\text{tr}} \cdot \mathbf{\Lambda}^{N-f} \right) \mathbf{L}^{\text{tr}} \quad (1.10) \\ &= -\tau \mathbf{L}^{-1,\text{tr}} \left( \sum_{N=1}^{\infty} \sum_{f=1}^N \frac{(-\tau)^{N-1}}{N!} \mathbf{\Lambda}^{f-1} \cdot \mathbf{M}^{\text{A}} \cdot \mathbf{\Lambda}^{N-f} \right) \mathbf{L}^{\text{tr}}. \end{aligned}$$

In the last step of this derivation the definition

$$\begin{aligned} \mathbf{M}^{\text{A}} &\equiv \mathbf{L}^{\text{tr}} \mathbf{F}^{\text{A}} \mathbf{L}^{-1,\text{tr}} \\ &= \mathbf{L}^{\text{tr}} \mathbf{F}^{\text{A}} \mathbf{L}^{-1,\text{tr}} \end{aligned} \quad (1.11)$$

is introduced.

Using the Yip and Case series  $\mathbf{J}$  [6, p. 643, formula 10], I get

$$\mathbf{Q} \equiv \frac{\partial E}{\partial \mathbf{R}} = -\tau \mathbf{L}^{-1, \text{tr}} \mathbf{J}(\mathbf{M}^{\mathbf{A}}, -\tau \mathbf{\Lambda}) \mathbf{L}^{\text{tr}}, \quad (1.12)$$

where  $\mathbf{J}$  is defined as

$$\mathbf{J}(\mathbf{A}, \mathbf{B}) = \begin{cases} \frac{\mathbf{A}_{ij}}{\mathbf{B}_{ii} - \mathbf{B}_{jj}} [\exp(\mathbf{B}_{ii}) - \exp(\mathbf{B}_{jj})] & \text{if } \mathbf{B}_{ii} \neq \mathbf{B}_{jj} \\ \mathbf{A}_{ij} \exp(\mathbf{B}_{ii}) & \text{if } \mathbf{B}_{ii} = \mathbf{B}_{jj}. \end{cases} \quad (1.13)$$

The next step is to calculate the partial derivative with respect to the distances  $r_{ij}$ :

$$\begin{aligned} \frac{\partial E}{\partial r_{ij}} &= \sum_{mn} \frac{\partial E}{\partial \mathbf{R}_{mn}} \frac{\partial \mathbf{R}_{mn}}{\partial r_{ij}} \\ &= \sum_{mn} \mathbf{Q} \frac{\partial \mathbf{R}_{mn}^*}{\partial r_{ij}}. \end{aligned} \quad (1.14)$$

The remaining formulas [1, (2.32-2.34)] stay unchanged.

### 1.2.1 Lemma 3

This derivation of lemma 3 (corresponding to lemma 2 [1, (9.6)], for the general case) uses the result of lemma 1 [1, (9.5)]. For the diagonal matrix  $\mathbf{A}^0$  the definition  $\mathbf{A}_l^0 \equiv \mathbf{A}_{ll}^0$  is used and furthermore the theorem  $\mathbf{R}^{\text{tr},N} = \mathbf{R}^{N,\text{tr}}$  is applied.

