TU | TECHNISCHE UNIVERSITÄT WIEN

D I P L O M A R B E I T

Accessing Backbone Dihedral Angle Populations of Intrinsically Disordered Proteins via NMR Spin Relaxation: A Maximum Entropy Approach

Ausgeführt am Institut für

Strukturbiologie und Computational Biology

der Universität Wien

unter der Anleitung von Univ.-Prof. Dr. Robert Konrat

durch

Clemens B. Kauffmann, BSc

Kundmanngasse 3/11, 1030 Wien

Wien, Oktober 2017

_____

Clemens B. Kauffmann, BSc

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

_____	_____
Date	Signature

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

_____	_____
Datum	Unterschrift

Danksagung

Eine wissenschaftliche Arbeit verrät naturgemäß nur wenig über jene, die sie ermöglicht haben. Mein Dank gilt:

- meinem Betreuer, Prof. Robert Konrat; für das Vertrauen, den Zuspruch und die intellektuellen Herausforderungen.

- Prof. Bojan Zagrovic; obwohl nur ein Gast seiner Gruppe, durfte ich mich wirklich zuhause fühlen.

- Prof. Gerhard Zifferer; für seine offene und freundliche Art, die mich - über Umwege - zu dieser Arbeit geführt hat. Ich bin dankbar, ihn noch kennengelernt haben zu dürfen.

- meinen Kollegen, Lukas Bartonek, David Weichselbaum und Mathias Kreuter; für ihre Hilfsbereitschaft, Geduld, Kompetenz und die anregenden Gespräche.

- meinen Eltern; für ihre emotionale, finanzielle und auch inhaltliche Unterstützung im Laufe meines Studiums.

- meiner Lebensgefährtin, Nina; für die gemeinsam verlebten Studienjahre.

Abstract

Unlike folded proteins, the dynamic structural ensembles of intrinsically disordered proteins (IDPs) do not possess fixed tertiary structures. To date, NMR spectroscopy is the most prevalent and versatile technique to address the analytical challenges involved in the characterization of IDPs. Still, the amount of experimental observables is inherently insufficient to capture the underlying ensembles in full detail. To characterize IDP backbone dihedral angles, previous studies have proposed a Maximum Entropy (MaxEnt) approach, which derives a representative solution from a convolution of specific prior assumptions and supplementary experimental constraints. This work promotes a more modest implementation of MaxEnt that attempts to distinguish more clearly between experimental and assumptive information content. The procedure is illustrated by assessing the theoretical feasibility of cross-correlated relaxation (CCR) experiments; a method still mostly unexplored in the study of IDPs. The proposed protocol can serve as a general framework to evaluate and implement combinations of NMR experiments not limited to CCR interactions. The particular set of CCR rates considered in this work prove highly proficient in probing IDP dihedral angle populations, indicating the unharnessed potential CCR experiments offer in the structural and dynamic characterization of IDPs.

Kurzfassung

Anders als gefaltete Proteine weisen die dynamischen Konformationsensembles von intrinsisch ungeordneten Proteinen (kurz: IDPs, englisch: *intrinsically disordered proteins*) keine festen Tertiärstrukturen auf. Kernresonanzspektroskopie (kurz: NMR, englisch: *nuclear magnetic resonance*) ist die aktuell gängigste und vielseitigste Methode, um den analytischen Herausforderungen in der Charakterisierung von IDPs zu begegnen. Dennoch ist die Zahl der experimentellen Parameter grundsätzlich nicht ausreichend, um die zugrunde liegenden Ensembles eindeutig zu bestimmen. In früheren Studien wurde zur Charakterisierung dihedraler Winkel von IDP Backbones das Maximum Entropy (kurz: MaxEnt, deutsch: *Maximale Entropie*) Kalkül vorgeschlagen. Dabei wird eine repräsentative Lösung aus a priori Annahmen und experimentellen Einschränkungen abgeleitet. Diese Arbeit vertritt eine konservativere Implementation des MaxEnt Kalküls, in der deutlicher zwischen experimenteller und a priori Information unterschieden wird. Illustriert wird das Verfahren anhand einer theoretischen Evaluation von kreuzkorrelierten Relaxations-Experimenten (kurz: CCR, englisch: *cross-correlated relaxation*); eine Methode, die bis heute kaum Anwendung bei IDPs gefunden hat. Das hier vorgestellte Protokoll kann in seiner Grundstruktur zur Bewertung und Implementation von Kombinationen aus NMR-Experimenten herangezogen werden, die sich nicht auf CCR-Interaktionen beschränken müssen. Die in dieser Arbeit erwogenenen CCR-Raten zeigen sich zur Populationsbestimmung von dihedralen Winkeln in IDPs sehr gut geeignet; CCR-Experimente besitzen demnach in der strukturellen und dynamischen Charakterisierung von IDPs noch unausgeschöpftes Potential.

Contents

Chapter 1

Introduction

For considerable time biochemistry has been dominated by the "structure - function" paradigm which assumes that protein function requires a well-defined 3D structure. While still highly influential, this concept started to be challenged in the late 90s [1–4]. Nowadays, the functional importance of conformational plasticity in proteins is indisputable: *Intrinsically disordered* regions of at least 30 residues in length are estimated to occur in 10-35 % of prokaryotic and 15-45 % of eukaryotic proteins [5, 6]. X-ray crystallography, the workhorse of structural biology, is not fit to deal with the structural flexibility of intrinsically disordered proteins (IDPs). NMR spectroscopy, the most important alternative in protein structure determination, does not require crystallization. Thus, it represents a key technique for studying structural and dynamics properties of IDPs [7, 8].

Unlike globular proteins, IDPs cannot be described by mapping experimental data onto a singular structure. Instead, the observables are averages of a dynamic conformational ensemble. As a result, IDPs yield more convoluted experimental data and are more challenging to analyze.

For lack of a singular tertiary structure, an evident approach is to instead characterize IDPs in terms of local secondary structure propensities, which in return can be used to derive ensemble representations. Different NMR parameters have been suggested and used in this regard [7, 9], e.g. chemical shifts [10], scalar couplings [11], residual dipolar couplings [12–14] and paramagnetic relaxation enhancement [15]. However, the number of experimental parameters is inherently insufficient to uniquely characterize the underlying ensemble; due to the large conformational freedom, the set of compatible ensembles is highly degenerate. Still, sophisticated methods exist to construct representative ensembles of IDPs from insufficient data, comprehensive overviews have been given in e.g. [7, 9, 16]. Bayesian statistics have also been suggested to tackle the underdetermined nature of this problem [17].

A more modest approach considers only the local backbone which most NMR parameters actually report on. Secondary structures can be characterized by their dihedral angles ϕ and ψ, a well-established, practical method applicable for both folded and disordered proteins [7, sec. 4.2]. Instead of a singular angle pair, an IDP-residue samples a distribution of dihedral angles in Ramachandran space. Still, even these backbone (ϕ,ψ)-populations cannot be determined unambiguously. In an approach closely related to Bayesian statistics, Massad et al. [18] and Mantsyzov et al. [19, 20] independently suggested different implementations of the Maximum Entropy (MaxEnt) calculus to define a representative solution for this underdetermined problem, using readily available NMR parameters like chemical shifts, scalar couplings and short range nuclear Overhauser effects.

This work suggests an extension of the established protocols in two general directions: The first one concerns the choice of NMR parameters. Scalar couplings in particular are used for their angular dependencies according to the Karplus relationship [21]. The need for parametrization [22] and the limited possibilities along the backbone lead to the development of cross-correlated relaxation (CCR) experiments as a complementary method in the late 90s [23, 24]. While arguably more complex, a multitude of possible interactions along the protein backbone can be probed, elucidating both structural and dynamic properties. Backbone dihedral angles of folded proteins have been shown to be accessible using suitable combinations of CCR experiments [25]. While promising first results do exist [26], a systematic inclusion of CCR rates in the characterization of IDPs has not been attempted to date.

The second suggestion concerns the implementation of the Maximum Entropy framework itself: A key characteristic of MaxEnt approaches is that experimental information needs to be processed relative to some initial assumptions encoded in the so-called prior. In the previous studies of Massad et al. [18] and Mantsyzov et al. [19, 20] only the convoluted results based on a fixed set of experiments and a singular prior have been presented; their individual contributions have not been assessed. Since no distinction is made between experimental and assumptive information content, it is not clear to what extent the experiments would be capable to contradict the prior assumptions. Here, a MaxEnt implementation is proposed which differentiates more intuitively between prior and experimental information. It can be used as a general framework for assessing and designing experimental protocols probing the protein backbone. The implementation is entirely in-silico and based solely on CCR experiments.

Chapter 2

Cross-Correlated Relaxation (CCR)

2.1 Spin relaxation theory

Relaxation is the dynamic process by which spins return to their equilibrium state. It is caused by fluctuating magnetic fields in the surroundings, the so-called *lattice*. In liquids these fluctuations generally result from dynamic reorientation of spins relative to the static magnetic field due to molecular motion. Thus, relaxation encodes structural and dynamic properties of the investigated system. Naturally, the local magnetic fields of spins can also be affected by spins in close proximity. Correlated motion of such spins gives rise to cross-correlated relaxation (CCR) which can be experimentally investigated.

The CCR experiments considered in this work probe the protein backbone to extract the dihedral angles ϕ and ψ. The corresponding physical models are based on Redfield's theory of relaxation [27, 28], which will briefly be touched upon in this section. It mainly follows the reviews of Kumar et al. [29] and Murali and Krishnan [30], supplemented by the books of Kowalewski and Mäler [31] and Cavanagh et al. [32].

Starting point is the Liouville-von Neumann equation (2.1). It describes the evolution of a spin system over time t:

$$\frac{d\hat{\rho}(t)}{dt} = -i[\hat{H}_0 + \hat{H}_1(t), \hat{\rho}(t)] \tag{2.1}$$

The spin system is represented by the density operator $\hat{\rho}$. $[\hat{H}_0 + \hat{H}_1(t), \hat{\rho}(t)]$ denotes the commutator of $\hat{\rho}$ and the Hamiltonian, which is assumed to consist of a large, unperturbed time-independent component $\hat{H}_0$ and a small, time-dependent perturbation $\hat{H}_1(t)$ causing the relaxation.

The first step in solving this equation is the so-called interaction frame transformation,

which removes the time-dependent variations of $\hat{\rho}(t)$ with $\hat{H}_0$.

$$\frac{d\hat{\rho}^*(t)}{dt} = -i[\hat{H}_1^*(t), \hat{\rho}^*(t)] \tag{2.2}$$

where

$$\hat{\rho}^*(t) = \exp(i\hat{H}_0 t)\hat{\rho}(t)\exp(-i\hat{H}_0 t)$$
$$\hat{H}_1^*(t) = \exp(i\hat{H}_0 t)\hat{H}_1(t)\exp(-i\hat{H}_0 t)$$

The differential equation (2.2) is then solved approximatively. The corresponding integral equation is:

$$\hat{\rho}^*(t) = \hat{\rho}^*(0) - i\int_0^t [\hat{H}_1^*(t'), \hat{\rho}^*(t')]dt' \tag{2.3}$$

Both sides contain the same, still unknown term $\hat{\rho}^*(t)$. By substituting $\hat{\rho}^*(t')$ with equation (2.3) again, an iterative scheme is obtained. Only the first three terms are kept, yielding a second order approximation:

$$\hat{\rho}^*(t) = \hat{\rho}^*(0) - i\int_0^t [\hat{H}_1^*(t'), \hat{\rho}^*(0)]dt' - \int_0^t \int_0^{t'} [\hat{H}_1^*(t'), [\hat{H}_1^*(t''), \hat{\rho}^*(0)]]dt''dt' \tag{2.4}$$

By taking the time derivative, an expression for $d\hat{\rho}^*(t)/dt$ is recovered:

$$\frac{d\hat{\rho}^*(t)}{dt} = -i[\hat{H}_1^*(t), \hat{\rho}^*(0)] - \int_0^t [\hat{H}_1^*(t), [\hat{H}_1^*(t'), \hat{\rho}^*(0)]]dt' \tag{2.5}$$

In the next step the spin systems are treated as an ensemble. $\overline{\hat{H}_1^*(t)}$ averaged over this "ensemble of ensembles" is zero. If not, the non-zero average component can always be incorporated into $\hat{H}_0$. This argument reflects the very nature of relaxation: The random, undirected fluctuating magnet fields are what equilibrates the sample.

The fluctuations are further assumed to occur on time scales much smaller than the relaxation time of the spin system. This way, $\hat{\rho}^*(0)$ and $\hat{H}_1^*(t)$ factorize and can be averaged independently. Consequently, the first commutator $[\hat{H}_1^*(t), \hat{\rho}^*(0)]$ in equation (2.5) vanishes.

In a similar argument, the slow variation of $\hat{\rho}^*(t)$ compared to $\hat{H}_1^*(t)$ and the short time period t justify the substitution of $\hat{\rho}^*(0)$ with $\hat{\rho}^*(t)$. In addition, the short correlation time allows for the extension of the upper integration limit to infinity. After substituting $\tau = t - t'$ and making a finite spin temperature correction $\hat{\rho}^*(t) = \hat{\rho}^*(t) - \hat{\rho}_0$, where $\hat{\rho}_0$ is the equilibrium density matrix, the *master equation* is obtained:

$$\frac{d\overline{\hat{\rho}^*(t)}}{dt} = -\int_0^\infty \overline{[\hat{H}_1^*(t), [\hat{H}_1^*(t-\tau), \hat{\rho}^*(t) - \hat{\rho}_0]]}d\tau \tag{2.6}$$

4

Equation (2.6) is now usually reformulated in one of two ways: The matrix formalism represents elements of $\overline{\hat{\rho}^*(t)}$ as eigenstates of the unperturbed Hamiltonian $\hat{H}_0$; the operator representation employs nuclear spin angular momentum operators. Both depictions are equivalent and require considerable algebraic effort. Which one is used usually depends on the the intended application and preference. Here, the operator formalism will be used to illustrate the essentials of CCR.

First, the relaxation Hamiltonian in the laboratory frame is written as a product of irreducible tensors:

$$\hat{H}_1(t) = \sum_q (-1)^q A_{-q} F_q(t) \tag{2.7}$$

A_q are time-independent operators acting only on the spin system, $F_q(t)$ are time-dependent, stochastic functions of the lattice variables and q denotes the rank of the tensors. Note that all dynamic relaxation contributions are attributed to the lattice.

The spin operators are then expanded in terms of basis operators, which exhibit transformation properties similar to equation (2.2).

$$\exp(i\hat{H}_0 t) A_q \exp(-i\hat{H}_0 t) = \sum_p A_{p,q} \exp(i\omega_{p,q} t)$$

$$\exp(i\hat{H}_0 t) A_{-q} \exp(-i\hat{H}_0 t) = \sum_p A_{p,-q} \exp(i\omega_{p,-q} t)$$

$$\hat{H}_1^*(t) = \sum_q \sum_p (-1)^q F_q(t) A_{p,-q} \exp(i\omega_{p,-q} t)$$

$\hat{H}_1^*(t)$ is then substituted into the master equation (2.6). Due to the double commutator, components of the obtained sum are modulated by an $\exp(i(\omega_{p,-q} + \omega_{p',-q'})t)$-term. The *secular approximation* considered here neglects rapidly oscillating contributions for which $\omega_{p,-q} + \omega_{p',-q'} \neq 0$. Omitting the ensemble average overbars, the master equation reads:

$$\frac{d\hat{\rho}^*(t)}{dt} = -\sum_q \sum_p [A_{p,-q}, [(-1)^q A_{p,q}, \hat{\rho}^*(t) - \hat{\rho}_0]] \int_0^\infty G_q(\tau) \exp(-i\omega_{p,q}\tau) d\tau \tag{2.8}$$

The integral is called the *spectral density function*. It is the Fourier transform of the *correlation function* $G(\tau)$:

$$G_q(\tau) = \overline{F_q(t)(-1)^q F_{-q}(t-\tau)} \tag{2.9}$$

The imaginary components of the spectral density functions cause the so-called *dynamic frequency shifts*, which are often small enough to be neglected. To further generalize, the Hamiltonian is now considered as a sum of n relaxation contributions:

$$\hat{H}_1(t) = \sum_n \hat{H}_{1,n}(t) \tag{2.10}$$

This leads to the appearance of cross-terms in the correlation function:

$$G_q(\tau) = \sum_n \overline{F_{q,n}(t)(-1)^q F_{-q,n}(t-\tau)} + \sum_{\substack{n<n',\\n\neq n'}} \overline{F_{q,n}(t)(-1)^q F_{-q,n'}(t-\tau)} \qquad (2.11)$$

The second term on the right-hand side of (2.11) vanishes, if the random motions facilitating n and n' are uncorrelated. If they are not, cross-correlated relaxation occurs.

The most common interferences are due to dipolar and anisotropic chemical shift interactions. Relaxation rates Γ can be predicted by plugging the corresponding Hamiltonians and tensors into the newly obtained master equation (2.8) [24, 29]. Skipping the extensive case-specific calculations, the obtained expressions exhibit a common general form:

$$\Gamma_{AB,CD}^{DP,DP}(\theta_{AB,CD}) = k_{DP}^2 \frac{\gamma_A\gamma_B\gamma_C\gamma_D}{\|\overrightarrow{AB}\|^3\|\overrightarrow{CD}\|^3} \times \frac{1}{2}(3\cos^2\theta_{AB,CD} - 1) \times J(\omega) \qquad (2.12)$$

$$\Gamma_{AB,X'}^{DP,CSA}(\theta_{AB,X'}) = k_{DP}k_{CSA}\frac{\gamma_A\gamma_B}{\|\overrightarrow{AB}\|^3}B_0\gamma_{X'}(\Delta\sigma_{X'}) \times \frac{1}{2}(3\cos^2\theta_{AB,X'} - 1) \times J(\omega) \quad (2.13)$$

$$\Gamma_{X',Y'}^{CSA,CSA}(\theta_{X',Y'}) = k_{CSA}^2 B_0^2\gamma_{X'}\gamma_{Y'}(\Delta\sigma_{X'})(\Delta\sigma_{Y'}) \times \frac{1}{2}(3\cos^2\theta_{X',Y'} - 1) \times J(\omega) \qquad (2.14)$$

- A,B, C and D are arbitrary nuclei subject to dipolar (DP) coupling.

- X' and Y' are nuclei exhibiting chemical shift anisotropy (CSA).

- k_{DP} and k_{CSA} are interaction-specific prefactors and physical constants.

- γ is the gyromagnetic ratio.

- B_0 is the field strength of the static magnetic field.

- $\Delta\sigma$ is the CSA tensor.

- θ is the projection angle.

- $J(\omega)$ is the frequency-dependent *spectral density function*.

$J(w)$ modulates Γ at tensor-specific frequencies ω [24]. Thus, it encodes the motions inducing the relaxation process. However, the dynamic information it provides requires careful consideration. In most cases the correlations functions, i.e. the underlying stochastic lattice functions $F(t)$, are not trivial to model. The commonly used "model-free" approach of Lipari and Szabo [33] even refrains from describing the dynamics in too much detail. At its heart, it assumes global and local motions to be factorized, which is a reasonable concept for globular proteins but inapplicable for IDPs. Since a generally accepted dynamic model for IDPs has not been devised to date, predicting and interpreting $J(w)$ is not a straightforward task. Spin relaxation experiments will certainly play a key

role in addressing this challenge, see e.g. [8, 34–36].

This work, however, focuses on the projection angle θ. It reflects the local molecular structure surrounding the probed spins, which allows for the intended extraction of the backbone dihedral angles ϕ and ψ.

