LBL-37265
UC-408

Lawrence Berkeley Laboratory

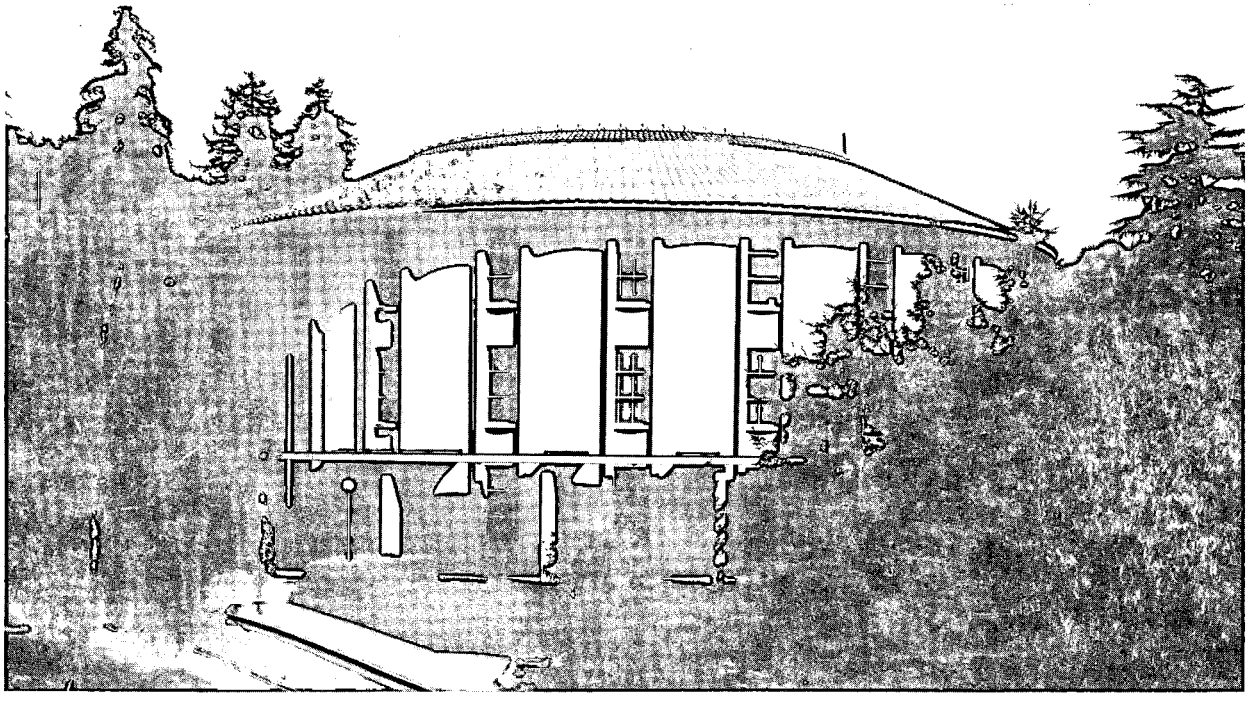
UNIVERSITY OF CALIFORNIA

STRUCTURAL BIOLOGY DIVISION

Solution Structure of Heat Shock Transcription Factor by Multidimensional Heteronuclear NMR

F.F. Damberger
(Ph.D. Thesis)

May 1995



REFERENCE COPY
Does Not
Circulate
Bldg. 50 Library.
LBL-37265
Copy 1

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL-37265
UC-408

Solution Structure of Heat Shock Transcription Factor by Multidimensional Heteronuclear NMR

F.F. Damberger
Ph.D. Thesis

Department of Biophysics
University of California, Berkeley

and

Structural Biology Division
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

May 1995

This work was supported by the Director, Office of Energy Research, Office of Health Effects Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Solution Structure of Heat Shock Transcription Factor by Multidimensional Heteronuclear NMR

Copyright © 1995

by

F.F. Damberger

The U.S. Department of Energy has the right to use this document
for any purpose whatsoever including the right to reproduce
all or any part thereof

LBL-37265
UC-408

Solution Structure of Heat Shock Transcription Factor by Multidimensional Heteronuclear NMR

by

Fred Franz Damberger

B.S. (University of Illinois at Urbana-Champaign) 1987

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor David E. Wemmer, Chair
Professor Alexander Glazer
Professor Nicholas Cozzarelli

1995

This work was supported by the Director, Office of Energy Research, Office of Health Effects Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Abstract

Solution Structure of Heat Shock Transcription Factor by Multidimensional Heteronuclear NMR

by

Fred Franz Damberger

Doctor of Philosophy in Biophysics

University of California at Berkeley

Professor David E. Wemmer, Chair

Heat shock transcription factor (HSF) regulates the heat shock response in eukaryotes and binds to the regulatory elements (heat shock elements) in a novel homotrimeric protein/DNA complex. The DNA-binding domain (DBD) of HSF shows a high degree of sequence conservation among a wide range of species indicative of a highly conserved structural motif for DNA recognition.

Multidimensional heteronuclear NMR experiments were used to determine the complete assignments and solution structure of the HSF DBD from the yeast, *Kluyveromyces lactis* (rmsd for backbone coordinates of 1.03 ± 0.19 Å). The topology of the secondary structure elements (H1-B1-B2-H2-H3-B3-B4) and their arrangement in three dimensions, places HSF DBD in the winged helix class of helix-turn-helix DNA-binding motifs. Helices H2 and H3 which are in the region of highest sequence conservation in the DBD, form the helix-turn-helix pair. Helix H2 contains a completely conserved proline, which creates a kink in the helix axis of H2, and prevented HSF from being identified as a

HTH protein based on sequence alignment. Two loops connect β-strands B1 and B2 (a short Ω-loop), and B3 and B4 (extended loop L1). The extended loop, L1, shows no long range NOEs and no evidence of regular structure. Comparisons to the crystal structure of the DBD from *K. lactis* HSF are made. The differences are modest (1.77 ± 0.20 Å rmsd) and are attributed to crystal packing forces.

The high resolution solution structure of HSF DBD was determined with the addition of new NMR experimental data to obtain stereoassignments and additional NOE constraints. The family of 33 structures (backbone rmsd 0.81 ± 0.16 Å) showed well ordered sidechains which allowed the identification of a highly conserved cluster of hydrophobic residues defining the packing arrangement between helices H2 and H3 which appear to stabilize the kink in helix H2. In addition dynamics data indicate that the extended loop, L1 participates in fast time scale dynamics and is unstructured in solution .

Dedicated to my family

whose love and support

I am eternally grateful for.

Table of Contents

Table of Contents iv

List of Figures viii

List of Tables x

Acknowledgements xi

Chapter 1 NMR Methodology 1

Introduction 1

Conventional Structure Determination of Proteins by 2D ^{1}H NMR 1

- 2D TOCSY 6
- 2D NOESY 8
- The Assignment problem 11
- Secondary Structure Determination 13
- Tertiary Structure Determination 13

Isotopic Labeling 16

2D Heteronuclear Correlation Experiments 18

- Heteronuclear Multiple Quantum Coherence (HMQC) 19
- Heteronuclear Single Quantum Coherence (HSQC) 19
- INEPT Transfer 21
- Constant-Time HSQC 22

Multidimensional Heteronuclear Experiments 23

Three-Dimensional ^{15}N-separated Experiments 23

- 3D ^{13}C HCCH-TOCSY 25

^{13}C,^{15}N, ^{1}H Triple Resonance Experiments 25

- CBCA(CO)NH 26
- Triple-resonance based Assignment 28

Floating Spin Systems 30
Structure Analysis using Multidimensional Heteronuclear Data 31
4D $^{13}C/^{13}C$-edited NOESY-HMQC 32
4D $^{13}C/^{15}N$-edited NOESY-HMQC 33
Generation of Distance Restraints 33
Distance Geometry Methods 33
Simulated Annealing Methods 34
Chapter 2 Medium Resolution Solution Structure of HSF 36
Introduction 36
The Heat Shock Response 36
Regulation of Heat Shock in Eubacteria 37
Regulation of Heat Shock in Eukaryotes 37
Domain Structure of the Heat Shock Factor 39
The Importance of Structural Studies of HSF 41
Materials and Methods 44
The Sequence and Expression of the DNA-binding domain of HSF 44
Sample Preparation 44
NMR Spectroscopy 47
^{1}H-^{1}H NMR 47
Heteronuclear 2D NMR 48
^{15}N-separated 3D NMR 50
Triple Resonance NMR 50
^{13}C-separated 3D experiments 51
4D $^{13}C/^{13}C$-edited NOESY spectrum 52
Chemical shift referencing 53
Structure Calculations 53

Results 54
Resonance Assignments 54
Assignments using Triple Resonance Data 57
Spin system Identification 57
Secondary Structure 60
Stereoassignments and χ_1 angles for Val, Leu, and Ile residues 66
Tertiary Structure 66
Discussion 70
Overall fold and Identification of the helix-turn-helix motif 70
Unusual Structural Features of HSF 73
Hydrophobic Sidechain Packing 74
Proposed DNA Interactions and Relationship with bacterial sigma factors 74
Comparison to other Structural Studies 75
RMSD comparison to X-ray model 76
Backbone NOE comparison to X-ray model 77
Comparison of χ_1 angles to X-ray model 78
Summary of comparison to X-ray model 79
Conclusion 80
Chapter 3 High resolution Solution Structure and Dynamics of HSF 81
Introduction 81
Materials and Methods 82
NMR spectroscopy 82
Stereoassignment experiments for Hβ protons 83
Stereoassignment experiments for Hδ protons 84
4D $^{13}C/^{15}N$ edited NOESY-HMQC experiment 85
Heteronuclear NOE experiment 85

Hydrogen bond assignments 86
Results 86
Determination of stereoassignments 86
Stereoassignment of Hδ protons 89
Analysis of $^{13}C/^{13}C$ and $^{13}C/^{15}N$ edited NOESY-HMQC experiments 92
Hydrogen Bond restraints 92
Structure calculations with new restraints 96
Dynamics from Heteronuclear NOEs 99
Discussion 102
Comparison to crystal structure 102
Comparison to homologous protein structures 103
Role of the Extended Loop and Dynamics 106
Bibliography 108
Appendix A ^{1}H, ^{15}N and ^{13}C resonance assigments for HSF 120
Appendix B Stereoassignment Table 125
Appendix C AMX-600 pulse programs 129
Appendix D Software for analysis 173
Appendix E Restraints for Structure Calculations 261
Appendix F NOEs,<rmsd>s, angular order parameters 293
Appendix G Table of Heteronuclear NOE data 295
Appendix H Table of Amide Exchange data 298

List of Figures

Fig. 1.1. Homonuclear 2D NMR pulse sequences 3

Fig. 1.2 Schematic of a 2D experiment 5

Fig. 1.3 Schematic of Sequential assignment 12

Fig. 1.4. The NOE diagram for Heat Shock Transcription Factor. 14

Fig. 1.5. Pulse sequences for heteronuclear correlation experiments 20

Fig. 1.6. Comparison of a 2D ^{1}H-^{1}H NOESY with a plane from a 3D ^{15}N separated NOESY-HMQC 24