$$\begin{aligned}
\sum_{kl} \mathbf{F}_{kl} \frac{\partial (\mathbf{R}^N)_{kl}}{\partial \mathbf{R}_{mn}} \mathbf{A}_l^0 &= \sum_{kl} \mathbf{F}_{kl} \left\{ \delta_{km} (\mathbf{R}^{N-1})_{nl} \right. \\
&\quad + \sum_{f=1}^{N-2} (\mathbf{R}^f)_{km} (\mathbf{R}^{N-1-f})_{nl} \\
&\quad \left. + \delta_{nl} (\mathbf{R}^{N-1})_{km} \right\} \mathbf{A}_l^0 \\
&= \sum_l \mathbf{F}_{ml} \mathbf{A}_l^0 (\mathbf{R}^{\text{tr},N-1})_{ln} \\
&\quad + \sum_{kl} \sum_{f=1}^{N-2} (\mathbf{R}^{\text{tr},f})_{mk} \mathbf{F}_{kl} \mathbf{A}_l^0 (\mathbf{R}^{\text{tr},N-1-f})_{ln} \\
&\quad + \sum_k (\mathbf{R}^{\text{tr},N-1})_{mk} \mathbf{F}_{kn} \mathbf{A}_n^0 \tag{1.15} \\
&= (\mathbf{F} \mathbf{A}^0 \mathbf{R}^{\text{tr},N-1})_{mn} \\
&\quad + \left( \sum_{f=1}^{N-2} \mathbf{R}^{\text{tr},f} \mathbf{F} \mathbf{A}^0 \mathbf{R}^{\text{tr},N-1-f} \right)_{mn} \\
&\quad + (\mathbf{R}^{\text{tr},N-1} \mathbf{F} \mathbf{A}^0)_{mn} \\
&= \left( \sum_{f=1}^N \mathbf{R}^{\text{tr},f-1} \mathbf{F} \mathbf{A}^0 \mathbf{R}^{\text{tr},N-f} \right)_{mn} \\
&= \left( \sum_{f=1}^N \mathbf{R}^{\text{tr},f-1} \mathbf{F} \mathbf{A} \mathbf{R}^{\text{tr},N-f} \right)_{mn}
\end{aligned}$$

## 1.3 Implementation Procedure, General Case

The sequence of operations for calculating the gradient of the penalty energy, in the general case not requiring symmetrization, according to the results of the previous chapter 'Theory, General Case', is the following:

1. Setting up the effective relaxation matrix  $\mathbf{R} \equiv \mathbf{R}^*$  and initial peak volume matrix  $\mathbf{A}^0$ , chemical exchange matrix  $\mathbf{K}$ , and the grouping matrix  $\mathbf{G}$  [1, (2.1-2.9)].
2. Calculating the eigenvectors  $\mathbf{L} \equiv \boldsymbol{\chi}^*$ , the inverse  $\mathbf{L}^{-1}$  and eigenvalues  $\boldsymbol{\Lambda} \equiv \boldsymbol{\lambda}^*$ .
3. Calculating the matrix of grouped peak volumes (1.2) for a given mixing time  $\tau_m$

$$\mathbf{A}_m^g = \mathbf{G}\mathbf{L} \exp(-\boldsymbol{\Lambda}\tau_m)\mathbf{L}^{-1}\mathbf{A}^0\mathbf{G}^{\text{tr}}. \quad (1.16)$$

4. Calculating the  $\mathbf{F}$  matrix [1, (2.21)] (for the given  $\tau_m$ )

$$\mathbf{F}_{kl} = \sum_{\alpha\beta} k_r w_{\alpha\beta} \begin{cases} -f_{small} & \text{if } cal \leq exp - err, \\ +1 & \text{if } cal \geq exp + err, \\ 0 & \text{else.} \end{cases} \quad (1.17)$$

$$\cdot \begin{cases} 1 & k \in \text{Peakgroup}_\alpha \text{ and } l \in \text{Peakgroup}_\beta, \\ 0 & \text{else.} \end{cases}$$

5. Calculating the matrix  $\mathbf{M}^A$ (1.11) (for the given  $\tau_m$ )

$$\mathbf{M}^A = \mathbf{L}^{\text{tr}}\mathbf{F}\mathbf{A}^0\mathbf{L}^{-1, \text{tr}}. \quad (1.18)$$

6. Calculating the matrix  $\mathbf{J}(\mathbf{M}^A, -\tau\boldsymbol{\Lambda})$  (1.13) (for the given  $\tau_m$ )

$$\mathbf{J}(\mathbf{A}, \mathbf{B}) = \begin{cases} \frac{\mathbf{A}_{ij}}{\mathbf{B}_{ii} - \mathbf{B}_{jj}} [\exp(\mathbf{B}_{ii}) - \exp(\mathbf{B}_{jj})] & \text{if } \mathbf{B}_{ii} \neq \mathbf{B}_{jj} \\ \mathbf{A}_{ij} \exp(\mathbf{B}_{ii}) & \text{if } \mathbf{B}_{ii} = \mathbf{B}_{jj}. \end{cases} \quad (1.19)$$

7. Calculating the matrix  $\mathbf{Q}$  (1.12) (for the given  $\tau_m$ )

$$\mathbf{Q} = -\tau\mathbf{L}^{-1, \text{tr}}\mathbf{J}(\mathbf{M}^A, -\tau\boldsymbol{\Lambda})\mathbf{L}^{\text{tr}}. \quad (1.20)$$

8. Calculating further matrices  $\mathbf{Q}$  for the remaining mixing times, starting at (1.16). Summing them up to gain the final matrix  $\mathbf{Q}$ .