2.2 CCR determination of backbone dihedral angles

Using CCR experiments to measure backbone dihedral angles was first proposed and implemented by Reif et al. [23], who probed $\Gamma_{H_i^\alpha C_i^\alpha, H_{i+1}^N N_{i+1}}$ to determine ψ_i. In the following years a multitude of additional experiments were devised to characterize the protein backbone in a similar fashion, see the review of Schwalbe et al. for a structured overview [24].

The general scheme is straightforward: The relaxation rate Γ is modulated by the angular-dependent second order Legendre polynomial $1/2(3\cos^2\theta - 1)$. By relating θ to the backbone geometry, Γ can be expressed as a function of ϕ and/or ψ depending on the particular interaction.
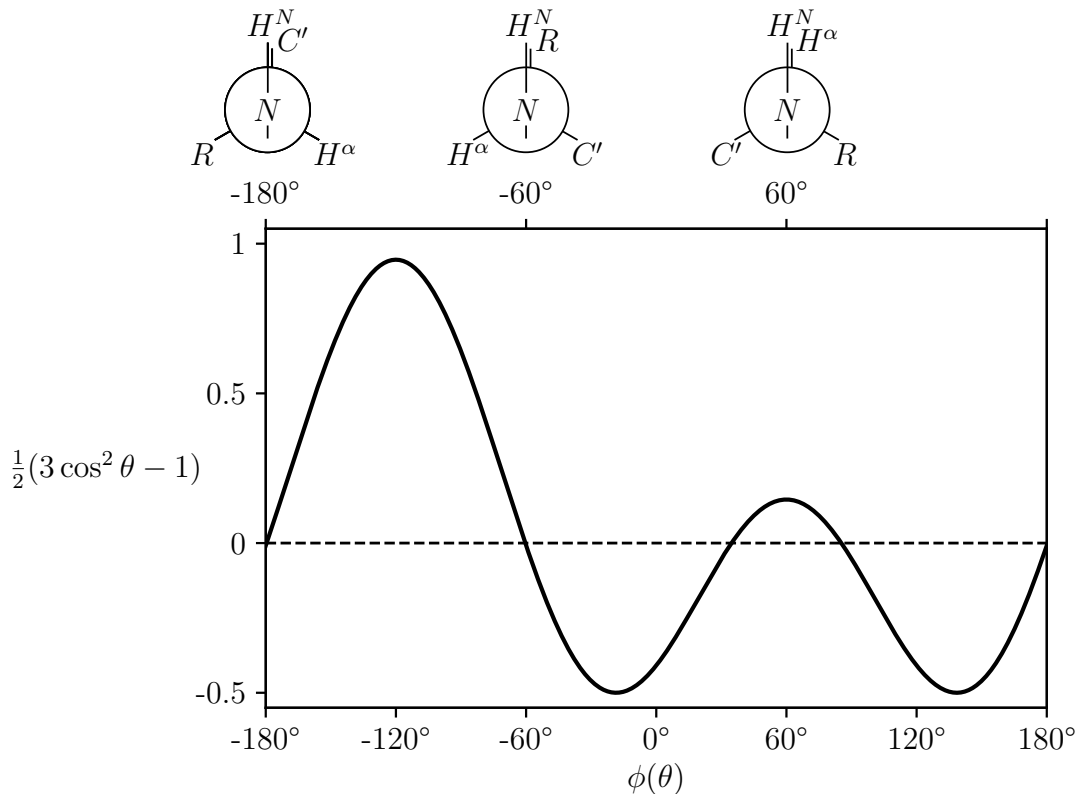

Figure 2.1: The angular modulation $1/2(3\cos^2\theta - 1)$ of the CCR rate $\Gamma_{H_i^N N_i, H_i^\alpha C_i^\alpha}(\phi)$ [37] with $\cos\theta = 0.163 - 0.819\cos(\phi - 60°)$. θ denotes the projection angle between the H_i^N-N_i and the H_i^α-C_i^α bonds, Newman projections of the corresponding backbone geometries are indicated above for selected ϕ-angles.

In many cases these functions bear a close resemblance to the Karplus relationship observed for scalar couplings [21]. However, CCR rates involve dynamic contributions, require no empirical parametrization [22], exhibit more combinatorial possibilities and can depend on more than one dihedral angle.

Figure 2.1 depicts the simple to illustrate relationship between the H_i^α-C_i^α and the H_i^N-N_i internuclear spin vectors, first investigated by Pelupessy et al. [37]. Newman projections of the bond geometry are sketched for $\phi = -180°$, $-60°$ and $60°$. Evidently, the function relating θ and ϕ is not bijective due to the ambiguous $\cos^2\theta$ term. Except for the maximum at $\phi = -120°$, multiple dihedral angles could lead to the same relaxation rate. Thus, if a given set of angles was to be characterized unambiguously, a suitable combination of CCR rates must be measured.

To address this shortcoming, Kloiber et al. [25] proposed a protocol employing five CCR rates and one scalar coupling constant: $\Gamma_{H_{i-1}^\alpha C_{i-1}^\alpha, C_{i-1}'}(\psi_{i-1})$ [38], $\Gamma_{C_{i-1}', H_i^\alpha C_i^\alpha}(\phi_i)$ [39], $\Gamma_{H_i^N N_i, H_i^\alpha C_i^\alpha}(\phi_i)$ [37], $\Gamma_{H_{i-1}^\alpha C_{i-1}^\alpha, H_i^N N_i}(\psi_{i-1})$ [23], $\Gamma_{H_{i-1}^\alpha C_{i-1}^\alpha, H_i^\alpha C_i^\alpha}(\phi_i, \psi_{i-1})$ [40] and $^3 J_{C', C'}(\phi_i)$ [41].

These experiments were shown to resolve the ambiguities in (ϕ_i, ψ_{i-1})-space for most of the accessible residues in Ubiquitin. Prolines, lacking the H^N-N bond, glycines, featuring two H-C^α bonds, and residues following glycines were excluded from the analysis. The spectral density functions were fitted using the model-free approach of Lipari and Szabo [33]. Omitting the details, the fitting procedure was found to be quite resilient: If not too extreme, singular deviations from predicted rates did not drastically affect the obtained results, indicating an inherent error correction resulting from the combination of multiple experiments. This effect is noteworthy as it might apply for other systems as well, possibly reducing the effect oversimplified dynamics might introduce.

2.3 Applicability of CCR experiments in IDPs

In the previous section CCR experiments were established as a tool to probe consecutive pairs of ϕ- and ψ-angles in the protein backbone. While characterizing globular proteins has proven successful, the dynamic and structural properties of IDPs are considerably more complex. The clear distinction between dynamics and structure is further challenged by the fact that the "structure" of an IDP should rather be thought of as a dynamic ensemble of interchanging conformations.

But not all complications can be addressed at once. Here, an initial simplification is made by assuming that all conformations are, at least locally, modulated by the same averaged spectral density function, effectively factoring out the dynamic contributions.

Clearly, dynamic modulations might differ for distinct regions in (ϕ,ψ)-space and even for coinciding (ϕ,ψ)-pairs; spatial, angular-dependent tensorial components could require time-dependent treatment. However, as first experiments suggest [8, 26], the results could also be dominated by the geometrical contributions as well as the afore mentioned cumulative error correction. The ignorance of the underlying dynamics could also be compensated by experimentally determined spectral densities.

Still, even when considering the CCR rates as dynamically unaffected, an important question remains: Could diverse structural ensembles even realistically be characterized?

Consider a given (ϕ,ψ)-pair rather as a discretized[1] probability distribution $p(\phi, \psi)$ with the normalization condition given as:

$$\sum_{\phi} \sum_{\psi} p(\phi, \psi) = 1 \tag{2.15}$$

A globular protein can essentially be treated as a singular structure; $p(\phi, \psi)$ resembles a narrow, localized distribution in (ϕ,ψ)-space. As shown by Kloiber et al. [25], such distributions can be determined unambiguously by measuring a suitable combination of (mostly) CCR rates Γ, which are in fact ensemble averages of $p(\phi, \psi)$:

$$\langle \Gamma(\phi, \psi) \rangle = \sum_{\phi} \sum_{\psi} \Gamma(\phi, \psi) p(\phi, \psi) \tag{2.16}$$

For an IDP this becomes more complicated. $p(\phi, \psi)$ corresponding to a heterogeneous structural ensemble populates many different regions in (ϕ,ψ)-space. If not a highly localized distribution is considered, the same set of experimental parameters cannot resolve the resulting ambiguities to uniquely characterize $p(\phi, \psi)$. There are too few observables for too many possible conformer populations: The problem is *underdetermined.*

This leads to the key question this work is trying to answer: *To what extent are arbitrary distributions in (ϕ,ψ)-space accessible using cross-correlated relaxation experiments?*

Put more abstractly: To what extent can arbitrary probability distributions be reproduced given only a set of probability-weighted average function values?

This raises a conceptual problem: We already established that an unambiguous relationship between a probability distribution and its corresponding set of mean function values does not exist for every case. But if we try to reproduce some sample distribution with its derived values, we already know it to be our "best" possible solution in advance. How do we prevent this knowledge from biasing our predictions?

---

[1] A continuous notation would be equally applicable. The discrete representation simply retains a higher consistency with the following notes and the subsequent numerical implementation.

Clearly, our target function should not include the sample distribution in any way. While we can assess the quality of a method by comparing it with the original distribution, we should employ only the derived constraints, not the distribution itself. Furthermore, introducing any kind of empirical parameters into the target function might bias our results towards the chosen set of sample distributions. We want the method to only reflect the information present in the set of experimentally accessible parameters.

The *Maximum Entropy principle* is a calculus that aims to provide a mathematical foundation for problems of this kind. While Maximum Entropy (MaxEnt) approaches have been used in IDP characterization before [18–20], CCR rates have not been considered until now. Furthermore, the particular MaxEnt implementation in this work differs from previous studies on a very fundamental level, as will be elaborated later on.

Chapter 3

The Maximum Entropy Framework

The *Maximum Entropy principle* was pioneered by E.T. Jaynes in 1957 [42] and builds on the foundations of Shannon's entropy measure introduced in 1948 [43]. In the previous chapter we asked: To what extent can arbitrary probability distributions be reproduced given only a set of probability-weighted average function values? Jaynes answer is: "In making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known." [42, p. 623] Naturally, the generality of this statement allows for a vast range of applications: Web of Science lists 21,687 results with the topic 'maximum entropy' between 1957 and 2016.

This chapter will explore the foundations that lead to the Maximum Entropy (MaxEnt) principle. It should be mentioned that derivations, axioms and interpretations of MaxEnt and entropy itself are topics in their own rights. This work is not meant to provide an exhaustive review or rigorous proofs. It will mostly follow Shannon and Jaynes in an information theory inspired narrative in discrete notation, because it provides a well established and intuitive notion of entropy as a numerical measure of uncertainty or missing information.

This notion, however, is not necessary to justify MaxEnt, in fact its applicability ranges beyond Jaynes' original intents and different, more abstract axiomatizations have been proposed. Entropy and MaxEnt are still an active field of research, but most improvements on Jaynes' original work do not offer particular advantages in a practical application, so only the necessary fundamentals will be covered. Noteworthy remarks and implications of later developments and controversies will be addressed in the last section of this chapter.

3.1 Maximum Shannon entropy

In "A Mathematical Theory of Communication" [43] Shannon considered the following: Given a set of n events occurring with probabilities $p_1, ..., p_n$, can we define a measure $H(p_1, ..., p_n)$ of how uncertain we are of the outcome?

Shannon argued that if such a measure H existed, it should satisfy a set of axiomatic continuity, monotonicity and consistency requirements. He could show that the resulting H is uniquely defined. Referring to its formal relationship with Boltzmann's H-theorem he called it *entropy*:

$$H(p_1, ..., p_n) = -K \sum_{i=1}^{n} p_i \log p_i \tag{3.1}$$

K is any positive constant and can be considered as a choice of base for log. Consistent with $\lim_{p \to 0} p \log p = 0$, the convention $0 \log 0 = 0$ is used.

While Shannon's axioms were the first ones used to derive equation (3.1), they are not unique. Later derivations with weaker hypothesis include prominent formulations by Khinchin [44, 45] and Faddeev [46] as well as generalizations by Rényi [47] and Tsallis [48], see e.g. [49, 50] for further reading. Faddeev's theorem does not contain Shannon's original monotonicity requirements and is often considered more intuitive in its characterization of H; slightly adapted from Conrad [51], Faddeev's axioms read:

For $n \geq 2$, let $\Delta_n = \{(p_1, ..., p_n) : 0 \leq p_i \leq 1, \sum p_i = 1\}$. Suppose on each Δ_n a function $H_n : \Delta_n \to \mathbb{R}$ is given with the following properties:

1. *Continuity: Each H_n is continuous.*

2. *Symmetry: Each H_n is a symmetric function of its arguments.*

3. *Consistency: For $n \geq 2$ and all $(p_1, ..., p_n) \in H_n$ with $p_n > 0$, and $t \in [0, 1]$,*
   $H_{n+1}(p_1, ..., p_{n-1}, tp_n, (1-t)p_n) = H_n(p_1, ..., p_{n-1}, p_n) + p_n H_2(t, 1-t).$

While the continuity and symmetry requirements seem intuitive, the consistency condition might need some interpretation: If we can describe the probability p_i of one event as two separate events with probabilities t and $1 - t$, then the total entropy equals the entropy of the combined events plus a p_i-weighted entropy for the separate events. Qualitatively speaking: No matter in what ways a given probability distribution can be dissected, its entropy has to be the same for every way possible.

These requirements suffice to derive the entropy expression (3.1) (e.g. [52]), the choice of constants is a matter of convention.

Following Shannon's original narrative again, we now examine the simple case of only two

possible events $H(p, (1 - p))$ in Figure 3.1 to illustrate a few properties of H which hold true for the general case:

- $H(p_1, ..., p_n) > 0$, if for $i \in n : 0 \leq p_i < 1$.
  If we are *uncertain* about the outcome, the entropy is always bigger than 0.

- $H(p_1, ..., p_n) = 0$, if for any $i \in n : p_i = 1$.
  If there is *no uncertainty* about the outcome, the entropy is 0.

- $H(p_1, ..., p_n)$ is *concave*.

- $H(p_1, ..., p_n)$ is maximal, if for all $i \in n : p_i = {}^1/n$.
  If all outcomes are equally likely, the entropy is maximal.
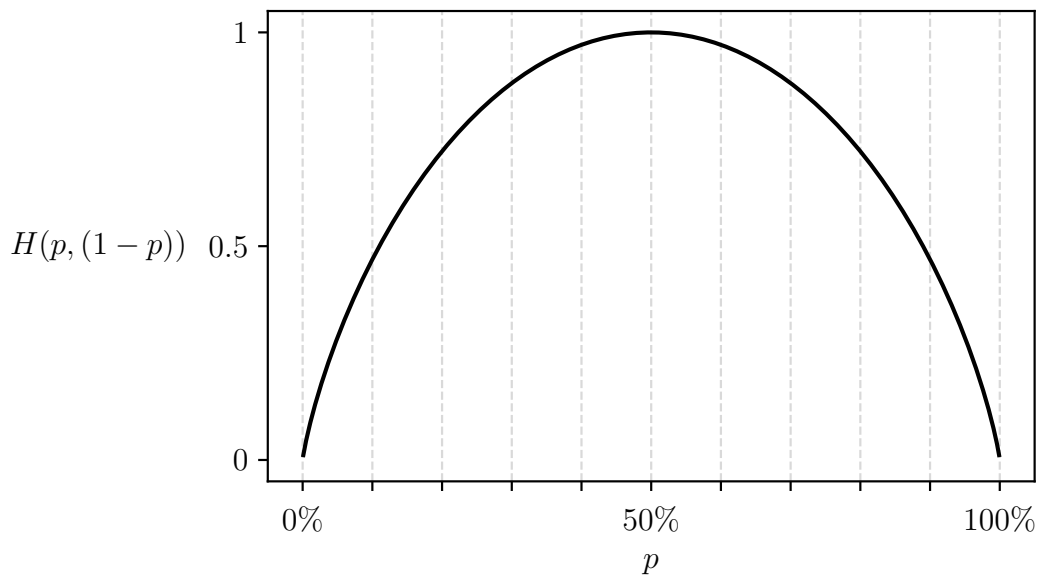

Figure 3.1: The binary entropy function of two possibilities p and $(1 - p)$ for log base 2.

Figure 3.1 can be seen as the representation of a coin toss: If it is a perfectly fair coin, our uncertainty about the outcome is highest. If it is biased in some way, our uncertainty is lower. If we know it always comes up one way, there is no uncertainty about the outcome.

It is not unexpected that out of all possible distributions over $p_1, ..., p_n$ the *uniform distribution* ($\forall i \in n : p_i = {}^1/n$) has maximum entropy. Shannon treated the uniform distribution as a special case and Khinchin even used this property as an axiom. Jaynes argued that this can be seen as a deeper foundation of Laplace's *principle of insufficient reason*: If there is no reason to think otherwise, we should assign equal possibilities to all events. But the notion of entropy as a numerical uncertainty measure allows us to extend this principle to cases where we do have reason to think otherwise. According to Jaynes, the probability distribution that uniquely reflects our limited information is the one which retains the highest uncertainty while reproducing the imposed constraints; any other dis-

tribution with lower entropy contains more information than originally supplied. This is the Maximum Entropy principle.

While Shannon acknowledged the formal relationship between his and Gibbs' entropy measure, Jaynes argued that they are in fact the same concept. According to Jaynes, Gibbs was the first one to apply MaxEnt for the special case of statistical mechanics [42, 53] when he chose "the distribution in phase which without violating this condition gives the least value of the average index of probability of phase" [54, p. 143]; "the average index of probability with its sign reversed corresponds to entropy" [54, p. 44].

However, this work does not focus on the relationship between statistical mechanics and information theory, but on MaxEnt as a general tool for probabilistic inference. So what are the features of a MaxEnt solution based on the concepts introduced up to this point?

Using the coin toss (Figure 3.1) as an example again, we have seen that without additional constraints MaxEnt predicts equal probabilities for heads and tails. This does not necessarily mean that the coin has to be fair of course. MaxEnt does not return the "true" but rather the most "ignorant" or "noncommittal" distribution based on the information supplied. If the experimental observations contradict the MaxEnt solution, it implies that we don't know enough about the coin: There might be unknown missing constraints.

What does a constrained prediction look like? Setting formalisms aside for now, we can qualitatively state the following: The general MaxEnt prediction is characterized as the "most uniform" distribution consistent with the imposed constraints; the unconstrained MaxEnt solution acts as a pressure towards uniformity.

Speaking in terms of Bayesian probability theory, this is of course based on the implicit assumption of equal *a priori* probabilities. While this choice might seem plausible in many cases, like a coin toss, we would not argue that this holds true for the general case. If we allow for the outcome of the coin landing on its edge, MaxEnt would predict a highly unintuitive *posterior* probability of $1/3$ for all three events. While we might not be able to quantify our *prior* "state of ignorance", the uniform distribution does not seem representative.

This notion was not anticipated by Jaynes in his original adoption of MaxEnt in 1957, but five years later [55]. While still insisting on the generality of MaxEnt, he identified the uniform distribution as a special prior contained in the definition of Shannon entropy; but by acknowledging that our measure of uncertainty is not absolute but relative, the question of how prior knowledge should be defined complicates Jaynes' simple and seemingly intuitive principle.

While MaxEnt and its derivatives are nowadays well established procedures in practice,

the points mentioned involve many epistemological implications. The resulting controversies surrounding MaxEnt will be addressed in the last section, for now we focus on establishing the necessary concepts: How can we formalize this generalization of MaxEnt?

