# NMR OF PROTEINS AND NUCLEIC ACIDS

KURT WÜTHRICH

# NMR of Proteins and Nucleic Acids

**Kurt Wüthrich**
Institute of Molecular Biology and Biophysics
Swiss Federal Institute of Technology
Zürich, Switzerland

**A Wiley-Interscience Publication**
**John Wiley & Sons**
New York • Chichester • Brisbane • Toronto • Singapore

***Library of Congress Cataloging in Publication Data:***

Wüthrich, Kurt.
NMR of proteins and nucleic acids.

(Baker lecture series)
"A Wiley-Interscience publication."
Bibliography: p.
Includes index.
1. Nuclear magnetic resonance spectroscopy.
2. Proteins—Analysis. 3. Nucleic acids—Analysis.
I. Title. II. Series. [DNLM: 1. Nuclear Magnetic Resonance—methods. 2. Nucleic Acids—analysis.
3. Proteins—analysis. QU 55 W973n]

QP519.9.N83W87 1986 574.19'245 86-7834
ISBN 0-471-82893-9

Printed in the United States of America

10 9 8 7 6 5 4 3 2

THE GEORGE FISHER BAKER
NON-RESIDENT LECTURESHIP
IN CHEMISTRY AT
CORNELL UNIVERSITY

# NMR of Proteins and Nucleic Acids

# Preface

The structure and function of proteins and nucleic acids are a long established, central theme of biological sciences, which has gained added interest by the fact that these biopolymers can now be modified efficiently to search for improved or new biological activities. Considering that the number of different variations for polynucleotide and polypeptide chains with $n$ building blocks are, respectively, $4^n$ and $20^n$, systematic screening of biopolymer sequences must be discarded as a practicable strategy for using genetic engineering or chemical synthesis in the design of new, improved compounds. More promising approaches are guided by basic knowledge of correlations of structure, conformation, and function of nucleic acids and proteins. This book treats the use of nuclear magnetic resonance for obtaining such information.

The potentialities and practice of NMR studies with proteins and nucleic acids were decisively changed during the past few years, so that perhaps the time has come to give a comprehensive introduction to the underlying principles and experimental procedures. It is the purpose of this volume to furnish such an account. In writing this book I have had in mind practicing scientists and students of biochemistry, chemistry, biophysics, and molecular biology who are concerned in their daily work with the structure and function of proteins and nucleic acids. This group can be expected to represent the primary users of the results of NMR studies on biopolymers. It is my hope that this text will provide a useful guide into the field, enabling NMR spectroscopists and nonspecialists alike to evaluate critically the potentialities and limitations of the method, and its applications in the primary literature.

KURT WÜTHRICH

*Zürich, Switzerland*
*March 1986*

# Acknowledgments

The incentive to prepare this text for publication has been provided by my tenure of the George Fisher Baker Nonresident Lectureship in Chemistry at Cornell University in September and October 1983. I wish to express my sincere thanks to Professor Roald Hoffmann and his colleagues in the Department of Chemistry for the invitation to present the Baker lectures and for the friendship and hospitality extended to me and my family during our period of residence in Ithaca.

My greatest debt is to the many friends, colleagues, and students in my group and around the globe, who have with their own work and their criticism over so many years provided inspiration and new incentives for continued activity. I would in particular like to mention the privilege of long standing, fruitful, and enjoyable collaboration with Professor R. R. Ernst and Professor G. Wagner, who also read large portions of this manuscript. Generous support in the preparation of this book was provided by my collaborators Dr. W. Braun, Dr. W. Chazin, Dr. T. Havel, Dr. A. Kline, Dr. W. Leupin, Dr. N. Müller, and Dr. E. Wörgötter and my students M. Billeter, G. Otting, T. Schaumann, P. Schultze, and H. Widmer, who gave permission to include unpublished results from their ongoing projects, and carefully read and made useful comments on various portions of the manuscript.

I am most grateful to those authors, publishers, and learned societies who gave permission to reproduce illustrations for which they hold the copyright. Due acknowledgment is made in each case in the text.