9. Calculating the matrix  $\mathbf{H}$  [1, (2.33)], considering potentially different relaxation matrices  $\mathbf{R}$  and so different  $S_{ij}^{\mathbf{R},s}$

$$\mathbf{H}_{ij} = -\frac{6}{r_{ij}} \left[ (\mathbf{Q}_{ii} + \mathbf{Q}_{jj}) S_{ij}^{\mathbf{R},s} + (\mathbf{Q}_{ij} + \mathbf{Q}_{ji}) \mathbf{R}^{\text{NOE}} \right] \quad (1.21)$$

$$\text{with } S_{ij}^{\mathbf{R},s} = \frac{1}{10} \gamma_i^2 \gamma_j^2 \left( \frac{\hbar \mu_0}{4\pi} \right)^2 r_{ij}^{-6} [J_{ij,0}^s + 3J_{ij,1}^s + 6J_{ij,2}^s].$$

10. Calculating the forces, like here for  $x$  [1, (2.34)].

$$\frac{\partial E}{\partial x_i} = \sum_j \mathbf{H}_{ij} \frac{x_i - x_j}{r_{ij}}. \quad (1.22)$$

This procedure is valid if several spectra have to be calculated only differing in mixing time. If any of the parameters influencing the relaxation matrices have changed - like the field, concentrations or kinetic constants - the effective relaxation matrix has to be set-up again and diagonalization is necessary. Then, the resulting  $\mathbf{H}$  matrices for different spectra should be summed before the calculation of the forces.

Numerical approximations, just as described in my thesis,<sup>1</sup> can be applied to speed up the diagonalization and the following matrix multiplications. This will involve computing only the eigenvectors to the the smallest eigenvalues and considering the reduced matrix sizes in the matrix multiplications.

## Chapter 2

# Multiplexing

INPHARMA is based on the protein-mediated magnetization transfer of two ligands binding competitively to the same binding pocket of a protein. During a NOESY experiment both ligands and the protein will be together in the same solution. From these spectra the molecular structure of the ligand-receptor complexes can be determined in the vicinity of the binding pocket.

It has been discussed to use several of these experiments, using different pairs of ligands, to gain further insight into the structure of the complexes. The idea of multiplexing is to have for example three ligands A,B,C binding to the same binding pocket of the protein T (Using a set of proteins is also conceivable). Now three samples can be prepared with the different compositions  $c$ :

1. A,B,T with the complexes AT and BT,  $c_1 = ABT$
2. A,C,T with the complexes AT and CT,  $c_2 = ACT$
3. B,C,T with the complexes BT and CT,  $c_3 = BCT$ .

$C$  is the set of all compositions  $c_\gamma$ . (A composition can also be understood as the vector of concentrations in a sample.)

From each (or some) of these samples INPHARMA (or generally NOESY) spectra will be recorded.

Under the assumption that the structures of the complexes only depend on the respective bound ligand, and not the free ligand in solution, several spectra encode the same structural information for the same complex. For example the spectra of the samples 1) and 2) will both depend on the very same conformation of the complex AT (and slightly on the conformation of the free ligand A).

By this procedure the amount of experimental data for structure determination can be extended.

Structural selection, similarly to the procedure first described in my diploma thesis,<sup>4</sup> is performed by generating structural models for the various complexes, simulating the pairwise spectra from them, gaining a score depending on the match of simulated and experimental pairwise data, and finally gaining a aggregated or total score involving the pairs of the other samples.

One example for such a total score or penalty energy is

$$E^+ = k^+ \left( \sum_{\gamma} E_{\gamma}^e \right)^{1/e}. \quad (2.1)$$

For the three given systems the individual, compositional energies  $E_{\gamma}$  will be calculated, an  $e$ -norm  $E^+$  of the (positive) compositional penalty energies  $E_{\gamma}$  will be calculated which is finally scaled with the weighting factor  $k^+$  to gain the total penalty energy.