Shannon entropy was derived under the key assumption that H is a function of $p_1, ..., p_n$ only, with no *explicit* mentioning of a prior distribution. The uniform prior is *implicitly* contained in the Shannon entropy by definition. Hence, in order to generalize our inference method, we need an entropy measure which explicitly incorporates arbitrary prior distributions without violating the established MaxEnt principle.

3.2   Minimum relative entropy

In 1951 Kullback and Leibler introduced a generalized expression of Shannon entropy [56]. Its first appearance can again be traced back to Gibbs [54, p. 136]. Following the discrete notation established in the previous section, it is defined by equation (3.2):

$$D[(p_1, ..., p_n)\|(q_1, ..., q_n)] = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{3.2}$$

D is short for Kullback-Leibler-Divergence, $\{p_1, ..., p_n\}$ and $\{q_1, ..., q_n\}$ denote probability distributions over the same set of n events. In the case of $q_i = 0$, D is only defined if $p_i = 0$ and the convention $0 \log 0 = 0$ applies. For our purposes we identify $\{q_1, ..., q_n\}$ as the *prior* and $\{p_1, ..., p_n\}$ as the *posterior* distribution.

First, with regard to the concepts introduced in the last section, the relationship with Shannon entropy is established. For this we set $\{q_1, ..., q_n\}$ as the uniform distribution:

$$
\begin{aligned}
D[(p_1, ..., p_n)\|(^1/n, ..., ^1/n)] &= \sum_{i=1}^{n} p_i \log \frac{p_i}{^1/n} \\
&= \sum_{i=1}^{n} p_i \log p_i + \sum_{i=1}^{n} p_i \log n \\
&= -H(p_1, ..., p_n) + \log n
\end{aligned} \tag{3.3}
$$

The Kullback-Leibler-Divergence with respect to the uniform prior equates to the negative Shannon entropy plus a constant; D contains H as a special case.

Due to this relationship, D is also called *relative entropy*. It is often illustrated as the information gain when updating from a prior to a posterior, Kullback referred to it as the *information divergence* of $\{p_1, ..., p_n\}$ with respect to $\{q_1, ..., q_n\}$. It serves as a *measure of difference* (or similarity) between probability distributions. However, it is not a distance

metric since it does not satisfy the symmetry condition and triangle inequality.

Expressed in these terms the Shannon entropy $H(p_1, ..., p_n)$ can be seen as a difference measure between $\{p_1, ..., p_n\}$ and the uniform distribution $\{1/n, ..., 1/n\}$. With regard to MaxEnt we can therefore conclude:

The probability distribution $\{p_1^*, ..., p_n^*\}$ that maximizes $H(p_1, ..., p_n)$ also minimizes $D[(p_1, ..., p_n)\|(1/n, ..., 1/n)]$.

D and H share other similarities as well, some noteworthy properties are stated without proof:

1. The Gibbs inequality: $D[(p_1, ..., p_n)\|(q_1, ..., q_n)] \geq 0$.

2. $D[(p_1, ..., p_n)\|(q_1, ..., q_n)] = 0$ if and only if $\{p_1, ..., p_n\} = \{q_1, ..., q_n\}$.

3. $D[(p_1, ..., p_n)\|(q_1, ..., q_n)]$ is convex.

Extending Maximum Shannon entropy to *Minimum relative entropy* was first proposed by Kullback in 1959 [57]. Jaynes adopted its formalism in 1963 to avoid the invariance problems of Shannon entropy in continuous notation [58].

As indicated in the previous section, the general process of minimizing D is straightforward: We infer a posterior distribution $\{p_1, ..., p_n\}$ by constrained minimization of $D[(p_1, ..., p_n)\|(q_1, ..., q_n)]$ where $\{q_1, ..., q_n\}$ is the prior distribution.

As mentioned before, the complicated matter is the choice of the prior. Jaynes argued that "prior probabilities should be those with the maximum entropy consistent with our prior knowledge" [55, p. 185]. By this logic the MaxEnt principle can still define an *objective* prior state of knowledge which allows for the inference of a posterior based on additional constraints. In terms of Bayesian probability theory the prior encodes *degrees of beliefs*; an inherently *subjective* notion of prior knowledge.

This work will not indulge in these debates. MaxEnt solutions occur in various contexts and scenarios with different interpretations, some of which have already been presented above. In some cases the notion of entropy as an uncertainty measure does not even occur.

Whether we consider the prior as subjective, objective or something different, the general mathematical framework is equivalent. Setting the epistemological implications aside for now, the general solution of constrained MaxEnt optimization will be derived in the next section.

3.3 Maximum Entropy distributions

The following derivations have been described by many other authors before. In particular, the depictions here are based on illustrations of Jaynes [53], Bricogne [59, 60] and the thorough write-up of Wu et al. [61].

The random variable X takes the values $x_1, ..., x_n$ with probabilities $p_1, ..., p_n$. As shown in (3.3), D and H are connected via opposite signs. To stay consistent with the MaxEnt terminology and conventions established by Jaynes, instead of minimizing $D[(p_1, ..., p_n)\|(q_1, ..., q_n)]$ we choose to maximize $-D[(p_1, ..., p_n)\|(q_1, ..., q_n)]$, where $\{q_1, ..., q_n\}$ is the fixed prior distribution over the same set of events. Thus, we can write $-D$ as a function of $p_1, ..., p_n$ alone and define $-D := S$.

$$\max_{p_1, ..., p_n} S(p_1, ..., p_n) \tag{3.4}$$

A general constraint of the maximization is given by the normalization condition:

$$F_0 = \sum_{i=1}^{n} p_i = 1 \tag{3.5}$$

Additional m expectation values of arbitrary functions $f_j(X)$ are denoted as equality constraints F_j with $j = 1, ..., m$:

$$F_j = \sum_{i=1}^{n} p_i f_j(x_i) = \langle f_j(X) \rangle \tag{3.6}$$

By introducing $m + 1$ Lagrange multipliers, we can write the constrained optimization problem as a Lagrangian:

$$\mathcal{L}(p_1, ..., p_n, \lambda_0, ..., \lambda_m) = -\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \sum_{j=1}^{m} \lambda_j (\sum_{i=1}^{n} p_i f_j(x_i) - F_j) + \lambda_0 (\sum_{i=1}^{n} p_i - 1) \tag{3.7}$$

For a particular distribution $\{p_1^*, ..., p_n^*\}$ to be a maximizer of $S(p_1, ..., p_n)$ it must be a stationary point of $\mathcal{L}$, so its partial derivatives with respect to p_i $(i = 1, ..., n)$ are zero.

This is of course only a necessary but not a sufficient condition for an absolute minimum. The uniqueness of this minimum will not be shown here, see e.g. [53, sec. 11.6] or [62] for a proof.

The partial derivative with respect to p_i is:

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\log p_i - 1 + \log q_i + \sum_{j=1}^{m} \lambda_j f_j(x_i) + \lambda_0 \tag{3.8}$$

On the stationary point this equates to zero and by rearranging we get:

$$\log p_i = \log q_i - 1 + \lambda_0 + \sum_{j=1}^{m} \lambda_j f_j(x_i) \tag{3.9}$$

Taking the exponential of both sides yields:

$$p_i = q_i \exp(-1 + \lambda_0) \exp(\sum_{j=1}^{m} \lambda_j f_j(x_i))$$

$$= \frac{q_i}{\exp(1 - \lambda_0)} \exp(\sum_{j=1}^{m} \lambda_j f_j(x_i)) \tag{3.10}$$

λ_0 can be written as a function of the other multipliers by employing the normalization condition (using k as an alternative index):

$$1 = \sum_{k=1}^{n} \frac{q_k}{\exp(1 - \lambda_0)} \exp(\sum_{j=1}^{m} \lambda_j f_j(x_k)) \tag{3.11}$$

$$\exp(1 - \lambda_0) = \sum_{k=1}^{n} q_k \exp(\sum_{j=1}^{m} \lambda_j f_j(x_k)) = Z \tag{3.12}$$

Z is called the *partition function*. Substituting it into equation (3.10) we get:

$$p_i = \frac{q_i}{\sum_{k=1}^{n} q_k \exp(\sum_{j=1}^{m} \lambda_j f_j(x_k))} \exp(\sum_{j=1}^{m} \lambda_j f_j(x_i)) \tag{3.13}$$

Equation (3.13) is the general form of a MaxEnt distribution or rather a "minimum relative entropy distribution". The main achievement of this derivation can already be seen: By expressing p_i as a function of $\lambda_1, ..., \lambda_m$, the number of dimensions of the original optimization can be drastically decreased; in fact only the number of constraints influences the dimensionality of the problem, not the number of possible events.

The resemblance of (3.13) with Gibbs' canonical ensemble is part of Jaynes' orginal narrative. In more general terms the MaxEnt distribution is part of the exponential family of probability distributions. Other noteworthy distributions in this family include the binomial, multinomial, exponential, chi-squared and normal distribution. In fact some members of the exponential family are MaxEnt distribution under specific constraints, the normal distribution is probably the most prominent example, see e.g. [51].

At the stationary point the partial derivative of the Lagrangian (3.7) with respect to λ_j equates by definition to the constraint F_j:

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = 0 = \sum_{i=1}^{n} p_i f_j(x_i) - F_j \tag{3.14}$$

Substituting p_i with the MaxEnt distribution in (3.13) we obtain the constraints F_j as functions of $\lambda_1, ..., \lambda_m$:

$$F_j = \sum_{i=1}^{n} f_j(x_i) \frac{q_i}{Z} \exp(\sum_{j=1}^{m} \lambda_j f_j(x_i)) \tag{3.15}$$

Equations (3.12),(3.13) and (3.15) are the so-called *entropy equations*. These m equations with m unknowns could already be used to determine the Lagrange multipliers $\lambda_1, ..., \lambda_m$ and $p_i(\lambda_1, ..., \lambda_m)$. However, the entropy equations exhibit additional properties worth investigating.

First we examine the corresponding maximized $S_{max} = [-\sum_{i=1}^{n} p_i \log p_i/q_i]_{max}$. As seen in (3.13), the probabilities p_i are a function of $\lambda_1, ..., \lambda_m$ alone, so S_{max} is a function of the constraints F_j only and not p_i. By substituting p_i with the MaxEnt distribution (3.13) and employing the normalization condition (3.5) we get:

$$\begin{aligned}
S_{max}(F_1, ..., F_m) &= -\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \\
&= -\sum_{i=1}^{n} p_i \log \frac{\frac{q_i}{Z} \exp(\sum_{j=1}^{m} \lambda_j f_j(x_i))}{q_i} \\
&= \log Z \sum_{i=1}^{n} p_i - \sum_{j=1}^{m} \lambda_j \sum_{i=1}^{n} p_i f_j(x_i) \\
&= \log Z - \sum_{j=1}^{m} \lambda_j F_j \tag{3.16}
\end{aligned}$$

In an equal manner the MaxEnt distribution (3.13) can be used to simplify the original Lagrangian stated in (3.7):

$$\mathcal{L}(p_1, ..., p_n, \lambda_0, ..., \lambda_m) = -\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \sum_{j=1}^{m} \lambda_j (\sum_{i=1}^{n} p_i f_j(x_i) - F_j) + \lambda_0(\sum_{i=1}^{n} p_i - 1)$$

Using the normalization condition (3.5) we can simplify and rearrange:

$$\mathcal{L}(p_1, ..., p_n, \lambda_0, ..., \lambda_m) = -\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \sum_{j=1}^{m} \lambda_j \sum_{i=1}^{n} p_i f_j(x_i) - \sum_{j=1}^{m} \lambda_j F_j$$

In analogy to (3.16) we substitute p_i with the MaxEnt distribution (3.13) which yields:

$$\begin{aligned}
\mathcal{L}(\lambda_1, ..., \lambda_m) &= \log Z \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} p_i \sum_{j=1}^{m} \lambda_j f_j(x_i) + \sum_{j=1}^{m} \lambda_j \sum_{i=1}^{n} p_i f_j(x_i) - \sum_{j=1}^{m} \lambda_j F_j \\
&= \log Z(\lambda_1, ..., \lambda_m) - \sum_{j=1}^{m} \lambda_j F_j \tag{3.17}
\end{aligned}$$

This is the same expression we derived for S_{max} in (3.16). This is no coincidence. The

connection becomes clearer by again investigating the partial derivative of $\mathcal{L}(\lambda_1, ..., \lambda_m)$:

$$\frac{\partial \mathcal{L}}{\partial \lambda_j} = 0 = \frac{\partial \log Z}{\partial \lambda_j} - F_j \tag{3.18}$$

We can calculate the derivative of $\log Z$ explicitly and compare with (3.15) to verify the equality above:

$$\begin{aligned}
\frac{\partial \log Z}{\partial \lambda_j} &= \frac{1}{Z}\frac{\partial Z}{\partial \lambda_j} \\
&= \frac{1}{Z}\sum_{i=1}^{n} f_j(x_i)q_i \exp(\sum_{j=1}^{m}\lambda_j f_j(x_i)) \\
&= F_j
\end{aligned} \tag{3.19}$$

This is an alternative way of writing the entropy equation in (3.15). Substituting the connection found in (3.19) into the expression for S_{max} in (3.16) we get:

$$S_{max}(F_1, ..., F_m) = \log Z(\lambda_1, ..., \lambda_m) - \sum_{j=1}^{m}\lambda_j \frac{\partial \log Z(\lambda_1, ..., \lambda_m)}{\partial \lambda_j} \tag{3.20}$$

S_{max} as a function of $F_1, ..., F_m$ equates to a function dependent only on $\lambda_1, ..., \lambda_m$. The partial derivatives establishing the connection between $S_{max}(F_1, ..., F_m)$ and $\log Z(\lambda_1, ..., \lambda_m)$ indicate an important *duality*: S_{max} and $\log Z$ are Legendre transforms of each other, either function gives a full description of the system.

The Lagrange multipliers $\lambda_1, ..., \lambda_m$ can therefore be expressed in an equal manner using the derivative of S_{max} with respect to F_j. Under consideration of (3.19) the partial derivative of (3.16) is:

$$\begin{aligned}
\frac{\partial S_{max}(F_1, ..., F_m)}{\partial F_j} &= \frac{\partial \log Z}{\partial F_j} - F_j\frac{\partial \lambda_j}{\partial F_j} - \lambda_j \\
&= \frac{\partial \log Z}{\partial \lambda_j}\frac{\partial \lambda_j}{\partial F_j} - F_j\frac{\partial \lambda_j}{\partial F_j} - \lambda_j \\
&= -\lambda_j
\end{aligned} \tag{3.21}$$

By substituting (3.21) into (3.16) and rearranging we obtain $\log Z(\lambda_1, ..., \lambda_m)$ as a function of $S_{max}(F_1, ..., F_m)$, the analogue expression to (3.20):

$$\log Z(\lambda_1, ..., \lambda_m) = S_{max}(F_1, ..., F_m) - \sum_{j=1}^{m}F_j\frac{\partial S_{max}(F_1, ..., F_m)}{\partial F_j} \tag{3.22}$$

This *duality* allows for a very convenient alternative to the constrained maximization of $S(p_1, ..., p_n)$. The potential represented by $\mathcal{L}(\lambda_1, ..., \lambda_m)$ in (3.17) can be shown to be strictly convex (e.g. [61]). If a stationary point exists, it is the unique absolute mini-

mum of $\mathcal{L}(\lambda_1, ..., \lambda_m)$. For a given a set of constraints $\{F_1^*, ..., F_m^*\}$ the unique minimizing set $\{\lambda_1^*, ..., \lambda_m^*\}$ equally solves the entropy equations in (3.15) and (3.19). In fact: $S_{max}(\lambda_1^*, ..., \lambda_m^*) = \mathcal{L}(\lambda_1^*, ..., \lambda_m^*)$, as the direct comparison of (3.16) and (3.17) shows.

The above transformation of a constrained into an unconstrained optimization problem follows a common procedure in convex optimization theory. By expressing the *primal* variables p_i as a function of the *dual* variables λ_j and substituting back into the *Primal* Lagrangian, the corresponding *Dual* is obtained. Whether the Dual represents a solution or only a higher/lower bound to the original problem depends above all on the convexity properties of the investigated function.

In the case of Minimum Kullback-Leibler-Divergence and Maximum Entropy respectively the Lagrange Dual was first described by Charnes and Cooper in 1975 [63] and extended in 1977-78 [64, 65]. More often cited is the independent report of Alhassid et al. in 1978 [62] who derived the connection from the Gibbs inequality alone without any notion of Lagrange or Legendre duality.

To summarize: The negative Kullback-Leibler-Divergence was introduced as a suitable measure for constrained optimization. Since we are investigating probability distributions, the normalization condition (zeroth moment) was identified as a general constraint. Additional equality constraints were introduced in the form of expectation values of arbitrary functions. This constrained maximization problem was then expressed as a Lagrangian. Holding the Langrange multipliers, i.e. the *dual variables*, fixed and examining the partial derivative with respect to the probability distribution, i.e. the *primal variables*, a general expression for the distribution was obtained: The Maximum Entropy distribution was found to be a function of the Lagrange multipliers alone, the dimensionality of the optimization problem does not depend on the number of possible events. As a consequence the maximized negative Kullback-Leibler-Divergence was identified as a function of the constraints only. Substituting the general Maximum Entropy distribution into both the negative Kullback-Leibler-divergence and the Lagrangian revealed an instance of *duality*: Following a general scheme in convex optimization, the problem of solving a system of linear equations posed by the *Primal* can be transformed into an unconstrained minimization of its corresponding *Dual*.

The minimization of the Dual is straightforward and can readily be implemented using Newton's method or other root-finding algorithms.

3.4   Why Maximum Entropy?

This work attempts to find a solution to an underdetermined inverse problem: Inferring a probability distribution from insufficient information. Naturally this raises the question of how we justify the selection of one out of many possible solutions. Why Maximum Entropy?

It turns out this cannot be answered unambiguously. In order to contextualize the specific use of MaxEnt in this work, a brief overview of MaxEnt rationales is given in this section.

While entropy functionals turned out to be very robust both in a mathematical and historical sense, it still depends on what justifications are considered plausible. In the first application of MaxEnt in 1957 Jaynes justified his choice of Shannon entropy based on its axioms, its interpretation as an uncertainty measure and analogies to statistical mechanics.

We followed this well established and somewhat intuitive narrative to introduce the Max-Ent method, but of course this reasoning is debatable; the Gibbs paradox or Maxwell's demon might come to mind.

Consequently, MaxEnt rationales and their extensions have become more and more sophisticated. Indicative of this debate's persistence, in 1991 Csiszár still addresses "the question of what selection rules are 'good'" [66, p. 2032] and concludes: "Unfortunately, it is hard to give a mathematical meaning to this question. It does not seem possible to define a general criterion by which the goodness of selection rules could be compared." [66, p. 2032-2033]

Csiszár chooses to "adopt an axiomatic approach and consider those selection rules as 'good' that lead to a logically consistent method of inference, in the sense of satisfying some natural postulates" [66, p. 2033]. Rather than basing the method on the axioms of the information measure, Csiszár opts for an axiomatization of the method itself. This follows an approach first devised by Shore and Johnson in 1980 [67], notable later alternatives include works by Skilling [68], Paris and Vencovská [69] and more recently Caticha and Giffin, who derived the compatibility of the "Maximum Relative Entropy" method and Bayesian Updating as a consequence [70–73]. Knuth and Skilling's "Foundations of Inference" [74] has an even wider scope, nonetheless it offers an axiomatic characterization MaxEnt and entropy measures in general. Both Caticha and Skilling comment on the unique role of the Kullback-Leibler-Divergence in this context, see e.g. [71, 75–77].