Fig. 1.7 CBCA(CO)NH pulse sequence. 27

Fig. 2.1. Bacterial sigma factors promoter recognition 38

Fig. 2.2. Minimal 3 GAA Box Heat Shock Element 40

Fig. 2.3. HSF domain structure and sequence alignment 42

Fig. 2.4. Sequence alignment of the helix-turn-helix region of HSF with that of the bacterial transcription factor σ^{32} 43

Fig. 2.5. MASS-SPEC of purified HSF samples and the M-28 problem 46

Fig. 2.6. HMQC assignments of HSF 55

Fig. 2.7. Strips from the CBCANH and CBCA(CO)NH of HSF. 56

Fig. 2.8. 3D HCCH-TOCSY planes of HSF. 58

Fig. 2.9. Strips from the 3D ^{15}N NOESY-HMQC of HSF 61

Fig. 2.10. Secondary structure diagram of HSF 62

Fig. 2.11. Schematic diagram of the β-sheet topology of HSF 64

Fig. 2.12. Cα coordinates of the heat shock factor solution structures 67

Fig. 2.13. Secondary structure Cα coordinates for family of HSF structures 68

Fig. 2.14. NOE distribution, angular order parameters and RMSDs of HSF by residue 71

Fig. 3.1. HNHB and DQF-COSY strips used for stereoassignments 87

Fig. 3.2. 2D LRCH spectrum used for stereoassignment of Hδ protons 90

Fig. 3.3. 2D LRCC spectrum used for stereoassignment of Hδ protons..... 91

Fig. 3.4. Backbone superposition of HSF high resolution structures 97

Fig. 3.5. Superposition of ordered sidechain heavy atoms in HSF..... 98

Fig. 3.6. Number of NOEs, rmsds, angular order parameters and $^{15}N\{^{1}H)$ NOEs vs residue number..... 100

Fig. 3.7. Conserved arrangement of a hydrophobic cluster between H2 and H3 105

List of Tables

Table. 2.1. $^3J_{NC}$ and $^3J_{CC}$ Coupling constants and Stereoassignment of methyl protons 65

Table 3.1 Structural statistics for high resolution structures 94

Table 3.2 R.m.s. differences for high resolution structures 95

Table A Resonance Assignments for HSF 120

Table B Stereoassignments data and analysis 125

Table E.1 Pseudoatom corrections to NOEs 262

Table E.2 NOE and H-bond restraints for Chapter 2 268

Table E.3 Dihedral restraints for Chapter 2 274

Table E.4 NOE and H-bond restraints for Chapter 3 278

Table E.5 Dihedral restraints for Chapter 3 287

Table F NOEs,<rmsd>s, angular order parameters for high resolution structures 293

Table G Heteronuclear NOE data 295

Table H Amide Exchange data 298

Acknowledgements*

Thanking those who contributed to this journey and made it rewarding in many ways is the most pleasant task of writing a dissertation. I would like to thank first of all my family, my mother and father, and my three brothers, Simon, Bernhard and Marcus, for their love and support and for encouraging my interest in learning and science. I also thank the many friends I have made during my stay in Berkeley whose company and friendship I will always value. I would particulary like to thank Stephen Rader, whom I have known since I was eight and who has pursued Biophysics graduate studies in parallel with me across the Bay, for his friendship and encouragement.

At the AXE house I made many friends over the years who found ways to divert me from my troubles at times in lab; Melissa Grush, Steve Lonsinger, John Daugherty, Andy Goodwin, Joe Ran, Gaby Carillo, Brian Deaton, Nick and LJ Materer, and especially Cathy Browning, Natalya Eliashberg, and Kimberly Clouser. I thank you all and hope we can continue to be friends over the years.

The Biophysics graduate students who started in the program my year and those who came in the years to come formed a close network of friends and activities which made the first difficult years better than bearable. Among those who made the Biophysics group a rewarding community to be involved in are Jeff and Ann King, Leslie Osborne, Michael O'Neil, Russel Heath, professor Bill Bialek my academic advisor, and of course Dulcynea Wilson, the graduate assistant the year I joined who kept the group functioning with her cheerful and efficient demeanor.

In the Calvin Laboratory ("the round house") I found a very special place in terms of the open community and interaction among so many different research groups. I would like to thank Gary Smith and Jane Coleman for keeping the computer facilities running smoothly in spite of all the "top hogs" in the trough whose company Gary assigned me to frequently. I also thank Phil Eggers and his succesor Vangie Peterson, the building

managers who performed the task of keeping the Calvin lab in good shape through such trials as the Loma Prieta earthquake (1989) and the coming of the Tiger Team. In the Kim group I made many good friends and had intellectual discussions during beer club with the likes of Steve Holbrook, Jay Pandit, Michael Milburn, Fan Jiang, Bill Scott, Milan Tomic, Andrew Bohm, Xiangwei Weng, Detlef Schmidt, Jaru Jancaric, Ursula Schultz-Gahmen and Vladimir. Schulmeister. Crystallography is pretty good too guys.

I would like to thank Professor David Wemmer for providing me with the opportunity to do research under his thoughtful mentorship which has been a truly rewarding educational experience. His patience, insight and knowledge of both NMR and biology have been both inspiring and helpful. Wemmer group members have also made this a special place to do research. I thank group members past and present, Joe Pease, Ann Caviani-Pease, Beth Snowden-Ifft, Vassant Kumar, Debbie Kallack, Mark Kubinec, Marty Pagel, Milton Werner, Valerie Copié , Rick Storrs, Bernhard Geierstanger, Brian Volkman, Pete Spielman, Patty Fagan, Andrew Lee, Tim Cole, Mark Coy, Jonathan Heller, Werner Kremer, Rafael Pelaez, Christian Seel, Qing Zhang, Stephanie Robertson, Mike Nohaile, Marc Ibáñez, Ho Cho, and Corey Liu.

Special thanks go to Andrew Lee for writing macros for handling 4D data within the FELIX program and John Somoza for his feedback on initial versions of this thesis and his friendship over the years. This project would have never seen the light of day without the curiosity and tenacity of David King. It has been a pleasure to know you over the years, DK, and I am excited to be able to help you write a paper on that bugbear M-28. You can expect to be receiving express packages with protein requiring your analytical expertise in the future. I thank Hillary Nelson for pursuing a collaboration on the HSF project with the "other technique" and the members of the Nelson laboratory who have given advise and help on HSF. I particularly thank Celia Harrison for providing samples and incentive to charge ahead with the NMR analysis, as well as Eva Grotttkop for preparing samples on

several occasions. Thanks are also due to other members of the Nelson group for their ideas and input; Ralph Peteranderl (yes Ralph, the trimerization domain is interesting too...), Susan Hubl, Karen Flick, and Otis Littlefield.

I owe a special debt of gratitude to Jeff Pelton who joined the project when I was still in the assignment doldrums and had very little experience with multidimensional heteronuclear NMR. With Jeff's help I was able to quickly learn a great deal about multidimensional heteronuclear NMR-based protein structure determination and we made rapid progress towards the final goal of the project. In addition, I learned from Jeff the enormous power of a simple work strategy; to always focus on what must be accomplished next and then sit down and do it. Finally, I would especially like to thank Kathryn Phillips for her love and encouragement through the final difficult year of my graduate career and for traveling all the way from Scotland to help me complete my dissertation.

May, 1995

* This work was supported by the Director, Office of Energy Research, Office of Health Effects Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Chapter 1 NMR Methodology

Introduction

The atomic structure of a protein allows one to understand the functioning of the molecule in a structural context, to formulate hypotheses for how the protein may interact with other biological molecules, and ultimately how it achieves its function. This type of information is important for developing an understanding of both the normal functioning of biological organisms at the molecular level and for the molecular basis of disease processes. It can ultimately be used to develop pharmaceutical products which interfere with abnormal functioning of such molecules in the case of hereditary diseases or to interfere with the progression of disease due to viruses or parasitic organisms.

Crystallography has developed over the last 50 years as a powerful method for determining the structures of proteins to atomic resolution. The insight gained from inspection of the atomic coordinates of a protein in developing an understanding of their functioning cannot be overemphasized . For example, atomic details of enzyme active sites help to explain substrate specificity, kinetic mechanisms and the effectiveness of inhibitors. The emergence of an alternative technique for structure determination based on Nuclear Magnetic Resonance (NMR) spectroscopy provides information on the structure of biological macromolecules in solution. NMR is sensitive to dynamics and kinetics of molecules which are an essential to their function. This technique has allowed the structure determination of numerous proteins which were not tractable by macromolecular crystallography.

Conventional Structure Determination of Proteins by 2D ^{1}H NMR

Methods developed in the early 1980s (Wüthrich, 1986) allow the determination of the structure of proteins up to approximately 10kD in size. These methods rely exclusively on 2D ^{1}H-^{1}H NMR experiments. In a 2D experiment, magnetization is prepared on a starting nucleus ^1H_a and labeled during an evolution period of incremented length t_1. A

combination of radio frequency pulses and delays (called the mixing period) transfers some of the magnetization to a second nucleus ^1H_b. During the acquisition period the frequency of this second spin **b** is detected. The experiment is run many times with an incremented t_1 value so that the evolution of spin **a** is documented by its modulation of the amplitude or phase of the magnetization of spin **b** at the start of the detection period (Fig. 1.1A). The simplest example of a ^1H-^1H 2D experiment is the COSY (Fig. 1.1B) (Marion & Wüthrich, 1983). This experiment consists of two 90° pulses separated by the incremented delay period t_1 followed by the detection period t_2. In the product operator formalism (Ernst, et al., 1990) magnetization transfer is achieved as follows:

$$^aI_z \quad \text{-}90_y \text{ pulse-}> \quad ^aI_x \qquad \text{eqn. 1.1}$$

Applying chemical shift and J-coupling evolution to this magnetization, we obtain:

$$\text{-chemical shift-}> \quad \underline{^aI_x \cos(\Omega_a t_1)} + {}^aI_y \sin(\Omega_a t_1) \qquad \textit{eqn 1.2a}$$

We follow only the underlined product operator term developed after chemical shift evolution.

-J coupling ->

$$^aI_x \cos(\Omega_a t_1) \cos(J_{ab}\pi t_1) + 2\,^aI_y\,^bI_z \cos(\Omega_a t_1) \sin(J_{ab}\pi t_1) \qquad \textit{eqn. 1.2b}$$

Hence during the evolution period t_1, the J-coupling generates magnetization on spin **a** which is antiphase with respect to spin **b** (the second term in *equation 1.2b*).