In the present strategy for INPHARMA duplexing sets of structural models for all species  $s \in \{A, B, C, AT, BT, CT\}$  are available. Although the free ligand conformations are considered in the simulation of the spectra and contribute to the compositional energies  $E_{\gamma}$  their influence is generally negligible due to their low size. The conformation of the free ligands can be considered to be identical to the conformation of the bound ligands. Nevertheless the actual conformation of the free ligands is considered in the simulations.

Let's assume we have the complex models AT, BT, CT with  $I, J, K$  many elements, respectively. Then  $I \cdot J \cdot K$  many different combinations can be created, each giving a new total score  $E_{ijk}^+$ :

$$E_{ijk}^+ = \left( E_{\gamma=1,ij}^e + E_{2,ik}^e + E_{3,jk}^e \right)^{1/e}. \quad (2.2)$$

This procedure is currently applied to create total scores and by this select a set of complexes, like  $\{AT_i, BT_j, CT_k\}$ , which best fulfills all available data.

Until now it is not possible to use these scores or energies for directed structural refinement of a given combination of structural models.

## 2.1 Theory

Here I will explain an extension to the multiplexing approach. It will allow for structural refinement of a given combination of sample compositions. Each set of spectra will be run on the identical sample and will consist of spectra with different mixing times.

Besides being applicable to multiplexing the described approach is also useful when spectra of the same sample or composition are recorded under varying conditions which ideally should not alter the complex conformations. The varying parameters may include concentrations or magnetic field or else. Then any of these conditions would be assigned a different, general composition  $c_\gamma$ .

Without loss of generality this derivation is exemplified on the penalty energy 2.1:

$$E^+ = \left( \sum_{\gamma} E_{\gamma}^e \right)^{1/e}. \quad (2.3)$$

In the case of a single spectrum the penalty energy was a multiple of an R-factor [1, (2.12)]. Its derivative with respect to the atomic coordinates was derived using the chain rule of differentiation [1, 2.18], 2.4:

$$\frac{\partial E}{\partial \mathbf{r}_i} = \sum_{\alpha\beta klmnj} \frac{\partial E}{\partial \mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \frac{\partial \mathbf{A}_{kl}}{\partial \mathbf{R}_{mn}^*} \frac{\partial \mathbf{R}_{mn}^*}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \mathbf{r}_i}. \quad (2.4)$$

The previous score, the penalty energy  $E$ , will be substituted by the total energy  $E^+$  of a multiplexing set, which in turn depends on the compositional energies  $E_\gamma$  of the samples.  $E_\gamma$  is corresponding to the energy  $E$  as used in my dissertation in the case of only one available composition. (This energy is not restricted to INPHARMA and can also depend on several proteins and ligands and any kind of chemical exchange matrix. Any number of species in solution with different concentrations is possible. The species are allowed to have arbitrary concentrations.)

Now the derivative will read

$$\frac{\partial E^+}{\partial \mathbf{r}_i} = \sum_{\alpha\beta\gamma klmnj} \frac{\partial E^+}{\partial E_\gamma} \frac{\partial E_\gamma}{\partial \mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \frac{\partial \mathbf{A}_{kl}}{\partial \mathbf{R}_{mn}^*} \frac{\partial \mathbf{R}_{mn}^*}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \mathbf{r}_i}. \quad (2.5)$$

The derivative of the previous, single compositional score is completely reused here;  $\partial E_\gamma / \partial \mathbf{r}_i$  is simply the gradient of the sample's non-duplexing energy. This gradient is the negative of the force which has to act on the involved atoms in a molecular dynamics simulation or during minimization. The procedure of calculating this term has already been described in my thesis and the formula can be simplified to

$$\frac{\partial E^+}{\partial \mathbf{r}_i} = \sum_{\gamma} \frac{\partial E^+}{\partial E_\gamma} \frac{\partial E_\gamma}{\partial \mathbf{r}_i}. \quad (2.6)$$

It should be noted that the different compositions of samples  $c_\gamma$  are represented in the computer with structural models of their respective chemical species  $s$ . If two different compositions share a common species, like one complex and one free ligand, then they will also share the identical coordinate and derivative matrices.