What unites these approaches is that entropy and its maximization are derived from requirements of consistent reasoning. As Caticha puts it: "*Entropy needs no interpretation. We do not need to know what 'entropy' means, we only need to know how to use it.*" [70,

p. 77]. While these approaches do not rely on a particular concept of entropy, their respective axioms can and have been objected [78–80].

Other justifications rely on relationships with other established principles in statistics. As Jaynes puts it, in a frequentistic view "the probability distribution which maximizes the entropy is numerically identical with the frequency distribution which can be realized in the greatest number of ways" [81, p. 8]; an observation dating back to Boltzmann and generalized in Sanov's theorem [82, 83]. Closely related is the finding of Van Campenhout and Cover [84] who showed the asymptotic equality between conditional and MaxEnt distributions. This was later extended by Csiszár in his *conditional limit theorem* [85]: Csiszár argues that the asymptotic agreement between MaxEnt and Bayesian Updating allows for a mutual justification of both methods [86]. Caticha points out that large ensemble scenarios might not be applicable to the general case, but the law of large numbers could still be used as a consistency argument [75].

However, the compatibility of Bayesian Updating and MaxEnt has already been indicated in 1980 by Williams [87] without referring to asymptotic behaviors. In 1973 Akaike linked Fisher's maximum likelihood method [88] with information theory and MaxEnt [89], which he related to Bayesian statistics in later works [90, 91]. In their "Alternative approach to maximum-entropy inference" [92] Tikochinsky et al. relate the MaxEnt distribution to Fisher's concept of sufficient statistics [88] and other more physical considerations without invoking any particular concept of entropy or information. A game theoretic viewpoint was given by Topsøe [93].

Recommendable background literature includes Caticha's lecture notes on Bayesian probability theory, entropy and inference [75], "Probability Theory: The Logic of Science" by Jaynes and Bretthorst [53], Csiszárs overview on MaxEnt [94] and entropy axiomatizations [50], Uffink's comment on the MaxEnt constraint rule and its relation to Bayesian updating [95] as well as the short write-up on entropy by Georgii [96].

To simplify the arguably complex epistemological debates, we now consider MaxEnt in the context of this work. We asked: To what extent can arbitrary probability distributions be reproduced given only a set of probability-weighted average function values?

We want to assess the cumulative information content provided by a set of ensemble averaged experimental observables. The problem is inherently underdetermined, but to what extent can we still draw conclusions about the underlying system? What answers can MaxEnt provide?

Jaynes argued as follows: "The principle of maximum entropy is not an oracle telling which predictions *must* be right; it is a rule for inductive reasoning that tells us which predictions *are most strongly indicated by our present information.*" [53, p. 370]

MaxEnt is used as a tool to assess information content. As Caticha puts it: "It is not just what you happen to know; you have to know the right thing. The constraints that should be imposed are those that codify the information that is relevant to the problem under consideration." [75, p. 88] If we don't know "the right thing", MaxEnt can guide us towards the design and implementation of experiments providing the missing information. This is the interpretation of MaxEnt we follow in this work. It is also the key difference to previous MaxEnt studies concerned with IDPs [18–20], which did not survey the amount nor the specific combination of experiments. Missing information was instead introduced by constructing a singular, highly sophisticated prior; its overall influence was not assessed.

However, the prior is arguably the most controversial parameter in the MaxEnt framework. Whatever we choose to "best" represent our prior knowledge, it is inherently subjective. Moreover, the use of only one prior makes it difficult to asses its dominance in the resulting predictions. This work opts for a different approach: The goal is to design and validate an experimental procedure. While the use of different priors arguably highlights the underdetermined nature of the underlying problem and the relativity of the resulting predictions, it allows for a better evaluation of the information content supplied by the experiments. The more the possibility space is restrained by experimental observables, the less pronounced the influence of the prior should become: If a prediction is dominated by the prior, it indicates a lack of experimentally supplied information. After all, experiments should be able to contradict possibly unjustified assumptions.

As indicated above, other interpretations of MaxEnt are equally justified; the mathematical framework is ultimately the same: To assess the validity of a specific combination of experiments the distribution with the lowest relative entropy with respect to a prior is chosen. While subjectivity cannot be eliminated, its culmination in form of the prior is clearly defined. The corresponding formalisms are well-established, simple, low-dimensional and can be computed highly efficiently.

Chapter 4

Methods and Implementation

4.1   Geometrical modulation of CCR rates

Not all CCR experiments and their angular dependencies investigated in this work have been described in full detail. They were either not thoroughly parametrized in previous studies or have not been devised yet. To investigate still unexplored CCR rates and to stay internally consistent, the geometrical dependencies were generated from scratch with a common protein geometry. Since there is no documented geometry to reuse, a backbone was constructed with Avogadro [97]. Bond lengths and angles were adapted from Momany et al. [98], slight angle deviations necessary to construct the backbone were kept to a minimum. The bond lengths can be found in Table 4.1. The protein with $\phi = \psi = \omega = -180°$ is depicted in Figure 4.1, its corresponding x,y,z-coordinates are documented in Table 4.2.
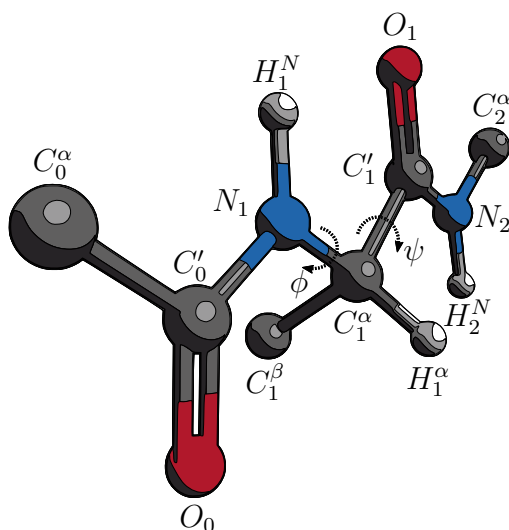

Figure 4.1: 3D representation of the model protein backbone with $\phi = \psi = \omega = -180°$ used to generate the geometrical modulations of arbitrary CCR rates.

Table 4.1: Bond lengths according to [98] used to construct the model protein backbone (Figure 4.1).

bond type	length [Å]
N-H	1.00
C-H	1.00
C-O	1.23
C-N	1.45
C-C	1.53

Table 4.2: The model protein backbone (Figure 4.1) in x,y,z-coordinates with $\phi = \psi = \omega = -180°$.

atom	x [Å]	y [Å]	z [Å]
C_0^α	3.65662	0.00927	-0.98006
C_0'	2.13500	0.01123	-1.13998
O_0	1.62493	0.02254	-2.25917
N_1	1.45000	0.00000	0.00000
H_1^N	1.87262	-0.08271	0.90253
C_1^α	0.00000	0.00000	0.00000
H_1^α	-0.33381	0.81771	-0.46895
C_1^β	-0.42991	-1.25812	-0.75711
C_1'	-0.52182	-0.01279	1.43821
O_1	0.26150	-0.02925	2.38639
N_2	-1.84661	-0.00472	1.55557
H_2^N	-2.47453	0.06788	0.78069
C_2^α	-2.48090	-0.01517	2.85943

With the protein backbone in Figure 4.1 as a template, different CCR rates were generated using Python scripts.

First the principal components of the carbonyl CSA were set in accordance with Teng et al. [99]: $\vec{C'}_z$ ($\sigma_{zz} = 90$ ppm) was defined as the cross product of the C'-O and the C'-C^α bond unit vectors, $\vec{C'}_x$ ($\sigma_{xx} = 244$ ppm) and $\vec{C'}_y$ ($\sigma_{yy} = 178$ ppm) as clockwise rotations of the C'-O bond unit vector around $\vec{C'}_z$ by 82° and -8° respectively, approximating the O-C'-N angle with 120°.

The rotations were implemented using the Euler-Rodrigues formula, describing a counterclockwise rotation by an angle θ around a unit vector (e_x, e_y, e_z) as a matrix R:

$$R := \begin{bmatrix} a^2 + b^2 - c^2 - d^2 & 2(bc + ad) & 2(bd - ac) \\ 2(bc - ad) & a^2 + c^2 - b^2 - d^2 & 2(cd + ab) \\ 2(bd + ac) & 2(cd - ab) & a^2 + d^2 - b^2 - c^2 \end{bmatrix} \tag{4.1}$$

where

$$a = \cos\frac{\theta}{2}$$

$$b = -\sin\frac{\theta}{2}e_x$$

$$c = -\sin\frac{\theta}{2}e_y$$

$$d = -\sin\frac{\theta}{2}e_z$$

Rotations around the C_1^α-N_1 and the C_1^α-C_1' bonds were performed in $1°$ steps, modifying the coordinates corresponding to changes in ϕ and ψ and calculating CCR rates for 360^2 angle pairs: $\{(\phi, \psi) \in \mathbb{Z}^2 \mid -180° \leq \phi, \psi < 180°\}$. The geometry-dependent components were calculated according to equations (4.2), (4.3) and (4.4):

$$\Gamma_{AB,CD}^{DP,DP} = \frac{1}{\left\|\overrightarrow{AB}\right\|^3 \left\|\overrightarrow{CD}\right\|^3} Y_2(\theta_{\overrightarrow{AB},\overrightarrow{CD}}) \tag{4.2}$$

$$\Gamma_{C',AB}^{CSA,DP} = \frac{1}{\left\|\overrightarrow{AB}\right\|^3} \sum_{i \in Q} \sigma_{ii} Y_2(\theta_{\vec{C'}_i, \overrightarrow{AB}}) \tag{4.3}$$

$$\Gamma_{C_1',C_2'}^{CSA,CSA} = \sum_{i \in Q} \sum_{j \in Q} \sigma_{ii}\sigma_{jj} Y_2(\theta_{\vec{C'}_{1i}, \vec{C'}_{2j}}) \tag{4.4}$$

with

- $Q = \{x, y, z\}$ containing the Cartesian coordinate indices,

- $Y_2(\theta_{\vec{a},\vec{b}}) = {}^1/_2(3\cos^2\theta_{\vec{a},\vec{b}} - 1)$ as the second order Legendre polynomial,

- $\cos\theta_{\vec{a},\vec{b}} = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\| \, \|\vec{b}\|}$ defining the projection angle and

- A, B, C and D denoting arbitrary nuclei subject to dipolar (DP) coupling.

Since only geometrical modulations will be surveyed in this work, the spectral density functions were not factored in. Corresponding with the $[-0.5, 1.0]$ interval of Y_2, the functions were instead normalized to their biggest positive value to ensure a stable numerical minimization later on; consequently, the carbonyl CSA components enter the evaluation not in absolute but relative terms.

4.2 Probability distributions generation

Different kinds of distributions in (ϕ,ψ)-space were employed in this work. The sample distributions used to generate constraints and to assess the quality of the MaxEnt solutions

are based on elliptical two-dimensional Gaussians:

$$G(\phi, \psi) = A \exp(-(a(\phi - \phi_0)^2 + 2b(\phi - \phi_0)(\psi - \psi_0) + c(\psi - \psi_0)^2)) \qquad (4.5)$$

where

$$a = \frac{\sin^2 \theta}{2\sigma_\psi^2} + \frac{\cos^2 \theta}{2\sigma_\phi^2}$$

$$b = \frac{\sin 2\theta}{4\sigma_\psi^2} - \frac{\sin 2\theta}{4\sigma_\phi^2}$$

$$c = \frac{\cos^2 \theta}{2\sigma_\psi^2} + \frac{\sin^2 \theta}{2\sigma_\phi^2}$$

with

- the center at (ϕ_0, ψ_0),

- the standard deviations σ_ϕ and σ_ψ,

- the amplitude A and

- a rotation by θ around the center, counterclockwise in the Ramachandran plot.

Different Gaussians were used to represent regions of interest. The study of Hollingsworth and Karplus [100] served as a useful reference in constructing the list of motives: right- and left-handed α-helical regions α_r and α_l, the canonical right-handed α_{rc} helix, a polyproline-II region $PP\text{-}II$ as well as β and extended β_{ext} regions at lower ψ values. Like the geometrical functions, the Gaussians were generated on 360 by 360 grids, the parameters are documented in Table 4.3. A graphical sketch of the motives can be found in the following chapter (Figure 5.1). To construct a sample distribution the Gaussians were generated with the desired amplitudes, summed up and normalized.

Table 4.3: Motive parameters employed for the generation of arbitrary sample distributions according to equation (4.5).

motive	ϕ_0 [°]	ψ_0 [°]	σ_ϕ [°]	σ_ψ [°]	θ [°]
α_l	65	35	10	25	30
α_r	-85	-20	13	40	40
α_{rc}	-63	-43	5	7	45
β	-130	145	17	32	40
β_{ext}	-130	75	10	18	0
$PP\text{-}II$	-75	155	13	22	40

Besides the sample Gaussians, two priors were used for the subsequent analyses, which are depicted in the next chapter (Figure 5.2).

The *uniform prior* simply assigns equal probabilities to all angle pairs: $p(\phi, \psi) = 1/360^2$, $\phi, \psi \in \mathbb{Z} : -180° \leq \phi, \psi < 180°$.

The *knowledge-based prior* is adapted from the library of Mantsyzov et al. [20]. It uses high resolution X-ray structures from the PDB ($\leq$ 2.0 Å, R factor lower than 23 %, maximum pairwise sequence identity of 50 %) from which all fragments were extracted that are at least three residues long and feature no backbone-backbone H-bonding (cut-off energy < -0.7 kcal/mol). While derived from structured proteins, this filter is an attempt to construct a prior representative for IDPs by capturing the "more unstructured" parts of the PDB.

The Mantsyzov library was downloaded from the MERA webpage [101]. Ill-defined angle-pairs of terminal residues were excluded as well as glycine and proline residues to reflect the experimental limitations of this work's setup, reducing the original 195859 entries to 147091. The angles were rounded to integers, transferred onto a 360 by 360 grid and normalized, yielding the *knowledge-based prior* employed in the following analyses.

4.3 Entropy optimization routine

The MaxEnt solutions were calculated in accordance with Section 3.3. The ensemble averages are given as

$$\langle \Gamma_k \rangle = \sum_\phi \sum_\psi \Gamma_k(\phi, \psi) p_s(\phi, \psi) \tag{4.6}$$

with

- $\phi, \psi \in \mathbb{Z} : -180° \leq \phi, \psi < 180°$,

- $p_s(\phi, \psi)$ as a given sample distribution and

- $\Gamma_k(\phi, \psi)$ as the k-th CCR rate, $k = 1, ..., m$.

The Lagrange potential is

$$\mathcal{L}(\lambda_1, ..., \lambda_m) = \log Z(\lambda_1, ..., \lambda_m) - \sum_{k=1}^m \lambda_k \langle \Gamma_k \rangle \tag{4.7}$$

where

$$Z = \sum_\phi \sum_\psi z(\phi, \psi) \tag{4.8}$$

$$z(\phi, \psi) = q(\phi, \psi) \exp(\sum_{k=1}^m \lambda_k \Gamma_k(\phi, \psi)) \tag{4.9}$$

with

- $q(\phi, \psi)$ as the prior distribution and

- λ_k as the k-th Lagrange multiplier, $k = 1, ..., m$.

For the Newton minimization of (4.7) the analytical first and second derivatives are used. The first derivative with respect to λ_i ($i = 1, ..., m$) is:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \lambda_i} &= \frac{\partial \log Z}{\partial \lambda_i} - \langle \Gamma_i \rangle \\
&= \frac{1}{Z} \frac{\partial Z}{\partial \lambda_i} - \langle \Gamma_i \rangle \\
&= \frac{1}{Z} \sum_\phi \sum_\psi \Gamma_i(\phi, \psi) z(\phi, \psi) - \langle \Gamma_i \rangle
\end{aligned} \tag{4.10}$$

Using product and chain rule, the second derivative with respect to λ_j ($j = 1, ..., m$) can be calculated as:

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \lambda_i \partial \lambda_j} &= \frac{\partial^2 \log Z}{\partial \lambda_i \partial \lambda_j} \\
&= \frac{1}{Z} \frac{\partial^2 Z}{\partial \lambda_i \partial \lambda_j} - \frac{1}{Z^2} \frac{\partial Z}{\partial \lambda_i} \frac{\partial Z}{\partial \lambda_j} \\
&= \frac{1}{Z} \sum_\phi \sum_\psi \Gamma_i(\phi, \psi) \Gamma_j(\phi, \psi) z(\phi, \psi) \\
&\quad - \frac{1}{Z^2} \sum_\phi \sum_\psi \Gamma_i(\phi, \psi) z(\phi, \psi) \sum_\phi \sum_\psi \Gamma_j(\phi, \psi) z(\phi, \psi)
\end{aligned} \tag{4.11}$$

It can easily be seen that the order in which the partial derivatives are taken is interchangeable, thus, the corresponding Hessian matrix is symmetric. With the Jacobian and the Hessian constructed from (4.10) and (4.11), the Lagrange potential $\mathcal{L}(\lambda_1, ..., \lambda_m)$ (4.7) is minimized using the SciPy [102] implementation of the Newton method.

The minimizing Lagrange multipliers $\lambda_1^*, ..., \lambda_m^*$ then define the MaxEnt distribution $p_{ME}^*(\phi, \psi)$ as:

$$p_{ME}^*(\phi, \psi) = \frac{q(\phi, \psi)}{\sum_\phi \sum_\psi q(\phi, \psi) \exp(\sum_{k=1}^m \lambda_k^* \Gamma_k(\phi, \psi))} \exp(\sum_{k=1}^m \lambda_k^* \Gamma_k(\phi, \psi)) \tag{4.12}$$

Chapter 5

Results

This work aims to explore experimental possibilities in a synthetic setting. MaxEnt was introduced as a general method with broad applicability and many degrees of freedom: We can vary the amount and combination of function constraints and use any number of prior and sample distributions. A systematic exploration of all possibilities is neither realistic nor desirable, the particular choices in the subsequent evaluations follow pragmatic considerations:

What sample probability distributions do we want to reproduce? Clearly, the narrow and localized distributions of globular proteins are not of primary interest. While steric exclusion is an evident restraint, it is not clear what probabilities we should assign within the allowed regions. The sample distributions to be reproduced can be anything considered interesting to characterize, which is of course subjective. In this work a set of differently weighted Gaussians in (ϕ,ψ)-space are used covering the regions of interest as illustrated in Figure 5.1. Details regarding their generation are described in Section 4.2. Of course other distributions with different features or alternative origins like MD simulations would be equally applicable. However, only the corresponding set of ensemble averaged constraints enter the subsequent mathematical treatment, the features of the sample distributions are not conserved.