Applying the second 90 pulse :

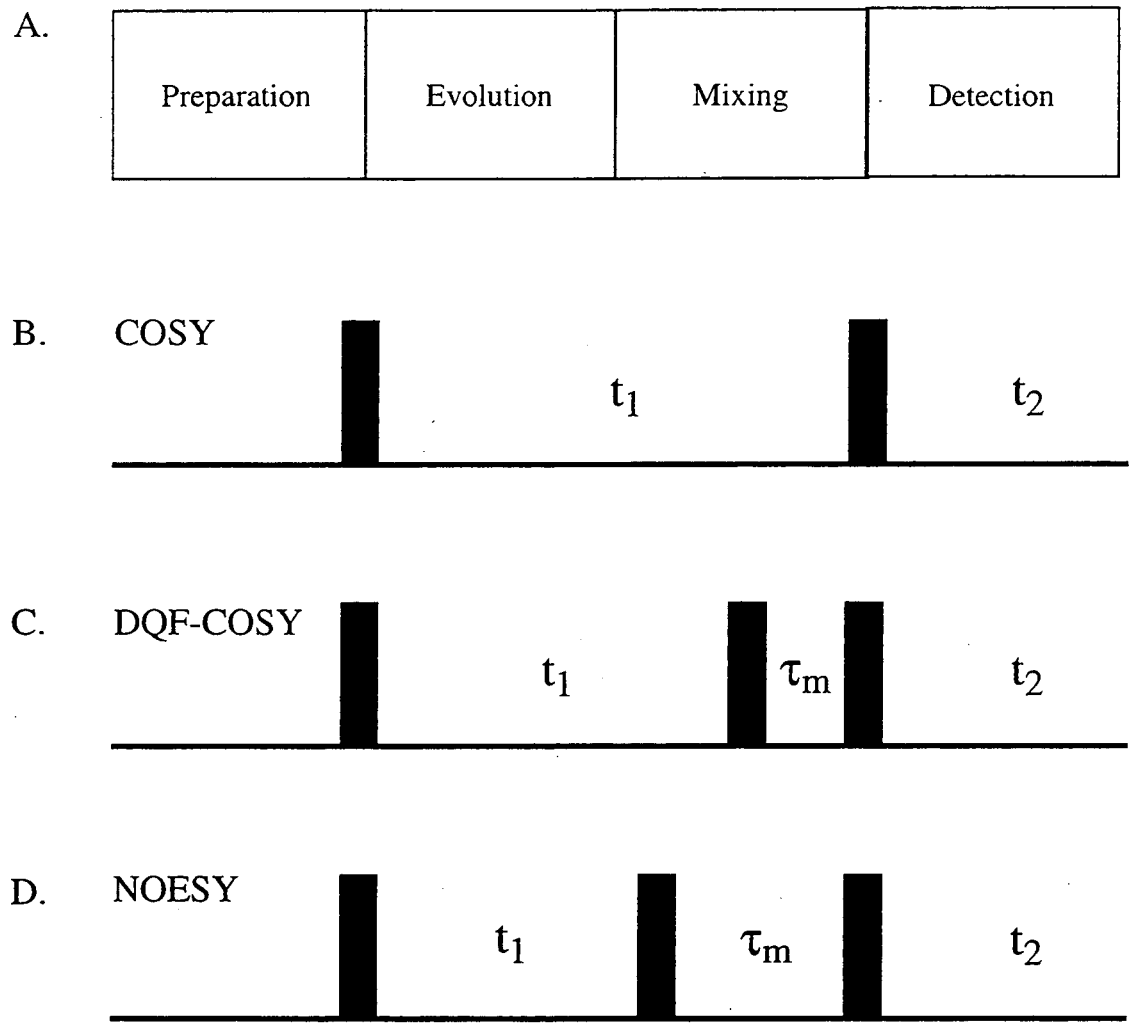


Figure 1.1 **A.** A schematic diagram of a homonuclear 2D experiment, which consists of a preparation period, an evolution period of incremented length t_1, a mixing period, and a detection period t_2. **B.** The pulse sequence for a Correlated Spectroscopy experiment (COSY). **C.** The pulse sequence for a double quantum filtered COSY experiment (DQF-COSY). The delay τ_m is very short (just long enough to allow the phase shift before the second pulse). **D.** The pulse sequence for Nuclear Overhauser Effect Spectroscopy (NOESY). The mixing period ranges from approximately 50 ms up to 450 ms depending on the sample and the purpose of the data (structural restraints, chemical exchange or assignment).

-90_x pulse-> $^aI_x \cos(\Omega_a t_1) \cos(J_{ab}\pi t) - 2^aI_z{}^bI_y \cos(\Omega_a t_1) \sin(J_{ab}\pi t_1)$ *eqn. 1.3*

We convert the second term in *equation 1.2b* from magnetization on spin **a** (antiphase to spin **b**) into magnetization on spin **b** (antiphase to spin **a**). This is the essence of the COSY experiment. Because of the coupling between spins **a** and **b** (J_{ab}) antiphase magnetization evolves during t_1 which is converted by the second pulse from being detectable on spin **a** to detectable on spin **b**. During the detection period t_2 the antiphase magnetization evolves into detectable ^bI_x under the influence of J-coupling evolution. This type of transfer of magnetization is known as a *COSY-type transfer* and is used in many different experiments more sophisticated than the simple COSY experiment.

The first term in *equation 1.3* is magnetization which was not transferred to spin **b** and will be detected at the frequency of spin **a** during the detection period. It appears at the same frequency in both dimensions of the experiment and will give rise to a diagonal peak in the two-dimensional spectrum. The second term will evolve into detectable magnetization on spin **b** during t_2 but was labeled at the frequency of spin **a** during the t_1 period. Hence it will give rise to a crosspeak which correlates two different frequencies (Ω_a in t_1 and Ω_b in t_2) in the COSY spectrum (Fig. 1.2). Because magnetization can also begin on spin **b** at the start of the experiment, a diagonal peak at Ω_b,Ω_b and symmetric crosspeak Ω_b,Ω_a will also be present in the spectrum. In general the COSY spectrum will show correlations between the frequencies of any two spins which have a large enough J-coupling. Thus the COSY family of experiments maps out all the J-couplings between ^1H spins by frequency correlation. Typically J-couplings between ^1H nuclei are detected in a COSY for protons separated by no more than three bonds in a protein under 10kD in size.

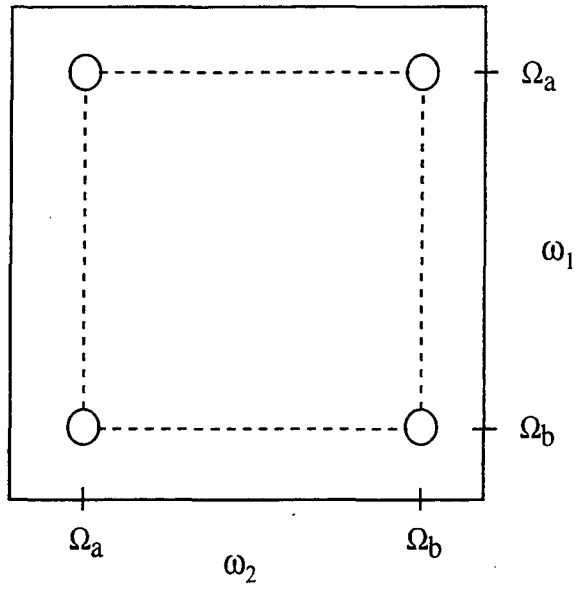


Figure 1.2 Schematic illustration of a 2D experiment. The diagonal crosspeaks correspond to magnetization which did not move from the starting spin in t_1. The crosspeaks correlate the frequencies of the spins which exchange magnetization during the mixing period of the 2D experiment (Ω_1 and Ω_2).

A variation of the COSY experiment which is more commonly employed than the two pulse experiment shown in Fig. 1.1B is the one shown in Fig. 1.1C which is called the double quantum filtered COSY experiment (DQF-COSY) (P.S. #1). [Numbered pulse sequences given in Appendix C are referred to by pulse sequence numbers in the text]. This experiment has the advantage over the conventional COSY experiment that it passes the magnetization developed during t_1 through a double quantum filter selecting only those terms in the magnetization which have evolved antiphase coupling to other spins. The crosspeaks and diagonal peaks have antiphase multiplet structure and both the diagonal and crosspeaks can be phased pure phase absorptive unlike the conventional COSY where the diagonal is dispersive making it difficult to observe crosspeaks close to the diagonal.

Because protons are typically only correlated in a COSY if they are separated by three or fewer bonds, the COSY crosspeaks can only connect protons within a single amino acid. This means that by following the correlations between protons through consecutive steps of three bonds one can map out the topology of a given set of protons within an amino acid to define a spin system (Fig. 1.3A). Examining the structures of the 20 amino acids one finds that there are only 10 discrete classes of spin systems (Wüthrich, 1986). Some spin systems such as AX_3 correspond to one amino acid type (Alanine) where as others such as AMX correspond to many. This identification of spin systems is the first step of solving the assignment problem using the conventional approach (Wüthrich, 1986). A second type of experiment which also probes the protons which are part of a spin-system is the TOCSY.

2D TOCSY

TOCSY stands for Total Correlation Spectroscopy and is also sometimes referred to as HOHAHA. As the name implies, in the TOCSY experiment (P.S. #2) all the protons in a given spin system are correlated with each other (Braunschweiler & Ernst, 1983; Bax

& Davis, 1985). This has the advantage over the COSY, that all the members of a spin system can be identified by locating crosspeaks to the frequency of the spin corresponding to the least crowded region of the spectrum. Often this pattern of frequencies is enough to identify the spin system. For example Valine typically shows correlations from the H_N to the αH (~3 ppm), βH (1-2 ppm), and two intense γH methyl resonances (0.4 - -1.0 ppm) (Fig. 1.3). In addition to identifying spin systems based on chemical shift pattern, the TOCSY spectrum also forms a convenient reference point for looking for correlations in the COSY spectrum which identify which spins are separated by three bonds in the spin system. The product operator description for the TOCSY experiment follows:

$$^{a}I_z \text{ -90y pulse-> } \quad ^{a}I_x \qquad \textit{eqn. 1.4}$$

Applying chemical shift and J-coupling evolution to this magnetization, we obtain:

$$\text{-chemical shift-> } \quad \underline{^{a}I_x \cos(\Omega_a t_1)} + {}^{a}I_y \sin(\Omega_a t_1) \qquad \textit{eqn. 1.5}$$

-J coupling ->

$$\underline{^{a}I_x \cos(\Omega_a t_1) \cos(J_{ab}\pi t_1)} + 2\,{}^{a}I_y\,{}^{b}I_z \cos(\Omega_a t_1) \sin(J_{ab}\pi t_1) \qquad \textit{eqn. 1.6}$$

After labeling the spins with their chemical shift frequencies in t_1, the spin-lock pulse sequence is applied. This spin-lock pulse sequence is optimized to prevent chemical shift evolution from occurring and results in the spins effectively evolving under an isotropic strong coupling Hamiltonian:

$$\mathbf{H} = h\,J\ (\,{}^{a}I_x\,{}^{b}I_x + {}^{a}I_y\,{}^{b}I_y + {}^{a}I_z\,{}^{b}I_z\,) \qquad \textit{eqn. 1.7}$$