For the given  $E^+$  (2.3) the following derivative with respect to the sample-based scores is valid

$$\begin{aligned} \frac{\partial E^+}{\partial E_\gamma} &= \frac{1}{e} \left( \sum_{\gamma} E_\gamma^e \right)^{\frac{1}{e}-1} \cdot e \cdot (E_\gamma)^{e-1} \\ &= \left( \sum_{\gamma} E_\gamma^e \right)^{\frac{1}{e}(1-e)} \cdot (E_\gamma)^{e-1} \\ &= (E^+)^{(1-e)} \cdot (E_\gamma)^{-(1-e)} \\ &= \left( \frac{E_\gamma}{E^+} \right)^{e-1}. \end{aligned} \quad (2.7)$$

In the case of an 1-norm with  $e = 1$

$$\begin{aligned} \frac{\partial E^+}{\partial E_\gamma} &= \left( \frac{E_\gamma}{E^+} \right)^{e-1} \\ &= 1 \end{aligned} \quad (2.8)$$

will be valid and all the individual gradients will be summed to gain the

total gradient  $E^+$ :

$$\begin{aligned} \frac{\partial E^+}{\partial \mathbf{r}_i} &= \sum_{\gamma} \frac{\partial E^+}{\partial E_{\gamma}} \frac{\partial E_{\gamma}}{\partial \mathbf{r}_i} \\ &= \sum_{\gamma} \frac{\partial E_{\gamma}}{\partial \mathbf{r}_i}. \end{aligned} \tag{2.9}$$

Using larger positive  $e$  will make bad models with large  $E_{\gamma}$  to dominate the energy and its gradient. In the infinity norm or maximum norm case with  $e = +\infty$  only the model of the composition with the highest energy at a time will contribute to the energy and to the forces. All other compositions will be not induce forces during minimization and molecular dynamics.

## 2.2 Implementation

The implementation of this protocol is straight forward.

1. Calculating the compositional penalty energies  $E_{\gamma}$  and gradients  $\partial E_{\gamma}/\partial \mathbf{r}_i$  for all sample compositions  $c_{\gamma}$ .
2. Calculating the total energy  $E^+$  and the weights  $\partial E^+/\partial E_{\gamma}$ .
3. For all  $\gamma$  add the gradients  $\partial E^+/\partial E_{\gamma} \cdot \partial E_{\gamma}/\partial \mathbf{r}_i$  to the total gradient  $\partial E^+/\partial \mathbf{r}_i$ .

Step 2. can be skipped in the case of  $e = 1$  and the compositional gradients can be immediately added to the total gradient.



# Chapter 3

## Dynamical Averaging

Normally the complete kinetic exchange matrix, involving the transitions between all relevant species, is considered for the simulation of NOESY spectra involving chemical exchange. This requires expensive diagonalization and inversion of large matrices.

Landy and Rao have developed the faster method called dynamical averaging to get similar results for calculating (multi-state) chemical exchange NOESY spectra.<sup>7</sup> Here an average relaxation matrix  $\mathbf{R}$  is constructed which is the weighted sum of the individual species' ( $s$ ) relaxation matrices  $\mathbf{R}^s$ . The weights  $p_s$  are the respective fractional concentrations of the species in solution. The initial magnetization matrix  $\mathbf{A}$  is the sum of the weighted initial magnetization of the species applied to the individual atoms (or dummy atoms).

The advantage of dynamical averaging is the use of smaller and symmetric relaxation matrices and the according gain of speed. As the method was demonstrated only for cases where the receptor conformation was neglected it is expected only to be valid for an initial buildup regime. Nevertheless it might be possible to consider the complete complex as one "species". Still the method will be an approximative solution to the problem of calculating NOESY spectra under chemical exchange conditions.

Here I want to present the foundation for gradient-based structural refinement using dynamical-averaging-based NOESY simulations.

### 3.1 Theory

A sample is composed of a ligand and a protein forming species of different properties. In the dynamic averaging approximation, for simplicity, only the

atoms of the ligand are assumed to form a species  $s$ . The different species will in general have different conformations and concentrations. Some of the species will represent bound, others the free state of the ligand.

The differential equation of the time evolution of magnetization is just the same as before [1, (2.1)]. Now, following the result of,<sup>7</sup> the relaxation matrix is

$$\mathbf{R} = \sum_s p_s \mathbf{R}^s \quad (3.1)$$

where  $p_s$  is the fractional concentration of the species  $s$  and defined as

$$p_s = \frac{t_s}{\sum_i^S t_i}. \quad (3.2)$$

The  $t_s$  are the respective life times of the species and  $S$  the number of all species.