The prior, on the other hand, is not just one out of many distributions we might want to reproduce. It is a *singular* distribution which should somehow represent the *plurality* of possible distributions. It strongly affects the predictions of the MaxEnt formalism, it is the standard by which we asses the predictive quality. The prior assumptions act as an additional source of information; any prediction without sufficient experimental constraints will be dominated by the prior. As will be illustrated later, the *uniform prior* offers particular advantages in this regard because it lacks any variation in density. Thus, any deviation from uniformity directly results from the introduced constraints; every fea-
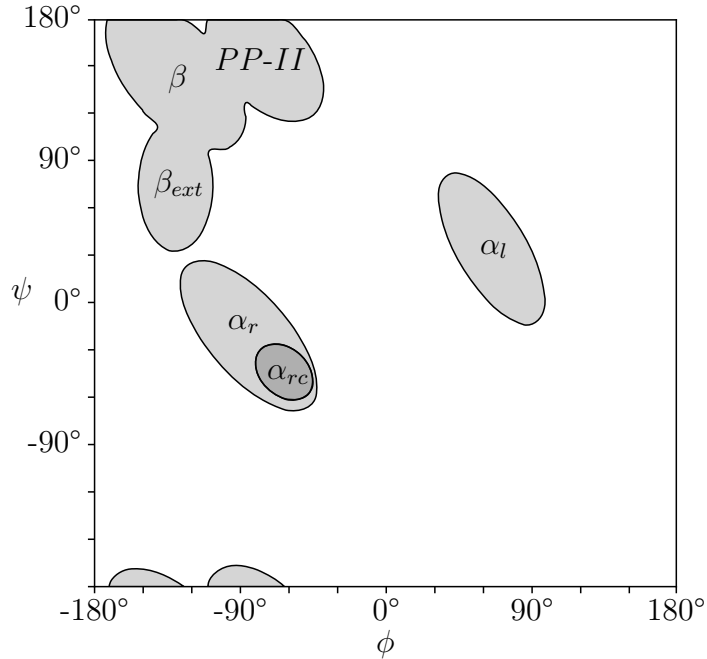
Figure 5.1: Illustration of the motives used to assess the MaxEnt formalism: right-handed α_r, left-handed α_l, canonical right-handed α_{rc}, polyproline-II $PP\text{-}II$, β and extended β_{ext}. Details are given in Section 4.2.

ture in the predictions represents the cumulative information content of the experiments. In a way the uniform prior is the most "ignorant" or "noncommittal" extreme to which other priors can be compared to.

What other priors could be used as comparisons? Again, both simple and sophisticated possibilities exist to derive what might be considered the most accurate representation of prior knowledge, even a MaxEnt prior could be used [18, 103]; but be it a prior either derived from first principles or a database, one could always debate whether the underlying assumptions are too specific, general or unrealistic in some way. The prior debate cannot be resolved here and this work does not attempt to do so. We are not concerned with inferring "true" distributions: An underdetermined problem cannot have a unambiguous solution. The ambiguities cannot be resolved, they can only be restrained. The MaxEnt distribution is first and foremost interpreted as a representation of information, which is always relative to its respective prior. Ideally a robust experimental setting should introduce enough restraint to yield comparable results even for different priors. Since we are attempting to asses the restraints introduced by CCR experiments and their potential in IDP characterization, prior assumptions will be treated as variables and evaluated as an additional degree of freedom.

The PDB-derived prior of Mantsyzov et al. [20] offers a convenient way to relate this work to previous studies. Again, whether it is the "best" available prior is not considered important at this point, it is merely a plausible option. Here, an adapted version excluding
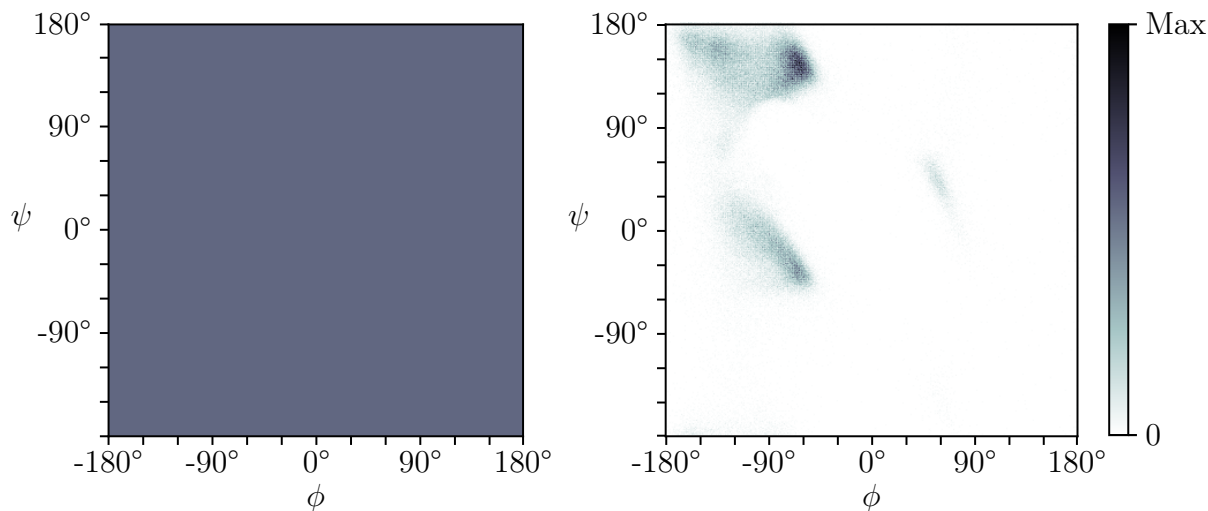
Figure 5.2: The two priors employed in the MaxEnt calculations: The *uniform* prior on the left-hand side, the *knowledge-based* prior on the right hand side. Details are given in Section 4.2.

glycine and proline residues will represent the *knowledge-based prior*. It is depicted in Figure 5.2 together with the uniform prior. Details are covered in Section 4.2.

With the prior and sample distributions defined, the essential question of this work can be addressed: What experimental constraints should be used to characterize the sample distributions? The protein backbone allows for a multitude of possible combinations of dipolar and/or CSA interactions, many of which have already been described in the literature. Employing already established experiments appears to be a sensible choice. Whether these suffice to characterize the chosen sample distributions remains to be seen of course: The MaxEnt formalism provides the guideline for possible deviations from established protocols.

5.1   Experimental procedure assessment

This section showcases the evolution of MaxEnt solutions with increasing amount of constraints. A simple sample distribution with equal ratios of α_l, α_r, β and $PP\text{-}II$ (Figure 5.3) is chosen for this purpose.

This procedure corresponds to the design of an experimental protocol similar to Kloiber et al. [25]: Possible interactions along the protein backbone are successively chosen to further improve the resulting prediction.
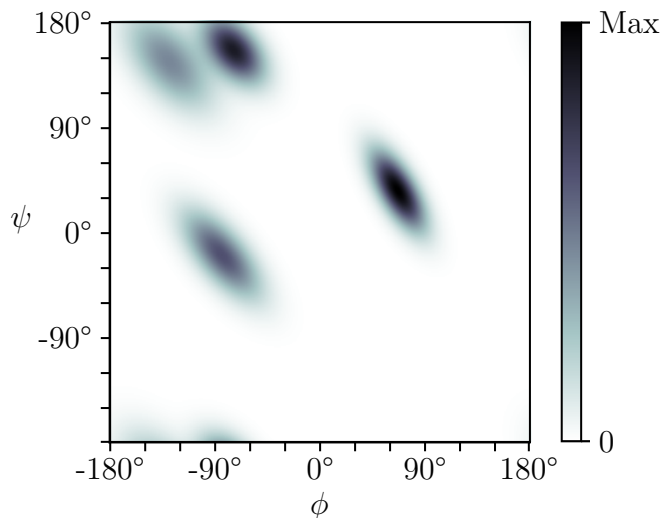
Figure 5.3: The sample distribution used to assess the quality of the MaxEnt predictions. It consists of equal ratios of α_l, α_r, β and PP-II, see Figure 5.1 and Section 4.2 for details.

5.1.1 MaxEnt solutions of existing experiments

Following the sequence of Kloiber et al. [25], we first examine four basic CCR experiments which relate the H^α-C^α bond with either a carbonyl CSA tensor or an H^N-N dipolar vector to probe ϕ or ψ (assuming $\omega = 180°$). Figure 5.4 shows the angular dependencies of $\Gamma_{H_i^\alpha C_i^\alpha, C_i'}(\psi)$ [38], $\Gamma_{C_{i-1}', H_i^\alpha C_i^\alpha}(\phi)$ [39], $\Gamma_{H_i^N N_i, H_i^\alpha C_i^\alpha}(\phi)$ [37] and $\Gamma_{H_i^\alpha C_i^\alpha, H_{i+1}^N N_{i+1}}(\psi)$ [23], details regarding their generation are covered in Section 4.1. Using these functions to generate constraints, their respective MaxEnt solutions can be calculated.

Figure 5.5 illustrates this process: The rows are labeled with I for "input", U for the "uniform prior" and K for the "knowledge-based prior". Referring to the plots by their row and column, $I0$ shows the sample distribution (Figure 5.3). The arrows point at particular MaxEnt solutions in rows K and U. $K0$ and $U0$ are the solutions with no constraints applied, which equates to the respective priors: $U0$ is the uniform, $K0$ the knowledge-based prior (Figure 5.2). Starting with $I1$, each $+$ sign along row I introduces a new constraint function from Figure 5.4. The corresponding expectation values calculated from $I0$ enter the MaxEnt solutions. Thus, the column number indicates the number and type of constraints entering the predictions.

A first important observation is the distinct influence of the prior on the predictions. This is a general feature of MaxEnt solutions: Since we are minimizing a distance measure, the prediction is in a way the distribution "closest" to the initial guess. If there is not much additional information supplied, the solution mostly reflects its prior. The uniform prior best illustrates this property: $U1$ adopts the shape of the constraint function $I1$. A single Lagrange parameter exponentially weighs $I1$, the assigned probabilities are highest or lowest at the corresponding extrema of $I1$. The result is highly unintuitive: The strong
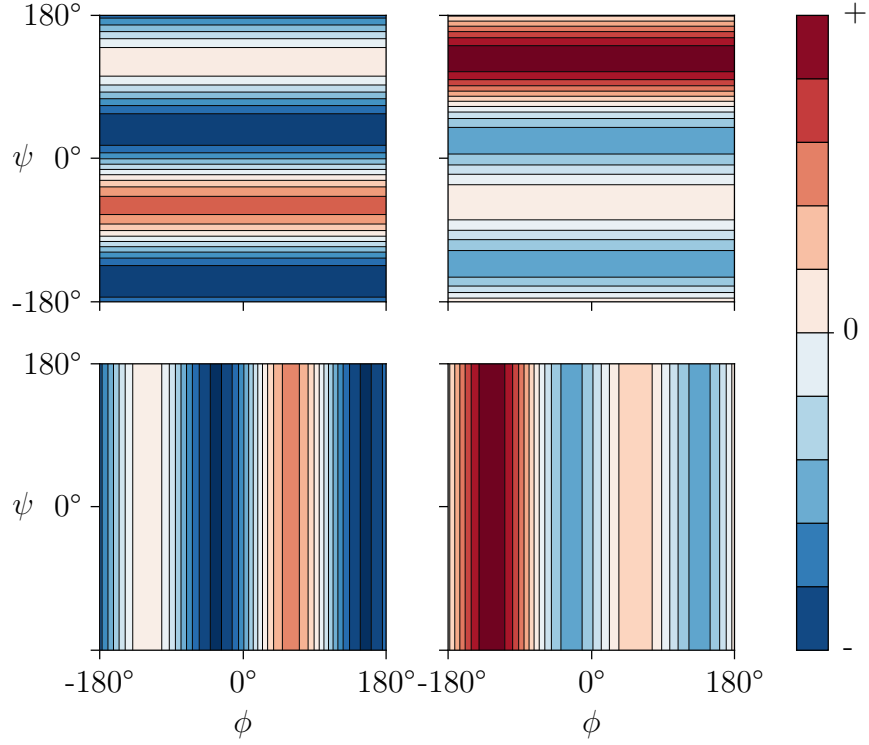
Figure 5.4: Geometrical modulations of different CCR rates according to Section 4.1. Upper left-hand side: $\Gamma_{H_i^\alpha C_i^\alpha, C_i'}(\psi)$. Upper right-hand side: $\Gamma_{H_i^\alpha C_i^\alpha, H_{i+1}^N N_{i+1}}(\psi)$. Lower left-hand side: $\Gamma_{C_{i-1}', H_i^\alpha C_i^\alpha}(\phi)$. Lower right-hand side: $\Gamma_{H_i^N N_i, H_i^\alpha C_i^\alpha}(\phi)$.
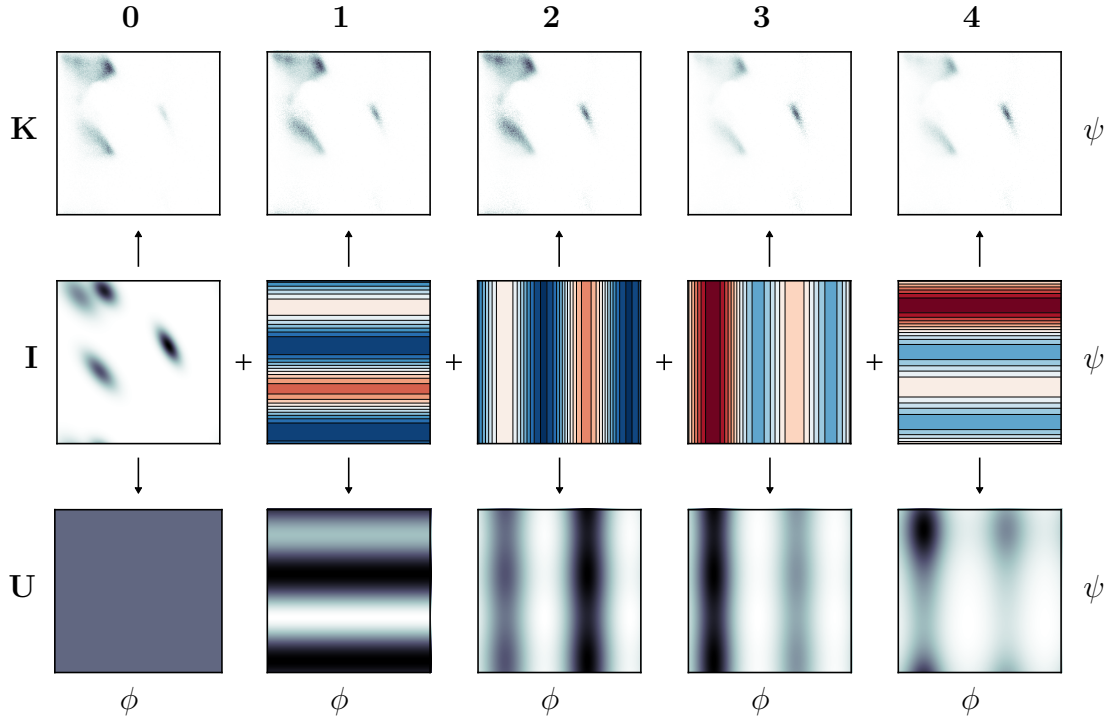

Figure 5.5: Illustration of the MaxEnt assessment: $I0$ is the constraint generating sample distribution (Figure 5.3), $I1$ to $I4$ indicates the consecutive inclusion of CCR ensemble averages (see Figure 5.4). Rows K and U are the corresponding *knowledge-based* and *uniform* prior MaxEnt predictions.

negative relaxation contribution of only 25 % α_l dominates the ensemble average to an extent that deems other motives unlikely.

The fine-grained shape of $K1$ might look more realistic, but this persuasive appearance is only due to the prior assumptions in $K0$. A single experiment yields no noteworthy restraints, the predictions are dominated by the prior. Additional constraints do lead to more distinct predictions, best seen in the evolution from $U1$ to $U4$. However, the highest densities still coincide with the extrema of the constraint functions, indicating that not enough conflicting information has been introduced up to this point.

This is not unexpected of course, Kloiber et al. used two additional experiments to investigate far less ambiguous distributions.

5.1.2   Extending established protocols

The above experiments (Figure 5.4) probe either ϕ or ψ, because the position of the H^α-C^α bond vector does not vary upon changes of the neighboring dihedral angles. While featuring an easy to interpret Karplus-like shape, this does of course mean that there is no pairwise distinction of regions in (ϕ,ψ)-space; functions of both ϕ and ψ arguably contain more information. To resolve this Kloiber et al. chose to probe $\Gamma_{H^\alpha_{i-1}C^\alpha_{i-1},H^\alpha_i C^\alpha_i}(\phi_i,\psi_{i-1})$ [40], an interaction extending to neighboring residues. This has two disadvantages, the first being an effective decrease of accessible residues due to the exclusion of glycine and proline. The second drawback lies again in the fact that H^α-C^α does not change upon variation of ϕ or ψ, limiting overall combinatorial possibilities: To yield a function of two dihedral angles H^α-C^α must be related to a H^α-C^α bond of a neighboring residue. C' and H^N-N, however, can be varied in three pairwise interactions as functions of ϕ and ψ. Thus, forgoing $\Gamma_{H^\alpha_{i-1}C^\alpha_{i-1},H^\alpha_i C^\alpha_i}(\phi_i,\psi_{i-1})$ allows for the inclusion of three alternative experiments: $\Gamma_{H^N_i N_i,H^N_{i+1}N_{i+1}}(\phi_i,\psi_i)$ [104], $\Gamma_{C'_{i-1},C'_i}(\phi_i,\psi_i)$ [105] and $\Gamma_{H^N_i N_i,C'_i}(\phi_i,\psi_i)$ [106], which are depicted in Figure 5.6.