Under this Hamiltonian, the first term in *equation 1.6* evolves as follows:

$$^{a}I_x \rightarrow 1/2\ ^{a}I_x\ (1 + \cos(2\pi Jt)) + 1/2\ ^{b}I_x\ (1 - \cos(2\pi Jt)) + (^{a}I_x\,^{b}I_z - ^{a}I_z\,^{b}I_y) \sin(2\pi Jt)$$

eqn. 1.8

Hence, $^{a}I_x$ is converted directly into $^{b}I_x$. With a larger number of spins in the coupling network, magnetization can be converted directly from in-phase magnetization on one spin in the network into in-phase magnetization on any other spin in the network. The rate at which magnetization evolves from one spin to another depends in a complicated way on the network of coupling constants connecting the two spins (Cavanagh, et al., 1990). Qualitatively, spins which are separated by smaller numbers of bonds pass magnetization to one another more quickly than those separated by large numbers of bonds. This means that longer mixing time will result in more remotely connected spins (for example H_N and H_δ from leucine) showing correlations to one another in a TOCSY. The length of the mixing time is limited by relaxation to equilibrium which occurs in the rotating frame ($T_{1\rho}$ relaxation) and reduces the signal to noise of the resulting crosspeaks. Typically mixing times in medium sized proteins (10kD) are no longer than 100 to 120 ms. In addition to the remotely coupled spins showing correlations in a TOCSY experiment, the other feature of the TOCSY experiment which is desirable is that the crosspeaks are all in-phase absorptive. Contrast this to the DQF-COSY experiment where the crosspeaks are anti-phase. Anti-phase peaks have the disadvantage that as line-widths become comparable in size to the coupling constants, the two opposite sign halves of the multiplet can cancel one another resulting in weak signal to noise or complete disappearance of peaks with smaller couplings (Rance, et al., 1983).

2D NOESY

Nuclear Overhauser Effect Spectroscopy (NOESY) (P.S. #3) is the structural workhorse in the NMR spectroscopist's set of tools (Neuhaus & Williamson, 1989). In

this experiment frequencies are correlated which correspond to spins which are close in space. In proteins of typical size investigated by NMR, close is defined as less than 5 Å. Although this is a rather crude and limited yardstick, it turns out to be enough to define the structures of proteins very accurately. This is because proteins are linear polymers which fold up into precisely defined compact globular structures. Such structures give rise to the close approach of many protons in the core of the protein. If many of these close approaches can be measured using the NOESY experiment they can be turned into distance constraints which, when combined with the constraints due to the well understood covalent geometry of the protein, are sufficient to define the three dimensional fold and packing of the protein down to less than an Ångstrom resolution (Wagner, et al., 1987; Powers, et al., 1993).

The NOESY experiment (Fig. 1.2D) starts with a preparation 90° pulse followed by a t_1 evolution period. This is followed by a 90° pulse, a delay τ_m (the mixing period) and a third 90° pulse followed by the detection period. During the mixing period, magnetization which was labeled with the frequency of one spin **a** during t_1, is transferred to a second spin **b** close in space, which is detected during t_2. The transfer of magnetization during the mixing period is accomplished by the nuclear Overhauser effect (NOE). The NOE is a perturbation in the population levels of one spin induced by a fluctuating magnetic field caused by a second spin close in space (Neuhaus and Williamson, 1989).

The product operator description of the NOESY follows:

$$^{a}I_z \xrightarrow{-90_y \text{ pulse}} {}^{a}I_x \qquad \textit{eqn. 1.9}$$

Applying chemical shift evolution to this magnetization, we obtain:

-chemical shift-> $^{a}I_x \cos(\Omega_a t_1) + {}^{a}I_y \sin(\Omega_a t_1)$ *eqn. 1.10*

To simplify matters we ignore scalar coupling in this case.

The second 90° pulse produces

-90y pulse-> $-{}^{a}I_z \cos(\Omega_a t_1)$ *eqn. 1.11*

This is longitudinal magnetization on spin **a** labeled with the frequency of spin **a**.

During the ensuing mixing period, the NOE causes some of this magnetization to be transferred to other spins which are close in space (such as spin **b**).

-NOE-> $-{}^{b}I_z \cos(\Omega_a t_1)$ *eqn. 1.12*

The longitudinal magnetization on spin **b** is converted into detectable magnetization with the third 90° pulse:

-90y pulse-> $-{}^{b}I_x \cos(\Omega_a t_1)$ *eqn. 1.13*

This magnetization evolves at the frequency of spin **b** during the detection period t_2:

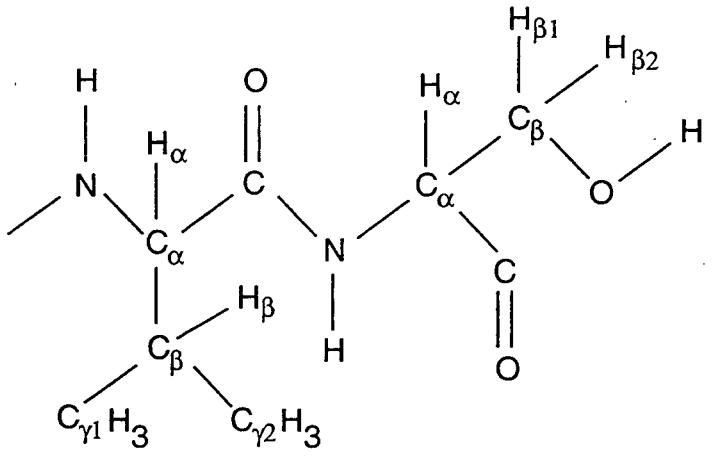
$-{}^{b}I_x \cos(\Omega_a t_1) \cos(\Omega_b t_2)$ *eqn. 1.14*

Thus the nuclear Overhauser effect gives rise to crosspeaks (Ω_a, Ω_b) between any two spins close in space. For a more detailed description of the NOE see Neuhaus and Williamson (1989).

The Assignment problem

The initial stage of NMR analysis of a protein is known as the assignment problem. The goal is to assign a frequency to every proton in the molecule. This facilitates the structural analysis since correlations between frequencies in experiments such as the NOESY then define spatial proximity of the corresponding protons. The first stage of the assignment problem is to identify all the spin-systems using the COSY and TOCSY experiments as described above. The spin systems are then connected to one another sequentially using NOESY data (fig. 1.3). In general the NOESY spectrum contains crosspeaks corresponding to pairs of protons which are close in space (< 5.0 Å). Some of these crosspeaks correlate the frequencies of protons which are in adjacent residues in the amino acid sequence. These cross peaks are referred to as sequential NOEs. Sequential NOE correlations between protons in two spin systems imply that the spin systems might correspond to adjacent amino acids in the sequence of the protein. The sequential NOEs usually relied upon to identify neighboring residues involve backbone protons and are referred to by the names of the protons involved: $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$. These connectivities are the primary source of sequential assignments because all residue pairs have αH, βH and NH protons (except Proline and Glycine which do not contain HN and βH protons respectively), and one of the distances $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ (often two) are close enough in all the possible protein backbone and $\chi 1$ angle conformations to give rise to reasonably strong NOE (Wüthrich, 1986). Once sufficiently long strings of spin systems are identified they can be placed uniquely in the sequence since a particular succession of spin systems occurs only once in the sequence. Finally all the fragments of polypeptides can be connected forming the complete protein sequence thus assigning all the proton frequencies in the protein (Wüthrich, 1986) .

A.



Val : A_3B_3MX Ser : AMX

B.

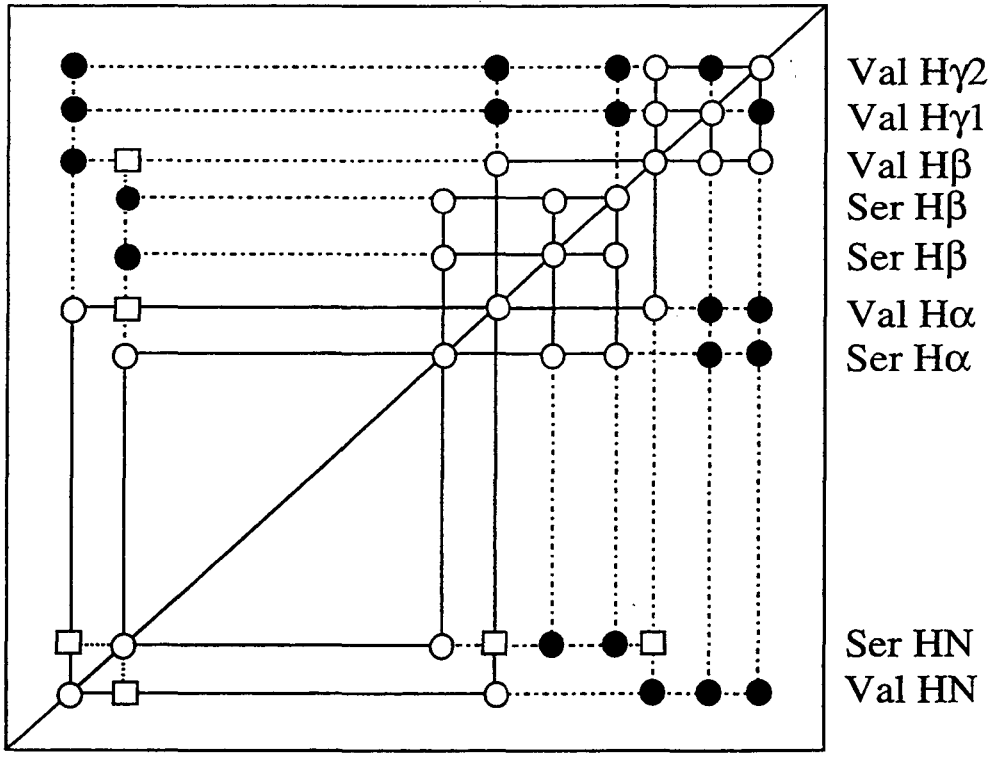


Figure 1.3 **A.** Dipeptide fragment Val-Ser along with the spin system types A_3B_3MX and AMX. **B.** Schematic diagram superimposing the crosspeaks which would ideally by observed for three 2D experiments: COSY; ○, TOCSY; ●, and NOESY; □. Note that TOCSY crosspeaks that duplicate COSY peaks are marked as COSY peaks. The connectivity of the two spin systems due to COSY is indicated by solid lines. Dashed lines show TOCSY connectivity and sequential NOEs.