The initial magnetization matrix is diagonal and defined as

$$\mathbf{A}^0 = \sum_s p_s \mathbf{A}^{0,s}. \quad (3.3)$$

Including the grouping matrix  $\mathbf{G}$ , not used in [7] but generally useful for NOESY spectra, the grouped peakvolumes for a given mixing time  $\tau$  are

$$\mathbf{A}^g = \mathbf{G} \exp(-\mathbf{R}\tau) \mathbf{A}^0 \mathbf{G}^{\text{tr}}. \quad (3.4)$$

If an penalty energy is defined as before for INPHARMA and general NOESY spectra the gradient of this energy in respect to the species' Cartesian coordinates is

$$\frac{\partial E}{\partial \mathbf{r}_i^s} = \sum_{\alpha\beta\epsilon\xi klmnj} \frac{\partial E}{\partial \mathbf{A}_{\alpha\beta}^g} \frac{\partial \mathbf{A}_{\alpha\beta}^g}{\partial \mathbf{A}_{kl}} \frac{\partial \mathbf{A}_{kl}}{\partial \mathbf{R}_{mn}} \frac{\partial \mathbf{R}_{mn}}{\mathbf{R}_{\epsilon\xi}^s} \frac{\partial \mathbf{R}_{\epsilon\xi}^s}{\partial r_{ij}^s} \frac{\partial r_{ij}^s}{\partial \mathbf{r}_i^s}. \quad (3.5)$$

In the dynamic averaging regime  $\mathbf{R}$  is always symmetric as long as only the ligand is considered. Nevertheless the following derivation of the gradient

will follow the general derivation of (ch. 1), allowing for an non-symmetric  $\mathbf{R}$ .

The result of the first three summations is known as (1.12) and is valid for a general relaxation matrix  $\mathbf{R}$ :

$$\mathbf{Q} \equiv \frac{\partial E}{\partial \mathbf{R}} = -\tau \mathbf{L}^{-1, \text{tr}} \mathbf{J} (\mathbf{M}^{\Lambda}, -\tau \mathbf{\Lambda}) \mathbf{L}^{\text{tr}}. \quad (3.6)$$

The used matriced  $\mathbf{L}$ ,  $\mathbf{\Lambda}$ , etc. are identically defined to the ones used for deriving (1.12), now using the newly defined  $\mathbf{R}$ .

Using 1.12) I yield

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{r}_i^s} &= \sum_{\epsilon \xi mnj} \mathbf{Q}_{mn} \frac{\partial \mathbf{R}_{mn}}{\mathbf{R}_{\epsilon \xi}^s} \frac{\partial \mathbf{R}_{\epsilon \xi}^s}{\partial r_{ij}^s} \frac{\partial r_{ij}^s}{\partial \mathbf{r}_i^s} \\ &= \sum_{\epsilon \xi mnj} \mathbf{Q}_{mn} \frac{\partial (p_1 \mathbf{R}_{mn}^1 + p_2 \mathbf{R}_{mn}^2 + \dots + p_s \mathbf{R}_{mn}^s)}{\partial \mathbf{R}_{\epsilon \xi}^s} \frac{\partial \mathbf{R}_{\epsilon \xi}^s}{\partial r_{ij}^s} \frac{\partial r_{ij}^s}{\partial \mathbf{r}_i^s} \\ &= \sum_{\epsilon \xi mnj} \mathbf{Q}_{mn} p_s \delta_{m\epsilon} \delta_{n\xi} \frac{\partial \mathbf{R}_{\epsilon \xi}^s}{\partial r_{ij}^s} \frac{\partial r_{ij}^s}{\partial \mathbf{r}_i^s} \\ &= \sum_{mnj} p_s \mathbf{Q}_{mn} \frac{\partial \mathbf{R}_{mn}^s}{\partial r_{ij}^s} \frac{\partial r_{ij}^s}{\partial \mathbf{r}_i^s} \end{aligned} \quad (3.7)$$

The next steps can be derived analogously to [1, (2.32-2.34)].