Finally I wish to thank Mrs. E. Huber and Mrs. E. H. Hunziker-Kwik, for having accepted the difficult task of typing the manuscript from my handwritten notes, Mr. R. Marani for the preparation of the reference list, and Mrs. E. H. Hunziker-Kwik for the carefully prepared illustrations.

K.W.

# Symbols and Abbreviations

Expressions, symbols, and units commonly used in nuclear magnetic resonance are introduced in Section 1.4 and at the outset of Chapter 5 [comprehensive listings may be found, e.g., in Ernst et al. (1986) and Jardetzky and Roberts (1981)]. The following are less common abbreviations:

| | |
|---|---|
| NOE | Nuclear Overhauser enhancement or nuclear Overhauser effect |
| TOE | Truncated driven NOE (a one-dimensional NMR experiment for measurement of NOE's) |
| 2D NMR | Two-dimensional nuclear magnetic resonance |
| COSY | 2D correlated spectroscopy |
| SECSY | 2D spin echo correlated spectroscopy |
| FOCSY | 2D foldover-corrected spectroscopy |
| NOESY | 2D nuclear Overhauser and exchange spectroscopy |
| 2D-*J* spectroscopy | 2D *J*-resolved spectroscopy |
| MQ, ZQ, 2Q | Multiple quantum, zero quantum, two quantum (e.g., MQ spectroscopy) |
| MQF, 2QF | Multiple quantum filter, two quantum filter (e.g., 2QF-COSY) |

All word combinations ending with COSY describe experiments for studies of scalar (through-bond) spin–spin couplings, for example,

| | |
|---|---|
| RELAYED-COSY | Relayed coherence transfer spectroscopy |

All word combinations ending with NOESY describe experiments for studies of dipolar (through-space) couplings manifested in nuclear Overhauser effects (or for studies of exchange phenomena), for example,

| | |
|---|---|
| $\omega_1$-decoupled NOESY | NOESY with homonuclear broadband decoupling along the $\omega_1$ frequency axis |

For the amino acids the IUPAC–IUB one-letter or three-letter symbols are used (Table 2.2). Individual atoms are identified using a prefix (e.g., NH is a backbone amide proton, 2H a hydrogen atom bound to the aromatic carbon atom 2, moreover $\alpha$H, $\beta$H, $\beta CH_2$, $\gamma CH_3$, 1NH, $\delta NH_2$). For the interpretation of certain NMR measurements, *pseudostructures* for amino acids are used, which contain the *pseudoatoms* K, L, M, P, Q, and QR as specified in Table 10.2. Polypeptide conformations are described by the torsion angles $\phi$, $\psi$, $\omega$, and $\chi^j$ (Fig. 7.2). The following abbreviations are used for proteins:

| | |
|---|---|
| BPTI | Basic pancreatic trypsin inhibitor from bovine organs |
| cBPTI | Cyclic analog of BPTI |
| Inhibitor K | Trypsin inhibitor homologue K from the venom of *Dendroaspis polylepis polylepis* |
| BUSI | Proteinase inhibitor IIA from bull seminal plasma |
| Tendamistat | $\alpha$Amylase polypeptide inhibitor from *Streptomyces tendae* |
| *lac* headpiece | DNA-binding domain 1–51 of the *lac* repressor from *E. coli* |
| metallothionein-2 | Metallothionein isoprotein 2 from rabbit liver |
| micelle-bound glucagon | Glucagon bound to perdeuterated dodecylphosphocholine micelles. |

For nucleosides or nucleotides the IUPAC–IUB one-letter symbols are used (Table 2.5). Individual atoms are identified using a prefix (Table 2.5, e.g., 1′H, 2H, 3NH, $6NH_2$). Polynucleotide conformations are described by the torsion angles $\alpha$–$\zeta$, $\nu_0$–$\nu_4$, and $\chi$ (Fig. 11.2).