Figure 5.7 shows the addition of the relaxation rates in Figure 5.6 to the MaxEnt evolution from above (Figure 5.5). Again the improvements are most apparent in row U as α_l and α_r are captured with higher accuracy, albeit with shifted positions due to the extrema of $I5$ and $I6$. Note that the maximum of $I5$ shifts β into the upper left corner. $I6$, however, proves incompatible within this region, correcting towards lower ψ angles. Changes in row K are again more subtle: The prior shifts the increase in α_l to the correct position, also a slight pronunciation of the β-region can be seen. This is due to the similarity between $I0$ and $K0$: As most of the prediction is already present in the prior, no strong changes are caused by the constraints. The uniform prior on the contrary is shaped considerably by the constraints, allowing for a better assessment of the experimental
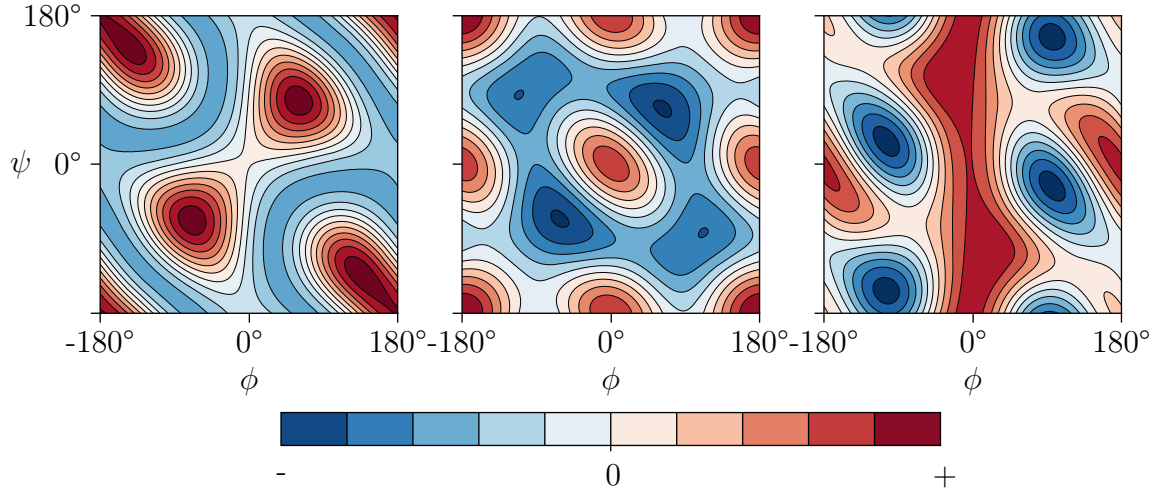
Figure 5.6: Geometrical modulations of different CCR rates according to Section 4.1. Left-hand side: $\Gamma_{H_i^N N_i, H_{i+1}^N N_{i+1}}(\phi, \psi)$. Middle: $\Gamma_{C'_{i-1}, C'_i}(\phi, \psi)$ Right-hand side: $\Gamma_{H_i^N N_i, C'_i}(\phi, \psi)$.
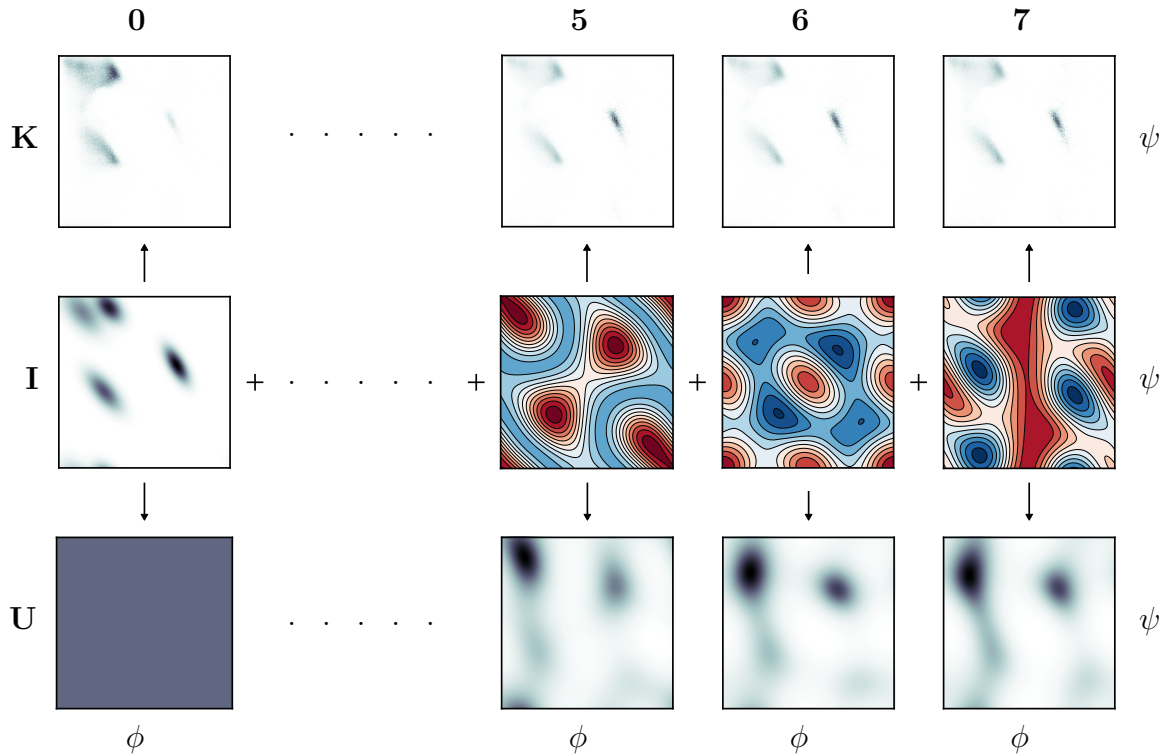

Figure 5.7: Extension of the MaxEnt assessment in Figure 5.5: $I0$ is the constraint generating sample distribution (Figure 5.3), $I5$ to $I7$ extends the consecutive inclusion of CCR ensemble averages (see Figure 5.6). Rows K and U are the corresponding *knowledge-based* and *uniform* prior MaxEnt predictions.

information content: At this stage the procedure still needs more restraints to allow for meaningful deviations from the prior. Similarities between an investigated distribution and its MaxEnt approximation might still be an artifact; additional experiments are necessary. However, all possible combinations of routinely used dipolar vectors and CSAs have been exhausted. The nitrogen CSA might come to mind as an additional possibility, but since it is virtually axial symmetric and its principal axis is quasi collinear with the

H^N-N bond [107] it offers no relevant additional information. The predictions can only be improved if more uncommon interactions are considered.

5.1.3 Suggesting new experiments

Options for CCR experiments are of course limited by the protein backbone, but there are dipolar vectors which are still mostly unexplored. $H^\alpha H^N$ is an example which has proven its viability in a first application by Crowley et al. [108] probing $\Gamma_{H_i^N N_i, H_i^N H_i^\alpha}(\phi)$. Unlike the dipolar vectors surveyed up to this point, $H^\alpha H^N$ is not parallel to a chemical bond. As a consequence, it exhibits an angular-dependent distance modulation and does not show the same symmetric properties of the previous relaxation interactions. This symmetry breaking property can act as an additional source of information which has not been utilized up to this point.

Which additional relaxation rates should be probed? Generally speaking, both $H_i^\alpha H_i^N$ and $H_i^\alpha H_{i+1}^N$ are valid options and many different interactions could be considered. However, an exhaustive combination of all available vectors and/or CSAs as above is not practically feasible: Due to spin relaxation, pulse sequences cannot be arbitrarily long and complex. Furthermore, their design and implementation takes considerable effort and measuring time is not an unlimited resource. Thus, a selection of few but informative experiments should be preferred.

Exhaustive sampling of all available combinations is of course difficult even with other possible dipolar vectors neglected. Possible interactions were instead assessed by their practical feasibility and their general effect on the MaxEnt predictions above (Figure 5.7). An optimum was found in *two interactions* which can be probed in a *single experiment*: $\Gamma_{C_i', H_i^\alpha H_{i+1}^N}(\psi)$ and $\Gamma_{C_i', H_i^\alpha H_i^N}(\phi, \psi)$, shown in Figure 5.8.

The dominant maximum in the important β and PP-II region is the most noteworthy property of $\Gamma_{C_i', H_i^\alpha H_{i+1}^N}(\psi)$. Since the relative positions of C_i' and H_{i+1}^N do not change with ψ, no negative relaxation can be observed. The shape is therefore defined by the angle-dependent distance modulation of $H_i^\alpha H_{i+1}^N$.

As a function of ϕ and ψ, $\Gamma_{C_i', H_i^\alpha H_i^N}(\phi, \psi)$ exhibits a more diverse pattern. Due to the higher distance between H_i^α and H_i^N with negative ϕ, the highly populated regions around α_r, β and PP-II show weak relaxation contributions. Both positive and negative extrema are found at positive ϕ angles, which are populated only by the α_l motive.

In Figure 5.9 the addition of the two experiments to the MaxEnt protocol from above (Figure 5.7) is shown. $\Gamma_{C_i', H_i^\alpha H_{i+1}^N}(\psi)$ ($I8$) proves to be highly specific, as can be seen in $U8$: The prediction of $U7$ proves incompatible with the additional constraint of $I8$, splitting
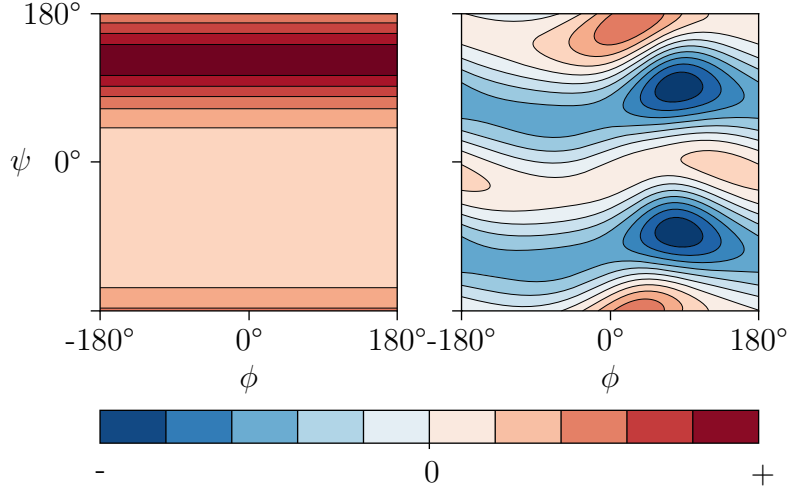
Figure 5.8: Geometrical modulations of different CCR rates according to Section 4.1. Left-hand side: $\Gamma_{C'_i, H^\alpha_i H^N_{i+1}}(\psi)$. Right-hand side: $\Gamma_{C'_i, H^\alpha_i H^N_i}(\phi, \psi)$.
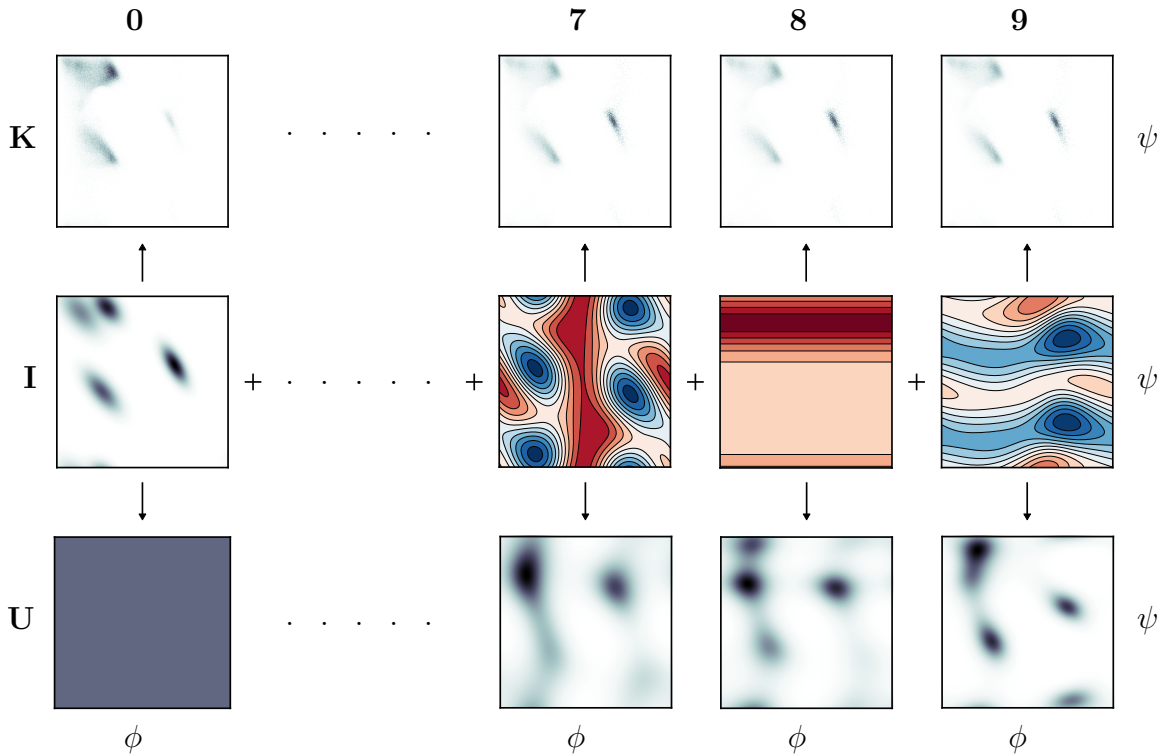

Figure 5.9: Extension of the MaxEnt assessment in Figures 5.5 and 5.7: $I0$ is the constraint generating sample distribution (Figure 5.3), $I7$ to $I9$ extends the consecutive inclusion of CCR ensemble averages (see Figures 5.6 and 5.8). Rows K and U are the corresponding *knowledge-based* and *uniform* prior MaxEnt predictions.

the single peak in the β-PP-II region of $U7$ in two. With the addition of $\Gamma_{C'_i, H^\alpha_i H^N_i}(\phi, \psi)$ ($I9$) the split is blurred and densities are shifted. Most notably the position of α_l moves towards more realistic ψ angles. In this case, not the extrema but their absence act as the source of information. An interaction without notable relaxation in the regions of interest might not be useful on its own, but can be relevant in conjunction with other experiments. While not considered crucial, the lacking resolution in the β and PP-II region definitely

leaves room for improvement. A selective relaxation pattern might possibly resolve this convoluted region. The emergent density in the lower right edge of $U9$ is caused by $I9$ and can safely be treated as an artifact, which could be suppressed by adequate prior assumptions as in $K9$.

The general effect of these assumptions, however, cannot be understated: The evolution of rows U and K differ substantially in shape and fine structure. Only if different priors are compared, meaningful conclusions can be derived. While certainly not the most realistic, the uniform prior proves useful as an "unbiased" internal standard.

Still, here we argue that based on the achieved similarity between the sample distribution $I0$ and its MaxEnt reproductions $U9$ and $K9$ it is safe to conclude: *CCR experiments can yield enough geometrical information to characterize the backbone dihedral angle populations of IDPs.*

To derive this general statement we relied on a particular set of prior assumptions and experiments as well as physical simplifications. Generalizing, implementing and optimizing this approach is arguably a complex goal and would go beyond the scope of this work. The suggested protocol is first and foremost a proof of concept, its characteristics and deficiencies are elaborated in the following section.

5.2 Protocol evaluation

This section showcases MaxEnt predictions of selected sample distributions based on the approach above: The uniform and the knowledge-based prior (Figure 5.2) are employed with all nine constraints from Figures 5.4, 5.6 and 5.8, namely $\Gamma_{H_i^\alpha C_i^\alpha, C_i'}(\psi)$, $\Gamma_{C_{i-1}', H_i^\alpha C_i^\alpha}(\phi)$, $\Gamma_{H_i^N N_i, H_i^\alpha C_i^\alpha}(\phi)$, $\Gamma_{H_i^\alpha C_i^\alpha, H_{i+1}^N N_{i+1}}(\psi)$, $\Gamma_{H_i^N N_i, H_{i+1}^N N_{i+1}}(\phi, \psi)$, $\Gamma_{C_{i-1}', C_i'}(\phi, \psi)$, $\Gamma_{H_i^N N_i, C_i'}(\phi, \psi)$, $\Gamma_{C_i', H_i^\alpha H_{i+1}^N}(\psi)$ and $\Gamma_{C_i', H_i^\alpha H_i^N}(\phi, \psi)$.

5.2.1 Resolving localized distributions

While designed to characterize broad (ϕ, ψ)-distributions, one might ask how the suggested procedure handles narrow ones? With three additional experiments it should at least match the results of Kloiber et al. Assuming a single α_{rc} peak in the canonical α region, Figure 5.10 shows the corresponding MaxEnt solutions. Again I indicates the "input", U and K stand for the "uniform" and the "knowledge-based" prior respectively.

Both U and K are very good approximations of I. The knowledge-based prior causes the fine grained structure of K, while the uniform prior determines the smooth appearance
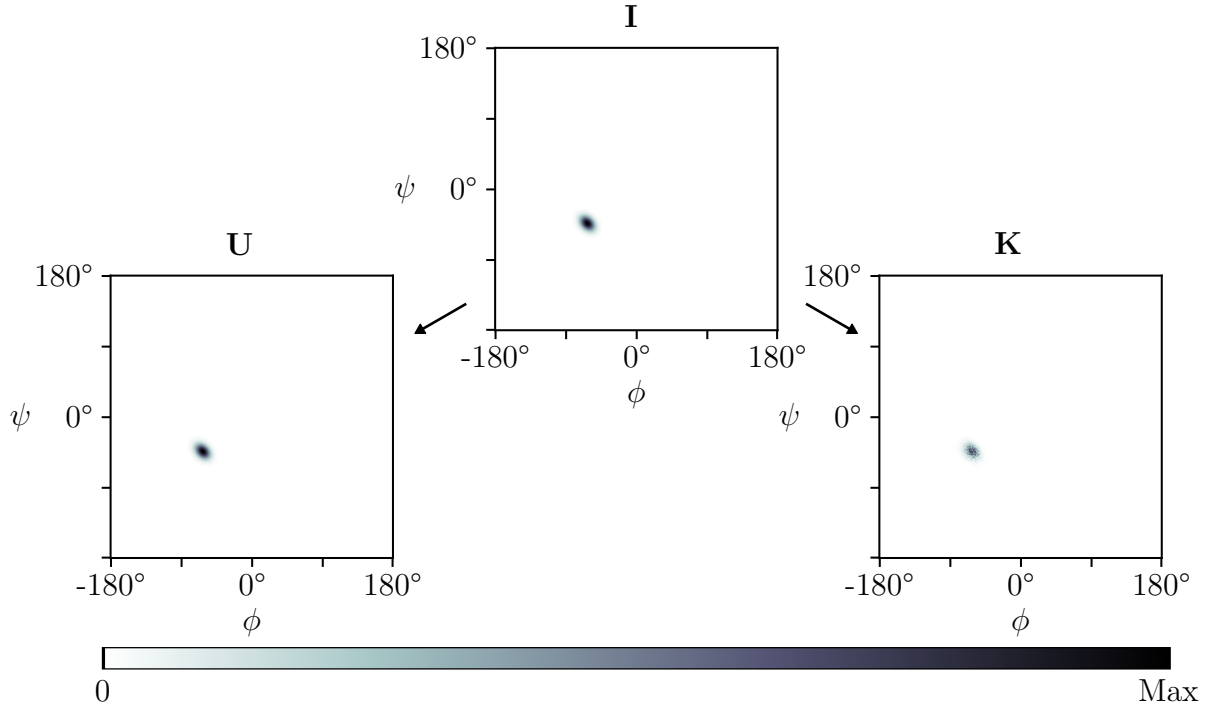
Figure 5.10: MaxEnt predictions U and K based on the input I ($100\,\%\;\alpha_{rc}$) employing all 9 CCR rates from Figures 5.4, 5.6 and 5.8. U and K stand for the *uniform* and the *knowledge-based* prior.

of U (Figure 5.2). Still, both solutions are mere representations of the same constraints. For localized distributions, such as this one, the procedure yields very good agreements.

5.2.2 Characterizing diverse distributions

As illustrated in Figure 5.11, more diverse distributions are clearly more challenging. The sample distribution I consists of equally weighted motives highly different in shape: α_l, α_r, α_{rc}, β, β_{ext} and PP-II.