$$\begin{aligned} \mathbf{H}_{ij}^s &\equiv \frac{\partial E}{\partial r_{ij}^s} = \sum_{mn} p_s \mathbf{Q}_{mn} \frac{\partial \mathbf{R}_{mn}^s}{\partial r_{ij}^s} \\ &= p_s - \frac{6}{r_{ij}} \left[ (\mathbf{Q}_{ii} + \mathbf{Q}_{jj}) S_{ij}^{\text{R},s} + (\mathbf{Q}_{ij} + \mathbf{Q}_{ji}) \mathbf{R}^s \right] \end{aligned} \quad (3.8)$$

with  $S_{ij}^{\text{R},s} = \frac{1}{10} \gamma_i^2 \gamma_j^2 \left( \frac{\hbar \mu_0}{4\pi} \right)^2 (r_{ij}^s)^{-6} [J_{ij,0}^s + 3J_{ij,1}^s + 6J_{ij,2}^s]$ .

Now the derivative with respect to the Cartesian coordinates of all atoms can be calculated like here demonstrated for  $x$ :

$$\frac{\partial E}{\partial x_i^s} = \sum_j \frac{\partial E}{\partial r_{ij}^s} \frac{\partial r_{ij}^s}{\partial x_i^s} = \sum_j \mathbf{H}_{ij}^s \frac{x_i^s - x_j^s}{r_{ij}^s} \quad (3.9)$$

In this derivation the indexing for the different matrices (like  $x^s$  and  $\mathbf{R}$ ) follow the same sequence of chemically corresponding atoms. The index  $s$  is indicating that these sequences apply on generally distinguishable atoms. This does not mean that they *must* have non-identical coordinate arrays.

## 3.2 Implementation, Dynamical Averaging

The sequence of operations for calculating the gradient of the penalty energy, in the case of dynamic averaging, according to the results of the previous chapter 'Dynamical Averaging', is the following:

1. Setting up the relaxation matrices  $\mathbf{R}^s$  and initial peak volume matrix  $\mathbf{A}^{0,s}$ , and the grouping matrix  $\mathbf{G}$  following the example of [1, (2.1-2.9)] for each species  $s$ .
2. Setting up the relaxation matrix  $\mathbf{R}$  (3.1)

$$\mathbf{R} = \sum_s p_s \mathbf{R}^s \quad (3.10)$$

and the initial peak volume matrix  $\mathbf{A}^0$  (3.3)

$$\mathbf{A}^0 = \sum_s p_s \mathbf{A}^{0,s}. \quad (3.11)$$

3. Calculating the eigenvector matrix  $\mathbf{L}$ , its inverse  $\mathbf{L}^{-1}$  and the eigenvalues  $\mathbf{\Lambda}$  of  $\mathbf{R}$ .
4. Calculating the matrix of grouped peak volumes (3.4) for a given mixing time  $\tau_m$

$$\mathbf{A}^g = \mathbf{G} \exp(-\mathbf{R}\tau_m) \mathbf{A}_0 \mathbf{G}^{\text{tr}}. \quad (3.12)$$

5. Calculating the matrix  $\mathbf{F}$  [1, (2.21)] (for the given  $\tau_m$ )

$$\mathbf{F}_{kl} = \sum_{\alpha\beta} k_r w_{\alpha\beta} \begin{cases} -f_{small} & \text{if } cal \leq exp - err, \\ +1 & \text{if } cal \geq exp + err, \\ 0 & \text{else.} \end{cases} \cdot \begin{cases} 1 & k \in \text{Peakgroup}_\alpha \text{ and } l \in \text{Peakgroup}_\beta, \\ 0 & \text{else.} \end{cases} \quad (3.13)$$

6. Calculating the matrix  $\mathbf{M}^A$  (1.11) (for the given  $\tau_m$ )

$$\mathbf{M}^A = \mathbf{L}^{\text{tr}} \mathbf{F} \mathbf{A}^0 \mathbf{L}^{-1, \text{tr}}. \quad (3.14)$$

7. Calculating the matrix  $\mathbf{J}(\mathbf{M}^A, -\tau\mathbf{\Lambda})$  (1.13) (for the given  $\tau_m$ )

$$\mathbf{J}(\mathbf{A}, \mathbf{B}) = \begin{cases} \frac{\mathbf{A}_{ij}}{\mathbf{B}_{ii} - \mathbf{B}_{jj}} [\exp(\mathbf{B}_{ii}) - \exp(\mathbf{B}_{jj})] & \text{if } \mathbf{B}_{ii} \neq \mathbf{B}_{jj} \\ \mathbf{A}_{ij} \exp(\mathbf{B}_{ii}) & \text{if } \mathbf{B}_{ii} = \mathbf{B}_{jj}. \end{cases} \quad (3.15)$$