Secondary Structure Determination

Once the assignments have been determined by assigning the spin systems using the COSY and TOCSY, and connecting them up using sequential NOEs from the NOESY, the next step is to determine the secondary structure of the protein. This is accomplished by examining the intensities of different classes of sequential NOEs and looking for NOEs separated by more than one residue. These are classified as "medium range" (between 2 and 5 residues separating the two protons involved), and "long range" (greater than 5 residues separating the two protons involved). The two types of regular secondary structure present in proteins (helices and beta sheets) have different patterns of sequential NOE intensities and differing medium and long range NOEs (Wüthrich, 1986). In helices, sequential NOEs between amide protons and β protons ($d_{\beta N}$) are medium to strong in intensity as are d_{NN} NOEs whereas $d_{\alpha N}$ NOEs are weak. In addition helices frequently show medium range NOEs between the H_α of one residue and the H_N of a residue three or four positions further in the sequence ($d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ respectively). Beta sheets, on the other hand have strong $d_{\alpha N}$ NOEs and show long range NOEs connecting the αH and H_N of adjacent strands (often indicated by $d_{\alpha N}(i,j)$). These patterns are a direct consequence of the repeating structure of the backbone of the protein which result in the close approach of the aforementioned protons. This type of analysis defines the location of secondary structural elements in the protein and is usually represented in a schematic diagram known as the NOE diagram (Fig. 1.4).

Tertiary Structure Determination

Once the NOESY experiment has been analyzed to identify secondary structure, the remaining NOEs can be interpreted in terms of the overall fold of the protein, since most of the remaining NOEs are long range and define the spatial arrangement of the secondary structural elements. All the NOEs can be converted into constraints defining the close

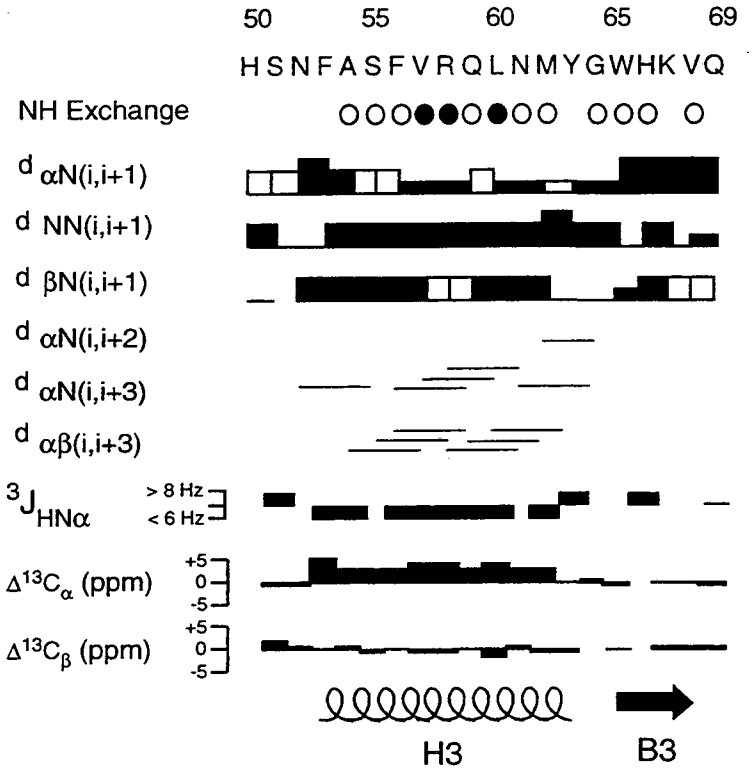


Figure 1.4 NOE diagram of the heat shock transcription factor helix 3 and beta strand 3. A schematic representation of the NMR data identifying two elements of secondary structure in the DNA-binding domain of the heat shock transcription factor from *Kluyveromyces lactis.* The data appearing by column are: amide proton chemical exchange rates; O indicates backbone H-N signals still present after 6 hours and ● indicates H-N signals still present 2 weeks after dissolution in D_2O. The next three rows show sequential NOE intensities $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ with strong medium and weak intensities indicated by small medium and large boxes. Open boxes show the maximum possible intensity in cases of overlap where only an upper limit could be placed on the NOE intensity. The next three columns show medium range NOEs which were observed that are indicative of helical secondary structure. The next row shows H_N-H_α coupling constant data qualitatively with boxes below the line indicating coupling constants under 6 Hz and boxes above the line indicating coupling constants above 8 Hz. The next two columns show deviations from random coil carbon chemical shifts (Spera & Bax, 1991) for the C_α and C_β chemical shifts. Finally secondary structure is indicated along the bottom of the diagram; helices by , beta sheet by .

approach of the protons involved. Typically, the NOEs are classified into three categories according to intensity (strong, medium, and weak), and constraints are defined with distance cutoffs appropriate to the NOE intensity (Appendix E) (Clore, et al., 1986; Kraulis, et al., 1989). For example, a strong NOE might have a upper bound for the distance between the two protons involved set at 2.4Å, and a weak NOE would have an upper bound of 5Å. The lower bound is normally set to the sum of the Van der Waals radii of the two protons (1.8Å). An algorithm is used to search for structures which are consistent with the distance constraints as well as consistent with known covalent geometry of polypeptides as observed in small molecule crystallography and high resolution protein crystal structures (Clore, et al., 1986; Wagner, et al., 1987; Nilges, et al., 1988). The result is a family of structures consistent with all that is known in general about protein structures as well as consistent with all of the distance constraints derived from analysis of the NOESY spectra.

The "conventional" approach described above for assignment and structure determination is viable for proteins up to about 10 kD in size. For larger proteins, the number of resonances becomes large enough so that a significant portion of the crosspeaks will overlap making interpretation of 2D ^1H-^1H spectra for assignment and structure determination difficult if not impossible. A second problem with larger proteins is that as the size of the molecule increases, the correlation time due to rotational diffusion increases and so relaxation becomes more rapid and lifetimes for spin states become shorter (Ernst, et al., 1990). This results in broader lines and further compounds the overlap problem in 2D spectra in addition to reducing signal to noise.

The size limitation of this "conventional approach" has spurred the growth of several new NMR techniques which make it possible to determine structures of proteins larger in size (up to 20 - 30 kD). This has expanded the range of problems which can be investigated using NMR to include many interesting biological systems such as DNA

recognition by transcription factors (Zhang, et al., 1994), signal transduction (Pelton, et al., 1991; Archer, et al., 1993), oligomerization domains and unfolded states in proteins (which typically have very poor signal dispersion) (Neri, et al., 1992).

Isotopic Labeling

Most of the recent advances in NMR methodology (late 80s and early 90s) have been drawn from the ability to label proteins with spin 1/2 isotopes such as ^{15}N and ^{13}C. This can been accomplished by incorporation of specific labels at interesting sites, or through uniform enrichment of the protein (McIntosh & Dahlquist, 1990). Proteins are made up primarily of hydrogen, carbon, nitrogen, and oxygen. Fortunately three of these elements have stable spin one half isotopes (^{1}H, ^{13}C, and ^{15}N). Hydrogen is 99.9 % ^{1}H (spin 1/2). Carbon and Nitrogen are approximately 1% spin 1/2 isotope at natural abundance. To increase the sensitivity of NMR experiments making use of these "heteronuclei" (^{13}C and ^{15}N), proteins must be isotopically enriched. This is usually accomplished by altering the growth media of bacteria expressing the protein.

If selective labeling is desired, rich media containing all the necessary amino acids can be used , with one or several amino acids replaced by isotope enriched amino acids. This will result in incorporation of the labeled amino acids anywhere they occur in the protein sequence. Amino acids are available which are uniformly enriched at all sites for a given nucleus, or specifically enriched at certain sites. For example one can buy valine with ^{13}C enrichment at the $C_{1\gamma}$ position to aid in assigning the two methyl groups stereospecifically.

Of more general use is the strategy of uniformly labeling all the sites in a protein. This is especially cost-effective for ^{15}N. One can buy 5 g of 99% ^{15}N enriched NH_4Cl for a few hundred dollars. This amount is sufficient to prepare enough ^{15}N labeled protein for an NMR sample by substituting the ^{15}N enriched ammonium chloride as the sole nitrogen

source in a minimal growth medium of bacteria which express the protein under the control of a strong promoter. Typically one can obtain tens of milligrams of NMR pure (>95%) protein with a 1-2 liter growth which, for a 11kD protein, is enough for one or two NMR samples. Carbon-13 labeling is usually combined with ^{15}N labeling to obtain "double" labeled protein. Such doubly labeled protein expands the range of possible experiments enormously and has been the most significant factor in obtaining complete proton assignments (including sidechains) and medium to high resolution structures of proteins over 15 kD (Powers, et al., 1993). Carbon-13 is a more expensive nucleus to label with (1 g of uniformly ^{13}C enriched glucose costs about 500 dollars) and uniform ^{13}C labeling is undertaken only when protein expression and sample conditions have been extensively optimized.

Why bother to isotope label? The gyromagnetic ratio for ^{13}C and ^{15}N are 1/4 and 1/10 of that for the proton respectively, making these nuclei much less sensitive for detection. The power of isotope labeling comes from several factors:

1) The heteronuclei resonate at distinct frequencies and can be manipulated independently using radio frequency (rf) coils tuned to resonance frequencies of the heteronuclei on separate rf channels;

2) The heteronuclei have large chemical shift ranges which can be used to spread out resonances due to different spins, reducing overlap problems;

3) Most heteronuclei are attached to one or more protons and the coupling constants to the attached protons are very large (on the order of 100 Hz) making transfer to the heteronucleus and back very efficient since it can occur very quickly before relaxation has much effect (in a few milliseconds);

4) These factors can be used to take advantage of the good chemical shift dispersion properties of heteronuclei without sacrificing sensitivity since one can use the larger

polarization of protons as a starting point and transfer to and from heteronuclei during evolution periods for frequency labeling by the heteronucleus;

5) Coupling constants between directly attached heteronuclei are also quite large (12-40 Hz) making it possible to transfer magnetization quickly and efficiently through the protein skeleton;

6) Scalar couplings between heteronuclei and other heteronuclei or protons separated by three bonds obey Karplus relations (Karplus, 1959; Bystrov, 1976) making them sensitive to the local dihedral angle and a useful supplement to three bond proton-proton couplings for defining dihedral angle and stereoassignments of stereopairs of protons (Zuiderweg, et al., 1985; Archer, et al., 1991; Clore, et al., 1991).

2D Heteronuclear Correlation Experiments

The simplest experiments using an isotope labeled sample are the heteronuclear correlation experiments. As might be expected, these experiments correlate the chemical shift of the heteronucleus labeled during the evolution period with the chemical shift of the directly attached proton during the detection period. In these experiments it is common to start with magnetization on the protons (taking advantage of the large proton gyromagnetic ratio which predicts a relatively large population difference between the spin up and spin down states). The magnetization is promptly transferred to the attached heteronucleus using either the INEPT trick or the HMQC sequence (see next section). Magnetization is allowed to evolve at the heteronuclear frequency during the evolution period (t_1). It is then transferred back to the proton prior to detection of the proton frequency in t_2 (Fig. 1.5). Note that both the initial excitation and final detection are accomplished using the proton so that the experiment does not sacrifice any sensitivity due to the inferior gamma for the heteronucleus. This 2D experiment is actually more sensitive than ^1H-^1H 2D experiments since the transfer efficiency which generates the crosspeaks of interest is nearly 100%.

Thus it is possible to perform a 2D heteronuclear correlation experiment with protein samples with concentrations as low as 0.3 mM.