*Spin systems* of the protons in amino acid residues and nucleotides are described by upper case letters (e.g., AX, AMX, $A_3B_3MX$; see Tables 2.2 and 2.5, Sections 2.1 and 2.2). The following symbols are used to identify types of spin systems representing different groups of amino acids:

| | |
|---|---|
| □ | AMX of $\alpha CH$–$\beta CH_2$ in Cys, Ser, Asp, Asn, Phe, Tyr, His, Trp |
| ■ | Pro, Met, Glu, Gln, Arg, Lys |
| ⧅ | Cys, Asp, Asn |
| ▲ | Met, Glu, Gln |
| ▼ | Arg, Lys |

*Proton–proton distances* have an important role in the analysis of the NMR data. In proteins, $d_{AB}(i,j)$ is the distance from proton A in amino acid residue $i$ to proton B in residue $j$ [e.g., $d_{\alpha N}(4,18)$, $d_{\alpha\beta}(i,i+3)$; $d_{\alpha N}(i,i+1) \equiv d_{\alpha N}$; see Section 7.1]. In nucleic acids the distances $d_i$(A;B), $d_s$(A;B), $d_{pi}$(A;B), and $d_{ps}$(A;B) are used as specified in Section 11.1.

The expressions CONFOR, DISGEO, and DISMAN indicate computer programs for the structural analysis of NMR data. The RMSD is the *root mean square distance* between different conformers.

THE GEORGE FISHER BAKER
NON-RESIDENT LECTURESHIP
IN CHEMISTRY AT
CORNELL UNIVERSITY

# NMR of Proteins and Nucleic Acids

# Contents


List of Tables xi

Symbols and Abbreviations xiii

CHAPTER 1. Introduction and Survey 1

PART I: THE FOUNDATIONS:
STRUCTURE AND NMR OF BIOPOLYMERS

CHAPTER 2. NMR of Amino Acid Residues and Mononucleotides 13

CHAPTER 3. NMR Spectra of Proteins and Nucleic Acids in Solution 26

CHAPTER 4. The NMR Assignment Problem in Biopolymers 40

CHAPTER 5. Two-Dimensional NMR with Proteins and Nucleic Acids 44

CHAPTER 6. Nuclear Overhauser Enhancement (NOE) in Biopolymers 93

**PART II: RESONANCE ASSIGNMENTS AND STRUCTURE DETERMINATION IN PROTEINS**

**CHAPTER 7. NOE-Observable $^1H$–$^1H$ Distances in Proteins** 117

**CHAPTER 8. Sequence-Specific Resonance Assignments in Proteins** 130

**CHAPTER 9. Polypeptide Secondary Structures in Proteins by NMR** 162

**CHAPTER 10. Three-Dimensional Protein Structures by NMR** 176

**PART III: RESONANCE ASSIGNMENTS AND STRUCTURE DETERMINATION IN NUCLEIC ACIDS**

**CHAPTER 11. NOE-Observable $^1H$–$^1H$ Distances in Nucleic Acids** 203

**CHAPTER 12. Resonance Assignments in Nucleic Acids Using Scalar Couplings** 224

**CHAPTER 13. Nucleic Acid Conformation, $^1H$–$^1H$ Overhauser Effects, and Sequence-Specific Resonance Assignments** 233