Unlike before (Figure 5.10), U can no longer resolve the sharp peak of α_{rc}. By including broader motives, the MaxEnt solution can better conform to the pressure towards uniformity: Without any motive preference, α_r and α_{rc} become blurred, much like β and PP-II which are too similar to be resolved. The region β_{ext} extended towards lower ψ angles is properly captured, the previously observed mirrored density of α_l at negative ψ is an artifact introduced from $\Gamma_{C_i',H_i^\alpha H_i^N}(\phi,\psi)$ (Figure 5.8). K on the other hand looks more accurate. This is of course related to the regions emphasized in the prior (Figure 5.2), which again highlights the relative nature of MaxEnt solutions. Still, diverse distributions such as I appear qualitatively accessible. A detailed, possibly quantitative characterization, however, will require sound assessments of the prior assumptions. Designing a more capable experimental protocol might prove worthwhile to circumvent this complication.
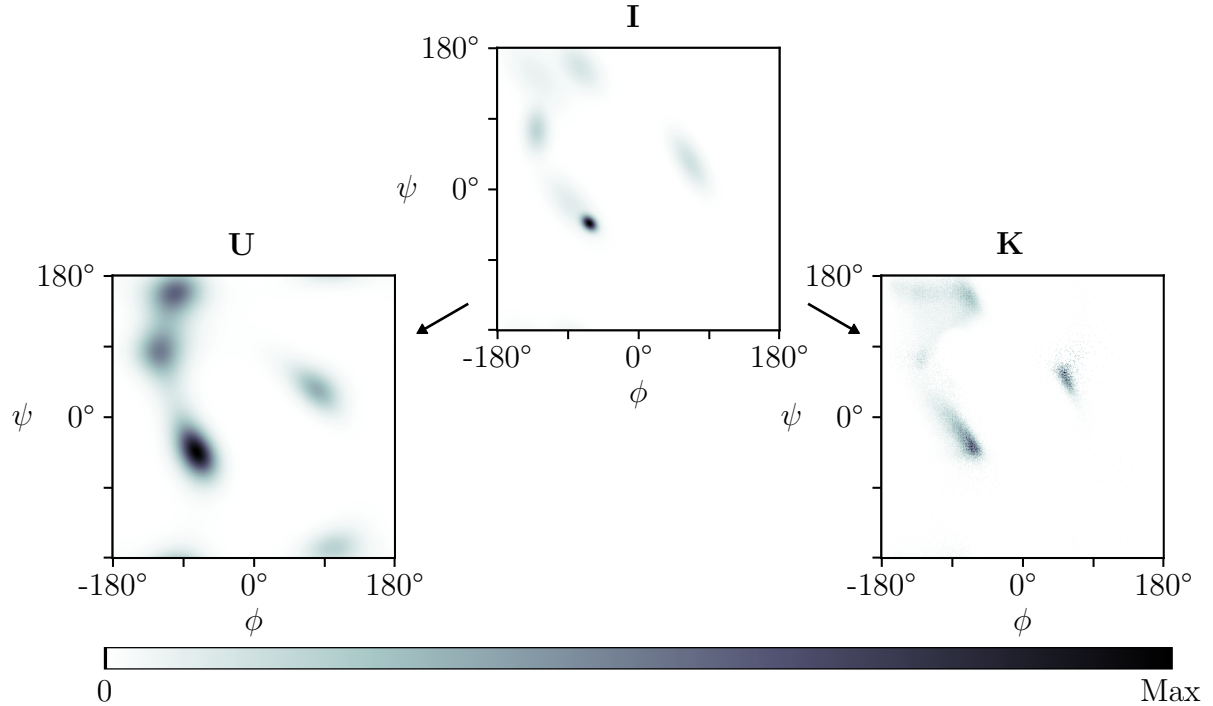
Figure 5.11: MaxEnt predictions U and K based on the input I (equal ratios of α_l, α_r, α_{rc}, β, β_{ext} and PP-II) employing all 9 CCR rates from Figures 5.4, 5.6 and 5.8. U and K stand for the *uniform* and the *knowledge-based* prior.

5.2.3 Distinguishing β and PP-II regions

As seen before, the MaxEnt solutions do not resolve the distinct β and PP-II regions of the sample distributions, instead the densities become blurred. This lack in resolution is an obvious shortcoming of the proposed protocol. In Figure 5.12 a compilation of predictions with different ratios of β and PP-II is shown.

At 0 and 100 % PP-II the predicted densities are highly accurate, the β region of the knowledge-based predictions K is slightly more spread out. For ratios in between the extrema, the distinct β and PP-II densities of I tend to merge. Still, the position of the merged densities in both U and K could be used as an indicator for the underlying ratios; while the representation might not be resolved in detail, it still offers a convoluted measure to draw conclusions from. Better resolution might be achieved with additional, ideally highly selective experiments.

5.2.4 Quantitative assessment

To asses the quality of the suggested protocol, sample and MaxEnt distributions were only compared by general appearance up to this point. Here, a simple quantification is shown for sample distributions with different ratios of α_r and β. Both are important secondary structure elements which can be clearly distinguished. Since one ratio determines the
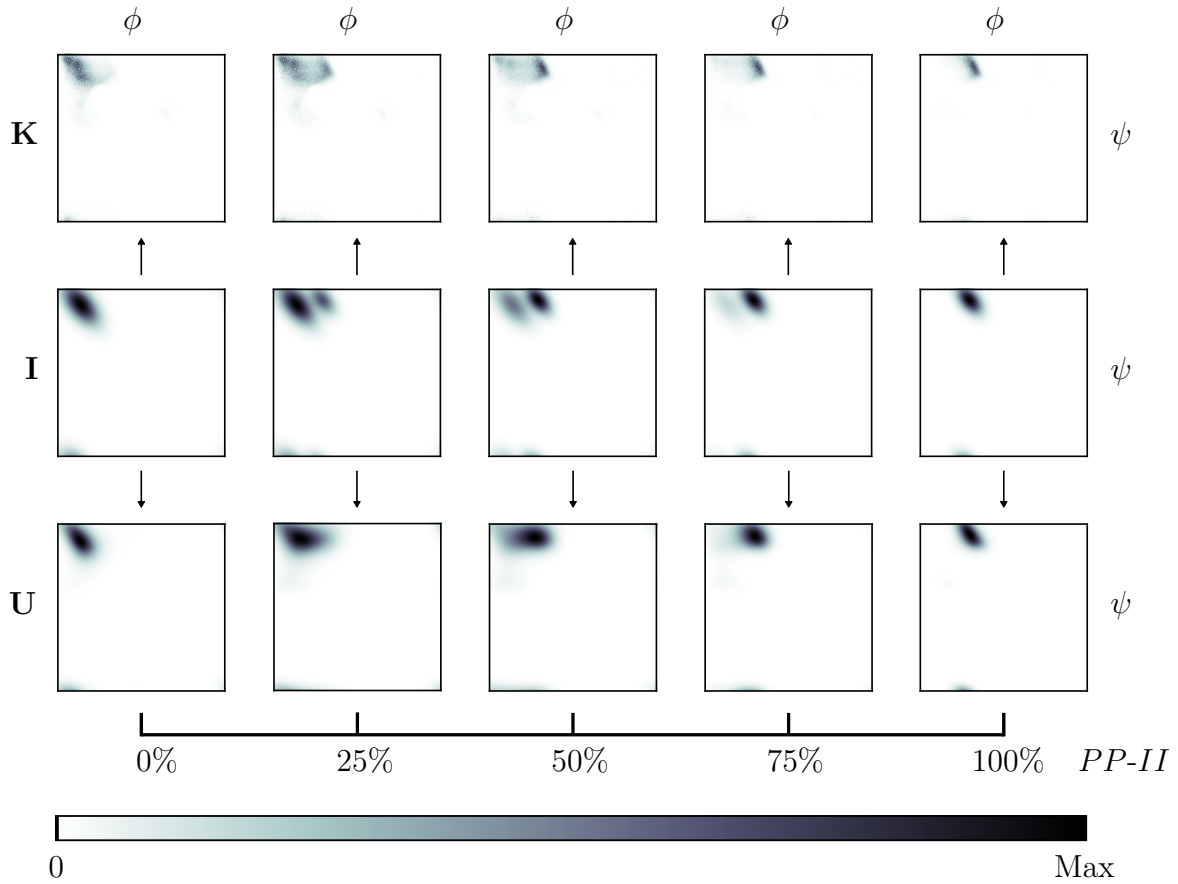
Figure 5.12: MaxEnt predictions in row K and U based on the inputs in row I with different ratios of PP-II and β employing all 9 CCR rates from Figures 5.4, 5.6 and 5.8. U and K stand for the *uniform* and the *knowledge-based* prior.

other, the results can be intuitively illustrated as a function of e.g. β [%].

Figure 5.13 shows the predicted densities for sample distributions from 0 to 100 % β in 10 % steps, the case of perfect agreement is indicated by the dashed $x = y$ line. Both the uniform- and knowledge-based solutions are very accurate. A general trend to be observed is the over-representation of the minority motive: Low ratios of β tend to be overestimated, high ratios are underestimated. This might likely be due to a preference for broader, higher entropy distributions. This trend is superimposed by the fact that all predicted ratios of β are lower for the knowledge-based prior than the uniform one. Again, without certainty, the uniform prior might favor the broader, higher entropy β region, while the α_r region occurs more dominantly in the knowledge-based prior. Still, these trends are of rather negligible influence.

Other important, albeit less intuitive, results are the Lagrange parameters. They are uniquely determined by the combination of relaxation rates, their ensemble averages and the prior define the shape of the MaxEnt solutions (equation 4.12).

Figure 5.14 depicts the variations of the Lagrange parameters with different ratios of β
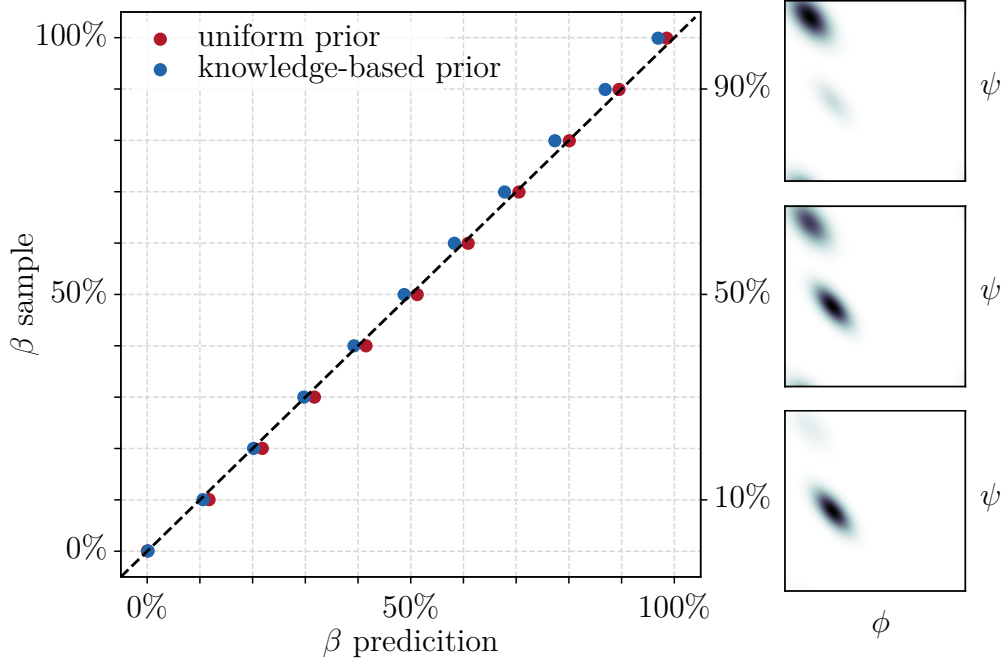
Figure 5.13: Comparison of β and $\alpha_r = 100\% - \beta$ ratios in sample distributions and their corresponding MaxEnt predictions employing all 9 CCR rates from Figures 5.4, 5.6 and 5.8.

for all employed relaxation rates. On the secondary ordinate the corresponding ensemble average is depicted, illustrating the linear dependency between the average rates and the changes in β ratio.

The first row of Figure 5.14 shows the rates dependent on ψ only. These are of course functions with high distinctive potential, as α_r and β lie in very different ψ-regions. For both the uniform and the knowledge-based prior the Lagrange parameters behave similarly, 0 and 100 % mark notable discontinuities. $\Gamma_{C_i', H_i^\alpha H_{i+1}^N}(\psi)$ (Figure 5.8), which is highly sensitive for β, shows this effect the strongest. Between 10 and 90 % the graphs are linear: When both motives are present, even with different ratios, the relaxation rates are weighted with comparable Lagrange multipliers. If only one motive is present, the Lagrange multipliers are evaluated differently. Since they are the partial derivatives of the entropy with respect to their constraints (equation 3.21), this indicates a significant change in entropy.

At first glance the second row of Figure 5.14 paints a less clear picture. The Lagrange parameters are considerably different for the two priors, still the rates appear to be weighted comparably, if not stronger, than the ones in the first row. However, these functions exhibit a far lower distinctive potential: $\Gamma_{C_{i-1}', H_i^\alpha C_i^\alpha}(\phi)$ and $\Gamma_{H_i^N N_i, H_i^\alpha C_i^\alpha}(\phi)$ (Figure 5.4) are functions of ϕ only. Since α_r and β do not differ strongly in ϕ, the cusps at 0 and 100% are also less pronounced. While the rates cannot distinguish between the two motives, they can still be incorporated to build up density in the relevant regions, best seen for $\Gamma_{H_i^N N_i, H_i^\alpha C_i^\alpha}(\phi)$. This effect is generally stronger for the uniform prior, as the knowledge-
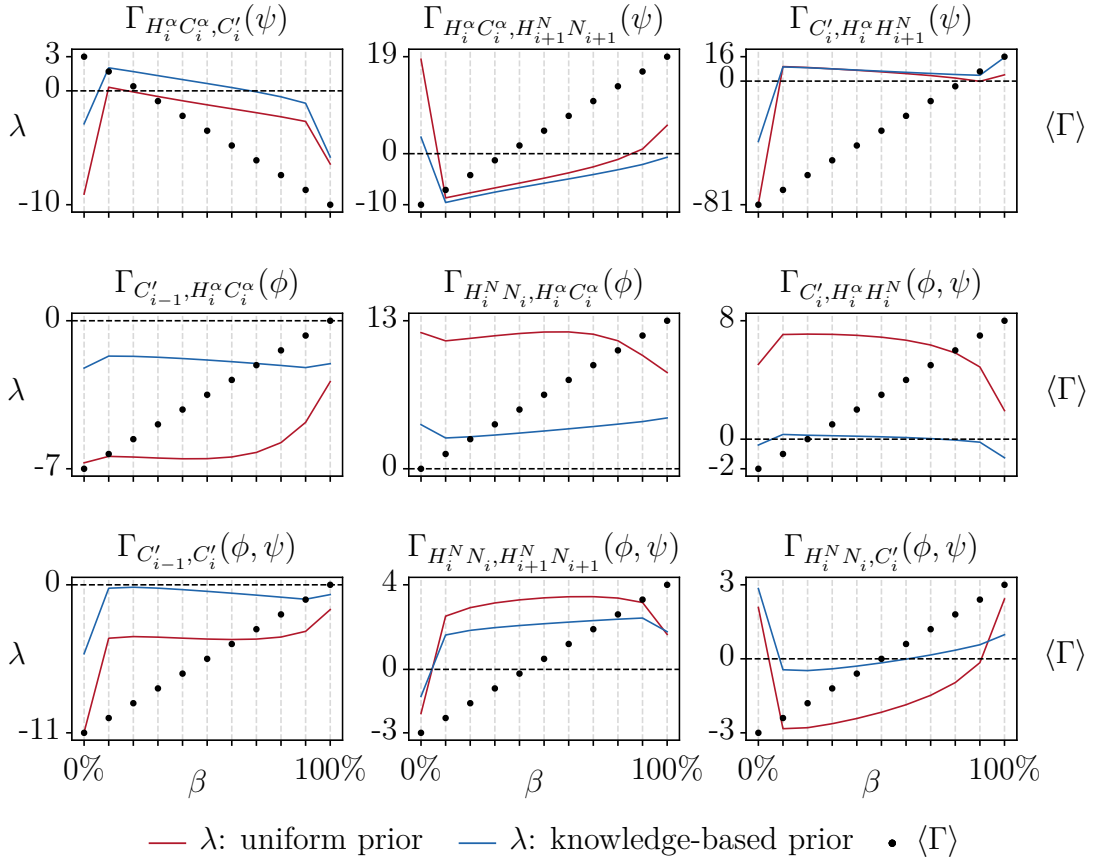
Figure 5.14: Evolution of the Lagrange parameters λ and ensemble averages $\langle\Gamma\rangle$ of all 9 CCR rates (Figures 5.4, 5.6 and 5.8) with different ratios of β and $\alpha_r = 100\% - \beta$ corresponding to Figure 5.13.

based one does not weigh all angle pairs equally. While $\Gamma_{C_i', H_i^\alpha H_i^N}(\phi, \psi)$ (Figure 5.8) is in principle a function of ψ as well, its extrema lie in positive ϕ-regions and α_r and β relax with similar rates. The knowledge-based solutions do not utilize this function strongly. In contrast, the uniform prior predictions weigh the on average negative rates with a positive multiplier to shift the density towards the less pronounced regions at negative ϕ.

The third row, featuring functions of both ϕ and ψ, exhibits the same properties as the first one, albeit less distinct due to their overall more complex shape. The uniform-based predictions again feature larger absolute multipliers to better distinguish between regions with positive and negative relaxation contributions. This substantiates the general observation that the uniform prior tends to incorporate more experimental input; an important property certainly worth exploiting.

As demonstrated, the assessment of the Lagrange multipliers is a useful tool for characterizing MaxEnt solutions. However, more complex distributions will inevitably lead to more convoluted and less intuitive parameters; interpreting them in practical applications might prove a nontrivial but worthwhile task.

Chapter 6

Summary

An experimental protocol for the investigation of IDP backbone dihedral angle distributions was derived from theory. The use of cross-correlated relaxation (CCR) for IDP characterization has been mostly unexplored to date, the proposed combination of experiments therefore serves as a general proof of concept. A real-life implementation still requires a sound assessment of the simplified dynamic models, pulse sequences adapted for IDPs and a practical routine to account for experimental uncertainties.

The inherent underdetermination of IDP dihedral angle populations, resulting from their high conformational freedom, was addressed using a Maximum Entropy (MaxEnt) approach, which identifies the highest entropy distribution as the most representative for the experimental constraints at hand. The particular implementation in this work differs from previous studies not only in its inclusion of CCR rates: It is argued that the initial assumptions encoded in the so-called prior should be treated with more care. The common use of a singular, highly sophisticated prior and a fixed set of NMR parameters hinders the distinction between experimental and assumptive information content. Instead, a step-by-step extension is proposed, evaluating combinations of experiments relative to different priors. This way, the viability and potency of the protocol can be assessed with more modesty. The classical uniform prior was highlighted in this regard: While neither epistemologically nor physically representative, its unbiased nature allows for a particularly simple and intuitive illustration of experimentally supplied information.

This framework was used not only to evaluate the applicability of existing CCR protocols but also to extend them. The addition of experiments not considered in previous protocols was found to improve the quality of the predictions considerably. The subsequent introduction of two entirely novel interactions lead to predictions with satisfactory accuracy. The MaxEnt framework was shown to offer a simple guideline in the design and assessment of such combinational experimental protocols. While derived exclusively for

CCR rates in this study, this observation applies universally.

The proposed combination of nine CCR rates is highly potent but of course still leaves room for improvement: While the theoretical capabilities are promising both qualitatively and quantitatively, there are obvious shortcomings regarding the distinction of highly diverse distributions as well as overall resolution; first and foremost the protocol is intended as a proof of concept. While presented in an exclusive framework here, CCR experiments can of course be used in combination with other NMR parameters as well.

Naturally, the next step involves a practical implementation of this approach. Not only experimental uncertainties need to be addressed but also the dynamic contributions which were omitted in this study. While dynamics of IDPs might pose additional challenges to overcome, the possibility of probing both structural and dynamic properties further highlights the unharnessed potential of CCR rates in the characterization of IDPs.

Bibliography

[1] A. K. Dunker et al. "Intrinsically disordered protein". In: *Journal of Molecular Graphics and Modelling* 19.1 (2001), pp. 26–59.

[2] V. N. Uversky, J. R. Gillespie, and A. L. Fink. "Why Are "Natively Unfolded" Proteins Unstructured Under Physiologic Conditions?" In: *Proteins: Structure, Function, and Bioinformatics* 41.3 (2000), pp. 415–427.

[3] P. E. Wright and H. J. Dyson. "Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm". In: *Journal of Molecular Biology* 293.2 (1999), pp. 321–331.

[4] P. Tompa. "Intrinsically unstructured proteins". In: *Trends in Biochemical Sciences* 27.10 (2002), pp. 527–533.

[5] P. Tompa. "Intrinsically disordered proteins: a 10-year recap". In: *Trends in Biochemical Sciences* 37.12 (2012), pp. 509–516.

[6] P. Tompa and A. Fersht. *Structure and Function of Intrinsically Disordered Proteins*. Boca Raton: CRC Press, 2009.