Heteronuclear Multiple Quantum Coherence (HMQC)

There are two classes of heteronuclear correlation experiment which derive their name from the state of the magnetization that exists during the evolution period used to label the heteronucleus frequency. The Heteronuclear Multiple Quantum Coherence (HMQC) experiment (P.S. #5) generates a multiple-quantum coherence between the heteronucleus and the attached proton which is allowed to evolve during t_1 (Mueller, 1979) . This coherence evolves at the sum and difference frequencies of the two nuclei (X and ^1H). To eliminate the evolution due to the ^1H during t_1, a 180° pulse occurs in the center of the t_1 period on the proton channel. This refocuses the proton evolution at the end of t_1 and leaves the X nucleus evolution unchanged. Hence the magnetization is labeled at the frequency of only the heteronucleus during the t_1 period. The multiple-quantum coherence is subsequently transferred back into transverse magnetization on the proton for the detection period t_2. Coupling to the heteronucleus during t_2 is eliminated using decoupling pulse sequence on the X-channel.

Heteronuclear Single Quantum Coherence (HSQC)

The second type of heteronuclear correlation experiment is called the Heteronuclear Single Quantum Coherence experiment (HSQC) (P.S. #6) and as its name implies it generates single quantum coherence on the heteronucleus to label the heteronuclear frequency during t_1(Bodenhausen & Ruben, 1980). This is achieved using a sequence called the INEPT. This stands for Incredible Natural Abundance Experiment for Polarization Transfer and has more to do with the origin of the experiment than how it functions (Morris & Freeman, 1979). Because the experiment generates single quantum coherence on the X-nucleus, it does not require a 180° pulse to ensure evolution at the X-nucleus frequency (Fig. 1.5B). However, to eliminate splittings due to the ^1H-X coupling

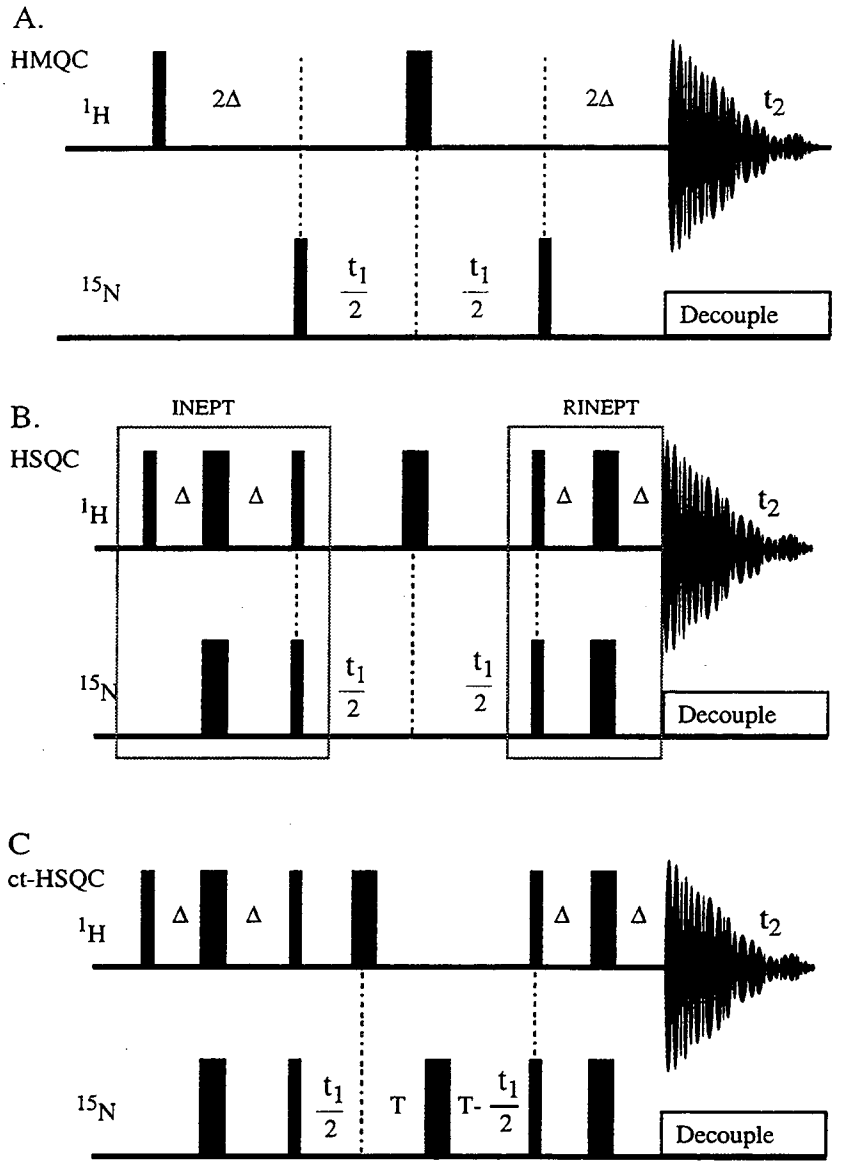


Figure 1.5 Heteronuclear Correlation Pulse sequences. Narrow boxes indicate high power 90° pulses. Wide boxes indicate high power 180° pulses. Incremented time periods are indicated by lower case t and refocussing delays by Δ (1/4J). All pulse sequences show the proton pulses in the upper row and ^{15}N pulses (or other heteronucleus) in the lower row. Broad-band decoupling sequences are usually applied during acquisition to remove the large one bond couplings to the heteronucleus. **A.** Heteronuclear Multiquantum Coherence (HMQC) experiment. **B.** Heteronuclear Single Quantum Coherence (HSQC) experiment. INEPT and reverse INEPT pulse sequences described further in the text are highlighted by grey boxes. **C.** constant-time Heteronuclear Single Quantum Coherence (ct-HSQC) experiment. The chemical shift of the heteronucleus is labeled by incrementing the 180° ^{1}H and ^{15}N pulses through the constant time period, 2T.

in t_1, the 180° pulse at the center of the t_1 period is still present, in this case refocusing evolution due to J-coupling with the attached proton.

INEPT Transfer

Because the INEPT transfer (Fig. 1.5B) is so commonly used in heteronuclear experiments, it warrants a detailed product operator description:

The first pulse of the ^1H channel generates transverse magnetization on a proton:

I_z -90y pulse-> I_x *eqn. 1.15*

The next part of the pulse sequence consists of a delay τ (1/2J), a pair of 180° pulses on both ^1H and X channels and another delay τ. Because of the pair of 180° pulses at the center of this evolution period the chemical shift evolution of the nuclei is refocussed . On the other hand the J-coupling between the ^1H and X nuclei is unaffected:

$-\tau$ -180° (^1H)-

-180°(X) $-\tau$ –> $I_x \cos(J\pi 2\tau) + 2I_y S_z \sin(J\pi 2\tau)$ *eqn. 1.16*

Substituting in the value 1/2J for τ, the first term of *equation 1.16* becomes zero and the second term becomes: $2I_y S_z \sin(\pi/2) = 2I_y S_z$ *eqn. 1.17*

Finally the simultaneous application of ^1H and X $90°_x$ pulses converts this magnetization on the proton which is antiphase with respect to the X nucleus into:

-90°x(^1H)-

-90°x (X) -> $-2I_z S_y$ *eqn. 1.18*

which is magnetization on the X nucleus antiphase with respect to the attached proton. This will evolve into in-phase magnetization on the X-nucleus which is single quantum coherence evolving at the frequency of the X-nucleus. Following t_1 evolution to label the heteronuclear frequency, the INEPT pulse sequence is reversed (reverse INEPT) followed by detection of the proton frequency.

Constant-Time HSQC

There are several tricks which can be used to eliminate couplings. Two we have already discussed are the refocussing 180° pulse at the center of an evolution period, and decoupling pulse sequences. A third method is the constant time (ct) period (Vuister & Bax, 1992). In this method, a time evolution period is not incremented but set to a fixed time period adjusted to a length which refocuses the coupling of interest. The chemical shift evolution is detected by placing a 180° pulse at the start of the evolution period (on the channel corresponding to the nucleus whose chemical shift is being labeled) and incrementing its position to later and later times (t_1) in the constant time period until it has marched to the other end of the constant time period (2T) (Fig. 1.5C).
Because the 180° pulse refocuses chemical shift evolution for a period of evolution equivalent to the delay since the start of the constant time period, the net evolution period for chemical shift evolution is different for each incremented position. This trick can be used in a constant time ^1H-^{13}C-HSQC (ct-HSQC) (P.S. # 12) to eliminate couplings between neighboring ^{13}C nuclei during t_1 evolution in a uniformly ^{13}C labeled protein.

If the ct period is chosen to be equal to $1/2J_{CC}$ then carbon nuclei will have resonances whose sign depends on whether they have an even or odd number of ^{13}C neighbors. This information can be useful in identifying resonances according to amino acid type. For example alanine β-methyl carbons have only one carbon neighbor (α-carbon) whereas isoleucine γ2-methylene has two neighbors (β-methine and δ-methyl), hence their signals will have opposite sign in a ct-HSQC tuned to $1/2J_{CC}$.

There are other advantages to the constant time trick. For example, during a constant time interval, several manipulations of the magnetization can occur at once. At the same time that the chemical shift of the relevant spin is labeled, magnetization can be transferred to another spin using either COSY-type or INEPT type magnetization transfer,

provided that the appropriate pulses occur on the channels corresponding to the frequencies of the spins of interest and the total evolution time of the chosen coupling results in complete magnetization transfer.

Multidimensional Heteronuclear Experiments

Isotope labeling methods would not have had nearly so dramatic an impact on structure determination by NMR if it were not for the combination of heteronuclei-based techniques with multidimensional NMR. The real power of labeling a protein becomes realized when heteronuclear correlation experiments are concatenated to normal homonuclear ^{1}H-^{1}H experiments, or when such experiments are combined with the ability to correlate three or even four chemical shifts at once.

Three-Dimensional ^{15}N-separated Experiments

The simplest examples of such experiments are probably the three-dimensional ^{15}N-separated experiments. One starts with an ordinary homonuclear 2D ^{1}H-^{1}H experiment such as a NOESY (Fig. 1.1D). After the NOESY mixing period an INEPT transfer is inserted which transfers magnetization from ^{15}N attached protons to the directly attached ^{15}N nucleus. This is followed by a second evolution period t_2 which labels the ^{15}N nucleus of the destination proton with the frequency of its directly attached ^{15}N. Then a reverse INEPT is applied, transferring the magnetization back to the proton for detection during t_3. Thus the experiment correlates the proton frequency of the starting proton with the ^{15}N and ^{1}H frequencies of the destination proton and its directly attached ^{15}N nucleus respectively. The heteronuclear part of the experiment spreads the information on the destination nucleus out into an extra dimension which is well resolved (namely the chemical shift of the ^{15}N nucleus attached to the destination proton).