**PART IV: WITH NMR TO BIOPOLYMER CONFORMATION AND BEYOND**

**CHAPTER 14. Conformation of Noncrystalline Proteins and Nucleic Acids** 259

**CHAPTER 15. NMR Studies of Intermolecular Interactions with Biopolymers** 266

**References** 271

**Index** 287

# List of Tables


**2.1.** Nuclear properties of selected isotopes. 14

**2.2.** Structure and $^1$H spin systems of the common amino acids. 15,16

**2.3.,2.4.** Random coil $^1$H chemical shifts of the common amino acid residues. 17,18

**2.5.** Structure and $^1$H spin systems of the common mononucleotides 21

**2.6.** $^1$H chemical shifts of DNA and RNA. 22

**7.1.** Short $^1$H–$^1$H distances in polypeptide secondary structures. 127

**8.1.** Statistics of short $^1$H–$^1$H distances in protein crystal structures. 133

**8.2.** Unique short peptide segments in globular proteins. 133

**8.3.** $^1$H chemical shifts in Tendamistat. 155

**9.1.** $^1$H–$^1$H distances for polypeptide secondary structure identification. 163

**9.2.** Helix identification by short sequential distances $d_{NN}$. 164

**9.3.** $\beta$-sheet identification by short sequential distances $d_{\alpha N}$. 165

**9.4.** Polypeptide secondary structure identification by spin–spin couplings $^3J_{HN\alpha}$. 168

**10.1.** NOE distance constraints in proteins. 177

**10.2.** Amino acid pseudostructures. 178

**10.3.** NOE distance constraints for protein structure determination. 181

**10.4.** Simulated NOE distance constraints derived from the crystal structure of BPTI. 192
**11.1.** Short covalent $^1H$–$^1H$ distances in the common nucleotides. 214
**11.2.** Short sequential $^1H$–$^1H$ distances in DNA. 215
**11.3.** Short sequential $^1H$–$^1H$ distances with labile protons in DNA. 217
**11.4.** Short interstrand $^1H$–$^1H$ distances in DNA duplexes. 218
**11.5.** Mean DNA double helix parameters. 220
**13.1.** $^1H$ chemical shifts in d(CGCGAATTCGCG)$_2$. 244
**13.2.** $^1H$ chemical shifts in d(GCAATTGTGACGG)·d(CCGCTCACAATTCC). 249

# CHAPTER 1
# Introduction and Survey

The first nuclear magnetic resonance (NMR) experiments with biopolymers were described over 30 years ago, and the potentialities of the method for studies of structure and dynamics of proteins and nucleic acids were long anticipated. However, in practice, initial progress was slow because of limitations imposed both by the available instrumentation and by the lack of suitable samples of biological macromolecules. Nuclear magnetic resonance of proteins and nucleic acids is represented in the literature by less than 30 papers during the years up to 1965, approximately 200 papers up to 1970, 4000 papers up to 1980, and of the order 500 to 1000 papers a year from 1981 onward. The work up to 1980 was covered in several textbooks (Dwek, 1973; Govil and Hosur, 1982; James, 1975; Jardetzky and Roberts, 1981; Wüthrich, 1976) and collections of review articles (e.g., Dwek et al., 1977; Opella and Lu, 1979; Shulman, 1979), and most papers published during the early 1980s used experimental procedures that are covered in these texts. This monograph describes a new approach for NMR studies of structure, dynamics, and intermolecular interactions of proteins and nucleic acids, which was developed since approximately 1977. It relies on sequence-specific resonance assignments obtained entirely from NMR experiments and knowledge of the chemical structure. The practicability of this approach was greatly aided by the fact that adaptation of two-dimensional (2D) NMR experiments for studies of biomacromolecules was started very shortly after the initial publications on 2D NMR (Aue et al., 1976a,b; Nagayama et al., 1977).

## 1.1. CONFORMATIONAL STUDIES BY NMR AND BY DIFFRACTION METHODS

Since the pioneering work of Perutz and Kendrew in the late 1950s, single crystal X-ray studies have set standards for protein and nucleic acid spatial structure determination and provided data on over 200 proteins, several tRNA's, and a selection of synthetic oligonucleotides (Richardson, 1981; Saenger, 1984; Schulz and Schirmer, 1979). The NMR approach described in this book is complimentary to X-ray crystallography in different ways:

*1.* The NMR studies use noncrystalline samples (e.g., solutions in aqueous or nonaqueous solvents, or detergent-solubilized biopolymers in mixed micelles). If NMR assignments and spatial structure determination by NMR can be obtained without reference to a corresponding crystal structure, a meaningful comparison of the conformations in single crystals and in noncrystalline states can be obtained.

*2.* Nuclear magnetic resonance can be applied to molecules for which no single crystals are available. Thus new structures can be obtained, which are not available from X-ray studies.