8. Calculating the matrix  $\mathbf{Q}$  (1.12) (for the given  $\tau_m$ )

$$\mathbf{Q} = -\tau\mathbf{L}^{-1, \text{tr}} \mathbf{J}(\mathbf{M}^A, -\tau\mathbf{\Lambda}) \mathbf{L}^{\text{tr}}. \quad (3.16)$$

9. Calculating further matrices  $\mathbf{Q}$  for the remaining mixing times, starting at (3.12). Summing them up to gain the final matrix  $\mathbf{Q}$ .

10. Calculating the matrices  $\mathbf{H}^s$  (3.8)

$$\mathbf{H}_{ij}^s = -\frac{6}{r_{ij}^s} \left[ (\mathbf{Q}_{ii} + \mathbf{Q}_{jj}) S_{ij}^{\text{R},s} + (\mathbf{Q}_{ij} + \mathbf{Q}_{ji}) \mathbf{R}^s \right] \quad (3.17)$$

$$\text{with } S_{ij}^{\text{R},s} = \frac{1}{10} \gamma_i^2 \gamma_j^2 \left( \frac{\hbar\mu_0}{4\pi} \right)^2 (r_{ij}^s)^{-6} [J_{ij,0}^s + 3J_{ij,1}^s + 6J_{ij,2}^s].$$

11. Calculating the forces, like here for  $x$  3.9.

$$\frac{\partial E}{\partial x_i^s} = \sum_j \frac{\partial E}{\partial r_{ij}^s} \frac{\partial r_{ij}^s}{\partial x_i^s} = \sum_j \mathbf{H}_{ij}^s \frac{x_i^s - x_j^s}{r_{ij}^s} \quad (3.18)$$

This procedure is valid if several spectra have to be calculated only differing in mixing time. If any of the parameters influencing the relaxation matrices have changed - like the field, concentrations or kinetic constants - the effective relaxation matrix has to be set-up again and diagonalization is necessary. Then, the resulting  $\mathbf{H}^s$  matrices for different spectra should be summed before the calculation of the forces.

Numerical approximations, just as described in my thesis,<sup>1</sup> can be applied to speed up the diagonalization and the following matrix multiplications. This will involve computing only the eigenvectors to the the smallest eigenvalues and considering the reduced matrix sizes in the matrix multiplications.



# Bibliography

- [1] M. Reese. Sensitivity Enhancement of Liquid-State NMR and Improvement of the INPHARMA Method, 2010, unpublished at 2010-06-01. iii, 2, 3, 5, 6, 7, 8, 11, 16, 17, 18, 19
- [2] R.E. London. Theoretical analysis of the Inter-Ligand Overhauser Effect: A new approach for mapping structural relationships of macromolecular ligands. *Journal of Magnetic Resonance*, 141(2):301–311, 1999. 1
- [3] C.Q. Liu. A method of symmetrization of asymmetric dynamical systems. *Shock and Vibration*, 12:309–315, 2005. 1
- [4] M. Reese. Entwicklung der INPHARMA Methode und Anwendung auf den Tubulin - Epothilon A Komplex, Diplomathesis, University of Göttingen, Germany, 2005. 2, 10
- [5] M. Reese, V.M. Sanchez-Pedregal, K. Kubicek, J. Meiler, M.J.J. Blommers, C. Griesinger, and T. Carlomagno. Structural basis of the activity of the microtubule-stabilizing agent epothilone A studied by NMR spectroscopy in solution. *Angewandte Chemie - International Edition*, 46(11):1864–1868, 2007. 2
- [6] P.F. Yip and D. Case. A new method for refinement of macromolecular structures based on nuclear overhauser effect spectra. *Journal of Magnetic Resonance*, 83:643–648, 1989. 5
- [7] Stevonn B. Landy and D. Nageswara Rao. Dynamical NOE in Multiple-Spin Systems Undergoing Chemical Exchange. *Journal of Magnetic Resonance*, 81:371–377, 1998. 15, 16