One can imagine this experiment as a concatenation of the NOESY experiment (Fig. 1.1D) with the HMQC or HSQC experiment (1.5AB) and these two versions of the experiment are referred to as the 3D ^{15}N-separated NOESY-HMQC (Zuiderweg & Fesik,

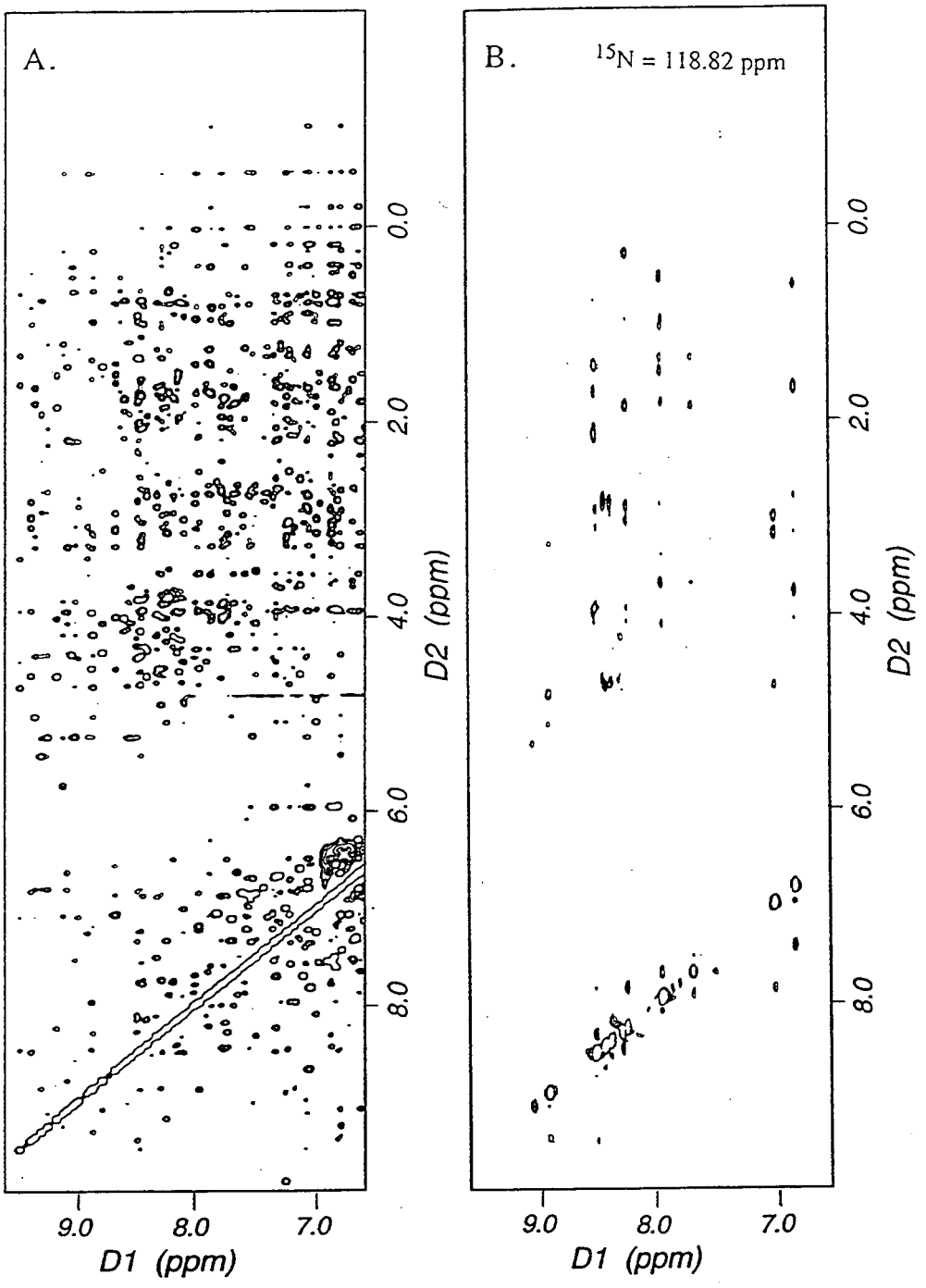


Figure 1.6 Comparison of a 2D 1H-1H NOESY **A.** with a plane from a 3D ^{15}N-separated NOESY-HMQC (^{15}N frequency = 118.92 ppm) **B.** Both spectra were collected on a sample of heat shock transcription factor DNA-binding domain with 100 ms mixing times and 14.12 ppm spectral widths in ^{1}H dimensions. Spectral width in the ^{15}N dimension of the 3D ^{15}N -NOESY-HMQC was 60.0 ppm.

1989; Marion, et al., 1989c) (P.S. #8) and 3D ^{15}N-separated NOESY-HSQC (Bodenhausen and Ruben, 1980) respectively. Fig. 1.6 is a comparison of a normal 2D ^{1}H-^{1}H NOESY spectrum with one ^{15}N plane from a 3D ^{15}N-separated NOESY-HMQC spectrum. It illustrates the degree to which overlap problems can be alleviated by using multidimensional heteronuclear experiments.

3D ^{13}C HCCH-TOCSY

In addition to concatenation, one can make use of the same magnetization transfer methods employed during the mixing periods of homonuclear 2D experiments to transfer magnetization between heteronuclei. A simple example of this is the 3D ^{13}C HCCH-TOCSY (Fesik, et al., 1990) (P.S. #14). In this experiment, magnetization generated on a proton is frequency labeled and then transferred to the attached ^{13}C nucleus. The ^{13}C frequency is labeled, and then the magnetization is transferred to all the other ^{13}C nuclei in the spin system using an isotropic mixing sequence employed on the ^{13}C channel. Finally the magnetization is transferred to the attached proton from the destination ^{13}C for detection. This version of the TOCSY takes advantage of the large one-bond couplings between the ^{13}C nuclei (~35 Hz in aliphatic residues) to rapidly and efficiently transfer magnetization between all the ^{13}C nuclei in the spin system. It also separates the ^{1}H-^{1}H planes according to chemical shift of the starting ^{13}C nucleus.

^{13}C,^{15}N, ^{1}H Triple Resonance Experiments

Using the idea of concatenation along with the ability to transfer magnetization between heteronuclei with the relatively large heteronuclear coupling constants, it becomes possible to move magnetization between sites on the protein backbone very precisely. This gives rise to a class of experiments which facilitate the assignment of the protein backbone without use of the NOE. To achieve this, one must correlate chemical shifts of nuclei in adjacent amino acids. Magnetization can be moved from one amino acid to the next by making use of one-bond and two-bond heteronuclear couplings between heteronuclei on

adjacent amino acids. Since such experiments require the ability to apply pulses on ^{1}H, ^{13}C and ^{15}N channels, they are referred to as triple resonance experiments (Ikura, et al., 1990). A powerful example of this class of experiments is the CBCA(CO)NH (Grzesiek & Bax, 1992b).

CBCA(CO)NH

The CBCA(CO)NH experiment (P.S. #10) correlates the C_β and C_α frequencies of a residue to the ^{15}N and ^{1}H of the following residue. This is achieved through the use of the large one bond couplings between ^{13}C and ^{1}H nuclei (~180 Hz), $^{13}C\alpha$ and $^{13}C'$ (~12 Hz), $^{13}C'$ and ^{15}N (~15 Hz), and ^{15}N and ^{1}H (~90 Hz). Because all the magnetization transfers are through large one bond coupling constants, the experiment is very sensitive and does not depend on the local backbone conformation unlike the NOESY experiment which is used to obtain sequential assignments in the conventional method as described earlier in this chapter.

When the data from the CBCA(CO)NH experiment are combined with analysis of the CBCANH experiment (Fig. 1.7A) (P.S. # 9) which correlates the amide ^{15}N and ^{1}H chemical shifts of a residue with its C_β and C_α chemical shifts, and the C_β and C_α chemical shift of the previous residue, assignments of the backbone resonances of intermediate size proteins can quickly be completed. Fig. 1.7B shows the pulse sequence of the CBCA(CO)NH and the transfer of magnetization is described below:

Magnetization is generated on the $H\beta$ and $H\alpha$ nuclei. This magnetization is INEPT transferred to the attached C_β and C_α nuclei respectively (position marked **a** in Fig. 1.7B) which are then frequency labeled during the constant time interval by incrementation of the refocussing 180° $C\beta/C\alpha$ pulse in t_1. The constant time interval is adjusted to be of length ($1/4J_{CC}$) so that at the end of the this period, $1/\sqrt{2}$ of the magnetization initially on C_β is now on C_α due to a COSY-type transfer between the ^{13}C nuclei, and $1/\sqrt{2}$ of the magnetization initially on C_α remains of C_α (**b**). The second function of the constant time