*3.* The solution conditions for NMR studies (e.g., pH, temperature, ionic strength, buffers) can usually be varied over a wide range. This opens the possibility for comparative studies under native and denaturing solution conditions, or for investigations of intermolecular interactions with other solute molecules.

*4.* For a characterization of the internal dynamics of biomacromolecular structures, NMR provides direct, quantitative measurements of the frequencies of certain high activation energy motional processes and at least semiquantitative information on additional high frequency processes. In comparison, X-ray structure determinations may include an outline of the conformation space covered by high frequency structural fluctuations. Furthermore, neutron diffraction in single crystals and NMR in solution can both be employed for studies of the exchange of labile protons, thus potentially enabling direct comparison of the molecular dynamics in the different states.

## 1.2. THE PIVOTAL ROLE OF SEQUENCE-SPECIFIC NMR ASSIGNMENTS

Since approximately 1981 methods have become available for obtaining nearly complete, sequence-specific resonance assignments in biopolymers. $^1$H NMR assignments have been described to date for more than 10 small proteins, a selection of synthetic DNA fragments and, to a more limited extent, for tRNA's and fragments of rRNA and mRNA. Only a fraction of the NMR data are needed for obtaining the resonance assignments, and once

these are available the information content of the remaining data is decisively increased.

As a first illustration we consider the structural analysis of intramolecular nuclear Overhauser effects (NOE) in a biopolymer chain. A NOE between two hydrogen atoms (or groups of hydrogen atoms) is observed if these hydrogens are located at a shorter distance than approximately 5.0 Å from each other in the molecular structure. Typically, a large number of NOE's are observed in globular proteins or in double-helical nucleic acids, indicating that these molecular structures contain numerous pairs of closely spaced hydrogen atoms (upper part of Fig. 1.1). Combined with resonance assignments these distance constraints can be attributed to specified sites along the polymer chain and hence the NOE experiments define the formation of loops by chain segments of variable lengths (Fig. 1.1). With model building or the use of suitable mathematical procedures, notably distance geometry (Blumenthal, 1970), the conformation space can then be searched for spatial arrangements of the polymer chain that are compatible with the experimental distance constraints. Sequence-specific resonance assignments thus provide a basis for systematic procedures toward spatial structure determination of noncrystalline biopolymers (Wüthrich et al., 1982).

As a second example we consider investigations of intermolecular interactions with biopolymers. Nuclear Overhauser effects can manifest short distances between nuclear spins located in the different, interacting molecules, and additional data on close intermolecular contacts in the complexes formed may be obtained from other experiments. Without sequence-specific resonance assignments such data can merely indicate that complex formation has occurred, but when combined with resonance assignments they identify the sites of intermolecular contacts in the polymer chain. As is

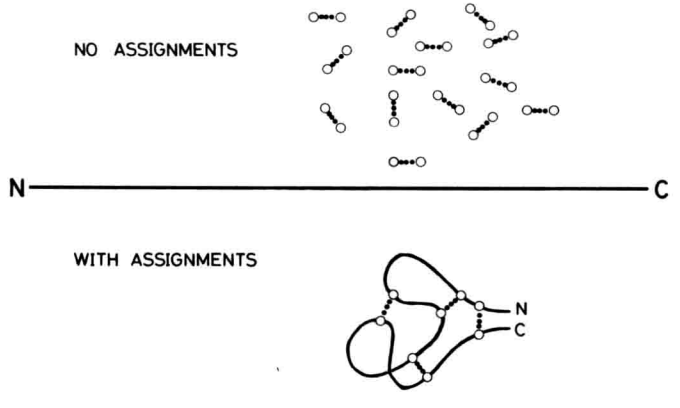


**Figure 1.1.** Information content of $^1H$–$^1H$ NOE's in a polypeptide chain with and without sequence-specific resonance assignments. Open circles represent hydrogen atoms of the polypeptide. The polypeptide chain is represented by the horizontal line in the center.