[7] I. C. Felli and R. Pierattelli, eds. *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*. Vol. 870. Advances in Experimental Medicine and Biology. Cham: Springer International Publishing, 2015.

[8] R. Konrat. "NMR contributions to structural dynamics studies of intrinsically disordered proteins". In: *Journal of Magnetic Resonance* 241 (2014), pp. 74–85.

[9] M. R. Jensen et al. "Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy". In: *Chemical Reviews* 114.13 (2014), pp. 6632–6660.

[10] M. R. Jensen et al. "Defining Conformational ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts". In: *Journal of the American Chemical Society* 132.4 (2010), pp. 1270–1272.

[11]   J. A. Marsh and J. D. Forman-Kay. "Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints". In: *Journal of Molecular Biology* 391.2 (2009), pp. 359–374.

[12]   P. Bernadó et al. "A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.47 (2005), pp. 17002–17007.

[13]   G. Nodet et al. "Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings". In: *Journal of the American Chemical Society* 131.49 (2009), pp. 17908–17918.

[14]   M. R. Jensen et al. "Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings". In: *Structure* 17.9 (2009), pp. 1169–1185.

[15]   D. Ganguly and J. Chen. "Structural Interpretation of Paramagnetic Relaxation Enhancement-Derived Distances for Disordered Protein States". In: *Journal of Molecular Biology* 390.3 (2009), pp. 467–477.

[16]   C. K. Fisher, A. Huang, and C. M. Stultz. "Modeling Intrinsically Disordered Proteins with Bayesian Statistics". In: *Journal of the American Chemical Society* 132.42 (2010), pp. 14919–14927.

[17]   C. K. Fisher and C. M. Stultz. "Constructing ensembles for intrinsically disordered proteins". In: *Current Opinion in Structural Biology* 21.3 (2011), pp. 426–431.

[18]   T. Massad et al. "Maximum entropy reconstruction of joint φ, ψ-distribution with a coil-library prior: the backbone conformation of the peptide hormone motilin in aqueous solution from φ and ψ-dependent J-couplings". In: *Journal of Biomolecular NMR* 38.2 (2007), pp. 107–123.

[19]   A. B. Mantsyzov et al. "A maximum entropy approach to the study of residue-specific backbone angle distributions in α-synuclein, an intrinsically disordered protein". In: *Protein Science* 23.9 (2014), pp. 1275–1290.

[20]   A. B. Mantsyzov et al. "MERA: a webserver for evaluating backbone torsion angle distributions in dynamic and disordered proteins from NMR data". In: *Journal of Biomolecular NMR* 63.1 (2015), pp. 85–95.

[21]   M. Karplus. "Contact Electron-Spin Coupling of Nuclear Magnetic Moments". In: *The Journal of Chemical Physics* 30.1 (1959), pp. 11–15.

[22] L. J. Smith et al. "Analysis of Main Chain Torsion Angles in Proteins: Prediction of NMR Coupling Constants for Native and Random Coil Conformations". In: *Journal of Molecular Biology* 255.3 (1996), pp. 494–506.

[23] B. Reif, M. Hennig, and C. Griesinger. "Direct Measurement of Angles Between Bond Vectors in High-Resolution NMR". In: *Science* 276.5316 (1997), pp. 1230–1233.

[24] H. Schwalbe et al. "Cross-Correlated Relaxation for Measurement of Angles between Tensorial Interactions". In: *Methods in Enzymology* (2001), pp. 35–81.

[25] K. Kloiber, W. Schüler, and R. Konrat. "Automated NMR determination of protein backbone dihedral angles from cross-correlated spin relaxation". In: *Journal of Biomolecular NMR* 22.4 (2002), pp. 349–363.

[26] J. Stanek et al. "Probing Local Backbone Geometries in Intrinsically Disordered Proteins by Cross-Correlated NMR Relaxation". In: *Angewandte Chemie International Edition* 52.17 (2013), pp. 4604–4606.

[27] A. G. Redfield. "On the Theory of Relaxation Processes". In: *IBM Journal of Research and Development* 1.1 (1957), pp. 19–31.

[28] A. G. Redfield. "The Theory of Relaxation Processes". In: *Advances in Magnetic and Optical Resonance* 1 (1965), pp. 1–32.

[29] A. Kumar, R. C. R. Grace, and P. K. Madhu. "Cross-correlations in NMR". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 37.3 (2000), pp. 191–319.

[30] N. Murali and V. V. Krishnan. "A Primer for Nuclear Magnetic Relaxation in Liquids". In: *Concepts in Magnetic Resonance Part A* 17A.1 (2003), pp. 86–116.

[31] J. Kowalewski and L. Mäler. *Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications*. Series in Chemical Physics 2. Boca Raton: CRC Press, 2006.

[32] J. Cavanagh et al. *Protein NMR Spectroscopy: Principles and Practice*. 2nd Edition. Burlington: Academic Press, 2007.

[33] G. Lipari and A. Szabo. "Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity". In: *Journal of the American Chemical Society* 104.17 (1982), pp. 4546–4559.

[34] A. Abyzov et al. "Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation". In: *Journal of the American Chemical Society* 138.19 (2016), pp. 6240–6251.

[35] N. Salvi, A. Abyzov, and M. Blackledge. "Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation". In: *The Journal of Physical Chemistry Letters* 7.13 (2016), pp. 2483–2489.

[36] G. Parigi et al. "Long-Range Correlated Dynamics in Intrinsically Disordered Proteins". In: *Journal of the American Chemical Society* 136.46 (2014), pp. 16201–16209.

[37] P. Pelupessy et al. "Simultaneous determination of Ψ and Φ angles in proteins from measurements of cross-correlated relaxation effects". In: *Journal of Biomolecular NMR* 14.3 (1999), pp. 277–280.

[38] D. Yang, R. Konrat, and L. E. Kay. "A Multidimensional NMR Experiment for Measurement of the Protein Dihedral Angle ψ Based on Cross-Correlated Relaxation between 1Hα-13Cα Dipolar and 13C' (Carbonyl) Chemical Shift Anisotropy Mechanisms". In: *Journal of the American Chemical Society* 119.49 (1997), pp. 11938–11940.

[39] K. Kloiber and R. Konrat. "Measurement of the protein backbone dihedral angle φ based on quantification of remote CSA/DD interference in inter-residue 13C'(i - 1)-13Cα(i) multiple-quantum coherences". In: *Journal of Biomolecular NMR* 17.3 (2000), pp. 265–268.

[40] E. Chiarparin et al. "Relative Orientation of CαHα-Bond Vectors of Successive Residues in Proteins through Cross-Correlated Relaxation in NMR". In: *Journal of the American Chemical Society* 122.8 (2000), pp. 1758–1761.

[41] J.-S. Hu and A. Bax. "Measurement of Three-Bond 13C-13C J Couplings between Carbonyl and Carbonyl/Carboxyl Carbons in Isotopically Enriched Proteins". In: *Journal of the American Chemical Society* 118.34 (1996), pp. 8170–8171.

[42] E. T. Jaynes. "Information Theory and Statistical Mechanics". In: *Physical Review* 106.4 (1957), pp. 620–630.

[43] C. E. Shannon. "A Mathematical Theory of Communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423.

[44] A. Y. Khinchin. "The concept of entropy in the theory of probability". [Russian]. In: *Uspekhi Matematicheskikh Nauk* 8.3(55) (1953), pp. 3–20.

[45] A. Y. Khinchin. *Mathematical Foundations of Information Theory.* Trans. by R. A. Silverman and M. D. Friedman. New York: Dover Publications, 1957.

[46] D. K. Faddeev. "On the concept of entropy of a finite probabilistic scheme". [Russian]. In: *Uspekhi Matematicheskikh Nauk* 11.1(67) (1956), pp. 227–231.

[47] A. Rényi. "On Measures of Entropy and Information". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. Berkeley: University of California Press, 1961, pp. 547–561.

[48] C. Tsallis. "Possible Generalization of Boltzmann-Gibbs Statistics". In: *Journal of Statistical Physics* 52.1-2 (1988), pp. 479–487.

[49] J. Aczél and Z. Daróczy. *On Measures of Information and Their Characterizations*. Vol. 115. Mathematics in Science and Engineering. New York: Academic Press, 1975.

[50] I. Csiszár. "Axiomatic Characterizations of Information Measures". In: *Entropy* 10.3 (2008), pp. 261–273.

[51] K. Conrad. *Probability distributions and maximum entropy*. URL: www.math.uconn.edu/~kconrad/blurbs (visited on 09/25/2017).

[52] A. Feinstein. *Foundations of Information Theory*. New York: McGraw-Hill, 1958.

[53] E. T. Jaynes. *Probability Theory: The Logic of Science*. Ed. by G. L. Bretthorst. Cambridge: Cambridge University Press, 2003.

[54] J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. New York: Charles Scribner's Sons, 1902.

[55] E. T. Jaynes. "New Engineering Applications of Information Theory". In: *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability*. New York: John Wiley & Sons, 1963, pp. 163–203.

[56] S. Kullback and R. A. Leibler. "On information and sufficiency". In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86.

[57] S. Kullback. *Information Theory and Statistics*. New York: John Wiley & Sons, 1959.

[58] E. T. Jaynes. "Information Theory and Statistical Mechanics (Notes by the lecturer)". In: *Statistical Physics*. Ed. by K. W. Ford. Vol. 3. Brandeis Lectures. New York: W. A. Benjamin, 1963, pp. 181–218.

[59] G. Bricogne. "Maximum Entropy and the Foundations of Direct Methods". In: *Acta Crystallographica Section A: Foundations of Crystallography* 40.4 (1984), pp. 410–445.

[60] G. Bricogne. "Direct Phase Determination by Entropy Maximization and Likelihood Ranking: Status Report and Perspectives". In: *Acta Crystallographica Section D: Biological Crystallography* 49.1 (1993), pp. 37–60.

[61] Z. Wu et al. "A Fast Newton Algorithm for Entropy Maximization in Phase Determination". In: *SIAM Review* 43.4 (2001), pp. 623–642.

[62] Y. Alhassid, N. Agmon, and R. D. Levine. "An upper bound for the entropy and its applications to the maximal entropy problem". In: *Chemical Physics Letters* 53.1 (1978), pp. 22–26.

[63] A. Charnes and W. W. Cooper. "Constrained Kullback-Leibler Estimation; Generalized Cobb-Douglas Balance, and Unconstrained Convex Programming." In: *Atti della Accademia nazionale dei Lincei. Rendiconti della Classe di scienze fisiche, matematiche e naturali* 58.4 (1975), pp. 568–576.

[64] A. Ben-Tal and A. Charnes. *A Dual Optimization Framework for Some Problems of Information Theory and Statistics.* Tech. rep. CCS 310. Austin: Center for Cybernetic Studies, 1977.

[65] A. Charnes, W. W. Cooper, and L. Seiford. "Extremal Principles and Optimization Dualities for Khinchin-Kullback-Leibler Estimation". In: *Mathematische Operationsforschung und Statistik. Series Optimization: A Journal of Mathematical Programming and Operations Research* 9.1 (1978), pp. 21–29.

[66] I. Csiszár. "Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems". In: *The Annals of Statistics* 19.4 (1991), pp. 2032–2066.

[67] J. Shore and R. Johnson. "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy". In: *IEEE Transactions on Information Theory* 26.1 (1980), pp. 26–37.

[68] J. Skilling. "The axioms of maximum entropy". In: *Maximum-Entropy and Bayesian Methods in Science and Engineering.* Ed. by G. J. Erickson and C. R. Smith. Vol. 1. Amsterdam: Kluwer, 1988, pp. 173–187.

[69] J. B. Paris and A. Vencovská. "A Note on the Inevitability of Maximum Entropy". In: *International Journal of Approximate Reasoning* 4.3 (1990), pp. 183–223.

[70] A. Caticha. "Relative Entropy and Inductive Inference". In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering.* Vol. 707. AIP Conference Proceedings. 2004, pp. 75–96. URL: arXiv.org/abs/physics/0311093.

[71] A. Caticha and A. Giffin. "Updating Probabilities". In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering.* Vol. 872. AIP Conference Proceedings. 2006, pp. 31–42. URL: arxiv.org/abs/physics/0608185.

[72] A. Giffin and A. Caticha. "Updating Probabilities with Data and Moments". In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering.* Vol. 954. AIP Conference Proceedings. 2007, pp. 74–84. URL: arxiv.org/abs/0708.1593.

[73] A. Caticha. "Information and Entropy". In: *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Vol. 954. AIP Conference Proceedings. 2007, pp. 11–22. URL: arxiv.org/abs/0710.1068.

[74] K. H. Knuth and J. Skilling. "Foundations of Inference". In: *Axioms* 1.1 (2012), pp. 38–73.

[75] A. Caticha. "Lectures on Probability, Entropy, and Statistical Physics". In: *eprint arXiv:0808.0012v1 [physics.data-an]* (2008). URL: arxiv.org/abs/0808.0012v1.

[76] J. Skilling et al. "Critique of Information Geometry". In: *AIP Conference Proceedings* 1636.1 (2014), pp. 24–29.

[77] J. Skilling. "Failures of Information Geometry". In: *AIP Conference Proceedings* 1641.1 (2015), pp. 27–42.

[78] S. N. Karbelkar. "On the axiomatic approach to the maximum entropy principle of inference". In: *Pramana - Journal of Physics* 26.4 (1986), pp. 301–310.

[79] J. Uffink. "Can the Maximum Entropy Principle be Explained as a Consistency Requirement?" In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 26.3 (1995), pp. 223–261.

[80] R. P. Kostecki. "On Principles of Inductive Inference". In: *AIP Conference Proceedings* 1443.1 (2012), pp. 22–31.

[81] E. T. Jaynes. "Prior Probabilities". In: *IEEE Transactions on Systems Science and Cybernetics* 4.3 (1968), pp. 227–241.

[82] I. Sanov. "On the probability of large deviations of random variables". [Russian]. In: *Matematicheskii Sbornik* 42(84).1 (1957), pp. 11–44.

[83] I. N. Sanov. *On the probability of large deviations of random variables*. Tech. rep. No. 192. Raleigh: North Carolina State University, Dept. of Statistics, 1958.

[84] J. Van Campenhout and T. Cover. "Maximum Entropy and Conditional Probability". In: *IEEE Transactions on Information Theory* 27.4 (1981), pp. 483–489.

[85] I. Csiszár. "Sanov Property, Generalized I-Projection and a Conditional Limit Theorem". In: *The Annals of Probability* 12.3 (1984), pp. 768–793.

[86] I. Csiszár. "An Extended Maximum Entropy Principle and a Bayesian Justification". In: *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting: September 6/10, 1983*. Amsterdam: Elsevier, 1985, pp. 83–98.

[87] P. M. Williams. "Bayesian Conditionalisation and the Principle of Minimum Information". In: *The British Journal for the Philosophy of Science* 31.2 (1980), pp. 131–144.

[88] R. A. Fisher. "On the Mathematical Foundations of Theoretical Statistics". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 222.594-604 (1922), pp. 309–368.

[89] H. Akaike. "Information Theory and an Extension of the Maximum Likelihood Principle". In: *Selected Papers of Hirotugu Akaike*. Ed. by E. Parzen, K. Tanabe, and G. Kitagawa. Springer Series in Statistics. New York: Springer, 1998, pp. 199–213.

[90] H. Akaike. "A Bayesian analysis of the minimum AIC procedure". In: *Selected Papers of Hirotugu Akaike*. Ed. by E. Parzen, E. Tanabe, and G. Kitagawa. Springer Series in Statistics. New York: Springer, 1998, pp. 275–280.

[91] H. Akaike. "Prediction and entropy". In: *Selected Papers of Hirotugu Akaike*. Ed. by E. Parzen, K. Tanabe, and G. Kitagawa. Springer Series in Statistics. New York: Springer, 1998, pp. 387–410.

[92] Y. Tikochinsky, N. Z. Tishby, and R. D. Levine. "Alternative approach to maximum-entropy inference". In: *Physical Review A* 30.5 (1984), pp. 2638–2644.

[93] F. Topsøe. "Information-theoretical optimization techniques". In: *Kybernetika* 15.1 (1979), pp. 8–27.

[94] I. Csiszár. "Maxent, Mathematics, and Information Theory". In: *Maximum Entropy and Bayesian Methods*. Ed. by K. M. Hanson and R. N. Silver. Fundamental Theories of Physics 79. Amsterdam: Springer, 1996, pp. 35–50.

[95] J. Uffink. "The Constraint Rule of the Maximum Entropy Principle". In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 27.1 (1996), pp. 47–79.

[96] H.-O. Georgii. "Probabilistic aspects of entropy". In: *International Symposium on Entropy*. Max Planck Institute for Physics of Complex Systems, Dresden, 2000. URL: www.mathematik.uni-muenchen.de/~georgii/papers/Dresden.pdf (visited on 09/25/2017).

[97] M. D. Hanwell et al. "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform". In: *Journal of Cheminformatics* 4.17 (2012), pp. 1–17.

[98] F. A. Momany et al. "Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids". In: *The Journal of Physical Chemistry* 79.22 (1975), pp. 2361–2381.

[99]   Q. Teng, M. Iqbal, and T. A. Cross. "Determination of the carbon-13 chemical shift and nitrogen-14 electric field gradient tensor orientations with respect to the molecular frame in a polypeptide". In: *Journal of the American Chemical Society* 114.13 (1992), pp. 5312–5321.

[100]   S. A. Hollingsworth and P. A. Karplus. "A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins". In: *Biomolecular Concepts* 1.3-4 (2010), pp. 271–283.

[101]   A. B. Mantsyzov et al. *MERA: Maximum Entropy Ramachandran map Analysis from NMR data.* 2015. URL: `spin.niddk.nih.gov/bax/software/MERA` (visited on 09/25/2017).

[102]   E. Jones, T. Oliphant, and P. Peterson. *SciPy: Open source scientific tools for Python.* 2001. URL: `www.scipy.org` (visited on 09/25/2017).

[103]   M. Mechelke and M. Habeck. "Bayesian Weighting of Statistical Potentials in NMR Structure Calculation". In: *PLOS ONE* 9.6 (2014), e100197.

[104]   P. Pelupessy, S. Ravindranathan, and G. Bodenhausen. "Correlated motions of successive amide N-H bonds in proteins". In: *Journal of Biomolecular NMR* 25.4 (2003), pp. 265–280.

[105]   N. R. Skrynnikov et al. "Relative Orientation of Peptide Planes in Proteins Is Reflected in Carbonyl-Carbonyl Chemical Shift Anisotropy Cross-Correlated Spin Relaxation". In: *Journal of the American Chemical Society* 122.29 (2000), pp. 7059–7071.

[106]   K. Kloiber and R. Konrat. "Peptide Plane Torsion Angles in Proteins through Intraresidue 1H-15N-13C' Dipole-CSA Relaxation Interference: Facile Discrimination between Type-I and Type-II β-Turns". In: *Journal of the American Chemical Society* 122.48 (2000), pp. 12033–12034.

[107]   M. D. Lumsden et al. "Carbonyl Carbon and Nitrogen Chemical Shift Tensors of the Amide Fragment of Acetanilide and N-Methylacetanilide". In: *Journal of the American Chemical Society* 116.4 (1994), pp. 1403–1413.

[108]   P. Crowley, M. Ubbink, and G. Otting. "φ Angle Restraints in Protein Backbones from Dipole-Dipole Cross-Correlation between 1HN-15N and 1HN-1Hα Vectors". In: *Journal of the American Chemical Society* 122.12 (2000), pp. 2968–2969.