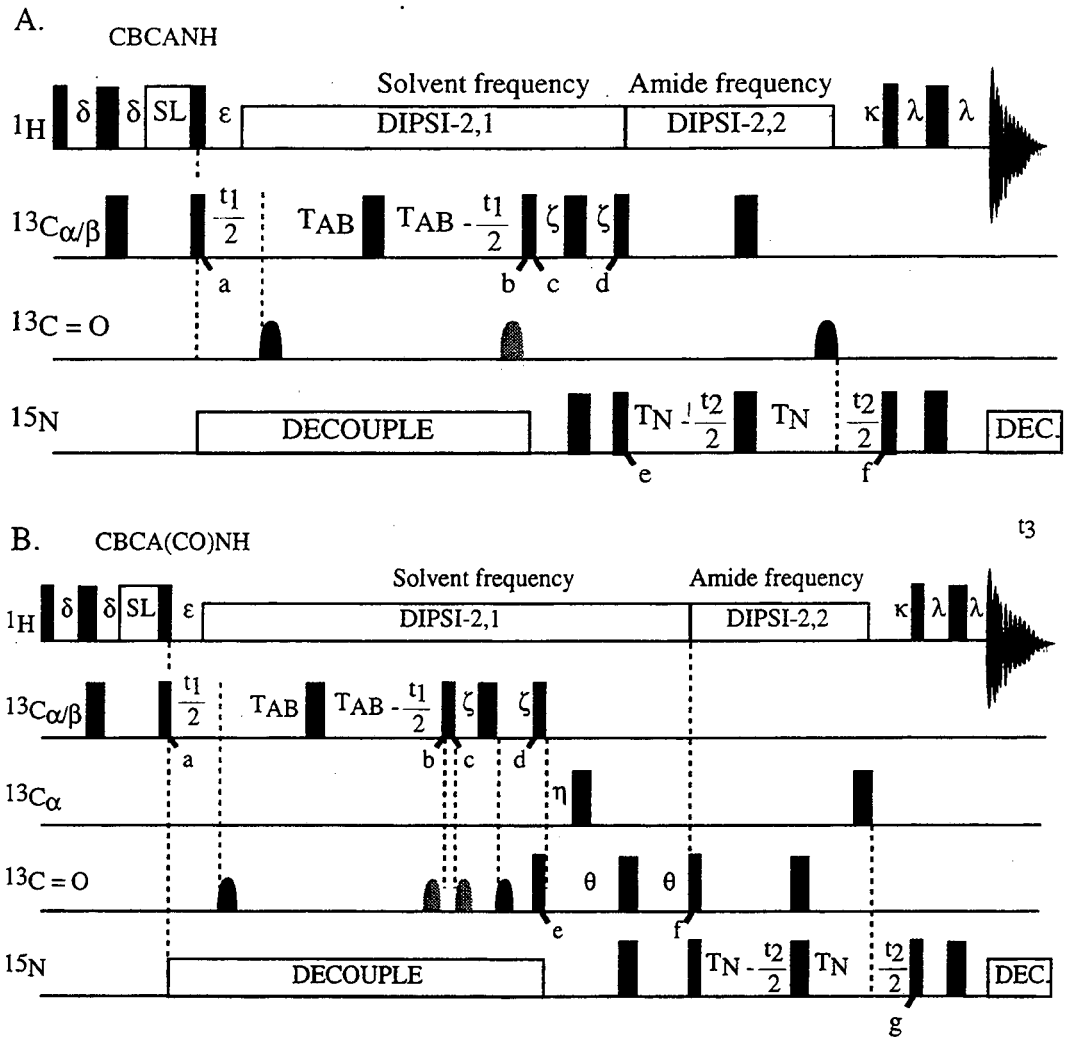


Figure 1.7 Triple resonance backbone assignment experiments. **A.** CBCANH and **B.** CBCA(CO)NH. 90° and 180° pulses are indicated by narrow and wide rectangles. Soft 180° pulses applied on carbonyl ^{13}C are indicated by rounded pulses. Grey soft pulses indicate Bloch Siegert compensation pulses as described in the text. Rectangular carbonyl pulses are applied with an RF field strength of 4.7 kHz resulting in a null in excitation for the C_α resonances. C_α/C_β pulses have an excitation null for the Carbonyl resonances. Both experiments begin with an INEPT transfer to C_α and C_β, **a**, followed by evolution of antiphase terms between C_α and C_β (and ^{13}C frequency labeling), **b**. A 90° pulse achieves COSY type transfer to C_α, **c**. At this point the two experiments diverge with CBCANH applying INEPT transfer to ^{15}N and the CBCA(CO)NH to carbonyl ^{13}C, **d**,**e**. The CBCA(CO)NH continues with a second INEPT to ^{15}N, **f**, and an INEPT to H_N, **g**, for detection. Both experiments frequency label ^{15}N and detect H_N . Delays are as follows: δ=1.5 ms, ε=2.1 ms (A) and 2.3 ms (B), T_{AB}=3.3 ms, ζ=11 ms (A) and 3.7 ms (B) T_N=11.2 ms (A) and 11.1 ms (B), κ=5.4 ms, λ =2.25 ms, η=4.5 ms, θ=11.4 ms.

period is simply to transfer all relevant magnetization to the Cα nucleus (**c**) preparing for the next step of the experiment which is the transfer to the C' nucleus (**d**,**e**) and from there to the ^{15}N nucleus (**f**). This magnetization is labeled with the ^{15}N frequency in t_2 before reverse INEPT transfer (**g**) to the amide ^{1}H for detection in t_3.

The pulse sequence is complicated by the need for selective 180° refocussing pulses on different nuclei to prevent evolution of couplings which would interfere with the magnetization transfer scheme. For example, due to the selective carbonyl ^{13}C 180° pulses during the constant time period which labels the Cβ/Cα nuclei, the coupling between these aliphatic nuclei and the carbonyl carbons does not evolve. Moreover the pulse sequence is written in such a way as to result in zero phase correction in the indirectly detected dimensions. To achieve this, no net chemical shift evolution can occur in the first t_1, t_2 increment. One source of potential phase changes is the presence of the carbonyl ^{13}C soft pulses during the Cβ/Cα evolution. Even though these pulses occur at frequencies different from the resonance frequencies of the Cβ/Cα nuclei, they can cause phase changes due to off-resonance effects (referred to as the Bloch-Siegert effect) (Freeman, 1988b). This evolution can be refocussed by matching the evolution under such a pulse before a refocussing 180° with a second "ghost" pulse after the refocussing 180°. This explains the presence of the second carbonyl ^{13}C pulse at the end of the Cβ/Cα evolution. Such "ghost" pulses are referred to as "Bloch-Siegert compensation" pulses.

Triple-resonance based Assignment

The correlations in the CBCANH experiment (Grzesiek & Bax, 1992a) are of two types: those which correlate C_β and C_α within a residue with its own N-H pair (Intraresidue), and those which correlate C_β and C_α of the previous residue in the sequence with the same N-H pair (Sequential). In principle one can assign the backbone of a protein from only the CBCANH experiment. Since the sequential crosspeaks tend to be weaker than the intraresidue crosspeaks (the magnetization is transferred to the neighbor residue

through the weak two-bond $^2J_{N_iC\alpha_{i-1}}$ coupling ~7 Hz), one can "walk" through the sequence residue by residue by matching Cβ/Cα resonance frequencies from intraresidue and sequential crosspeaks. Unfortunately the $C\alpha_{i-1}$ -> N_i transfer is the weak link in the experiment and often one or both sequential Cβ/Cα correlations are missing. Although this experiment may be sufficient for some intermediate sized proteins, those with significant aggregation or somewhat larger molecular weights will have incomplete CBCANH spectra. Consequently the fragments of connected amino acids which can be assigned are too short to connect them all together and sequence specifically assign the backbone of the protein (Wüthrich, 1986).

A more robust strategy is to combine the information in the CBCANH with that in the CBCA(CO)NH. The CBCA(CO)NH experiment shows only sequential correlations, positively identifying the sequential Cβ/Cα pair in the CBCANH. Because the CBCA(CO)NH exclusively makes use of single bond couplings for magnetization transfer, the experiment has very high sensitivity and forms a powerful complement to the CBCANH since nearly all the sequential correlations appear in the CBCA(CO)NH spectrum. Combined analysis of these two spectra is a reliable approach to assigning large stretches of the backbone to sequential fragments.

A method has been proposed for translating these fragments into sequence specific assignments by making use of correlations between Cβ/Cα chemical shift and amino acid type (Grzesiek & Bax, 1993). Another approach is to use the N-H pair as a starting point for identifying the proton chemical shifts in the spin system by matching the N-H chemical shifts to appropriate plane and H_N frequency (N-H strip) in a 3D ^{15}N-separated TOCSY-HMQC. The TOCSY correlations should include the $H\alpha$ and $H\beta$ chemical shifts. Next one can look for correlations further out on the sidechain by looking at strips corresponding to the $C\beta$-$H\beta$ pair and $C\alpha$-$H\alpha$ pair in an 3D HCCH-TOCSY. Often the sidechain proton chemical shift pattern is enough to make a good guess at the spin-system type (or even the

amino acid type). The sidechain proton chemical shift pattern can be used to extend the carbon assignments of the spin-systems to the remainder of the sidechain.

Because carbon chemical shifts correlate fairly well with chemical environment in the amino acid (Wishart, et al., 1991) one can predict the chemical shift range of the sidechain carbons fairly accurately. Then one can look for matching proton chemical shifts in appropriate carbon planes. Hence at each stage the assignments are extended by making use of an "anchor" proton-heteronucleus pair which directs which strip is searched for characteristic ^{1}H chemical shifts; then these characteristic chemical shifts are used as a search pattern to extend the assignments to other strips thus extending the assignments to a new "anchor pair". This process culminates in fairly complete assignments of the ^{1}H and X nuclei for the amino acids in each fragment. During the assignment extension, the amino acid type is identified. Once a sufficient number of amino acids are identified in a fragment it can be matched to a unique location in the known amino acid sequence for the protein, thus sequence-specifically assigning all the resonances in that fragment. Another consequence of assignment extension of the CBCANH/CBCA(CO)NH fragments using 3D ^{15}N-separated TOCSY-HMQC and 3D HCCH-TOCSY data is that $C\beta/C\alpha$ correlations missing in triple resonance data are identified in the TOCSY experiments allowing sequential fragments to be extended by searching for new $C\beta/C\alpha$ matches. Eventually all the fragments are connected to one another and this completes the process of sequence-specific resonance assignments.

Floating Spin Systems

There are a few resonances which cannot be sequentially assigned with the above analysis. These are sidechain spin systems which are not connected to the backbone by observable correlations in the TOCSY and COSY experiments. Examples are the εCH_3 of methionine, the δNH_2 and γNH_2 of Asn and Gln residues respectively, and the aromatic spin-systems of Tyr, Phe, His, and Trp. The Met, Asn, and Gln spin-systems do not show

correlations to the backbone because protons in these spin-systems are removed by more than three bonds from the spins which do show correlations to the backbone in the TOCSY and COSY experiments. They do not show correlations in the HCCH TOCSY because they are separated from the backbone by non-carbon nuclei. The aromatic spin systems are also separated by more than three bonds from protons which show correlations to the backbone. So ^{1}H-^{1}H scalar coupling cannot generate correlations to the backbone protons. The aromatic carbon nuclei resonate very far upfield (at 180 ppm) of the aliphatic carbons (10-70 ppm), consequently the homonuclear decoupling sequence used for isotropic mixing of the aliphatic carbon nuclei in the HCCH-TOCSY cannot be applied at high enough power to include the aromatic carbon nuclei. As a result, there are no correlations between the aromatic spin-systems and the backbone resonances in these experiments either.

To assign these "floating" spin systems one must make use of through-space correlations of the NOESY. The H_δ protons in all the aromatic amino acids usually show an intense correlation to at least one proton in the following group: H_N, H_α, H_β. Thus one can make use of H_δ-H_N correlations in the 3D ^{15}N-separated NOESY-HMQC or H_δ-H_α and H_δ-H_β correlations in the 3D ^{13}C-separated NOESY-HMQC (or 4D 13C/13C-edited NOESY-HMQC: see below) to assign the H_δ proton and C_δ carbon. The rest of the aromatic spin system is connected to the backbone through these NOEs to the H_δ.

Structure Analysis using Multidimensional Heteronuclear Data

Once the assignments have been obtained through analysis of the triple resonance backbone assignment experiments coupled with 3D HCCH-TOCSY, 3D ^{15}N separated TOCSY-HMQC, and COSY experiments which allow side chain assignments to be determined, the next step is structure determination. In the conventional approach this is already partially realized once assignments are obtained since sequential and medium range NOE intensities have already been determined and their intensities are diagnostic of secondary structure. Using multidimensional heteronuclear methods, one relies on 3D and

4D heteronucleus separated NOESY experiments to define secondary and tertiary structure. A common starting point is the ^{15}N-separated NOESY-HMQC experiment. Potential degeneracies in the amide proton chemical shifts are removed by correlating their chemical shift with the chemical shift of the attached ^{15}N nucleus in a third dimension. This results in reliable determination of NOE intensitities diagnostic of secondary structure type. For example, strong d_{NN} and $d_{\beta N}$ connectivities and $d_{\alpha N}(i,i+3)$ connectivities are diagnostic of helical backbone geometry where as strong $d_{\alpha N}$ and cross strand $d_{\alpha N}(i,j)$ and $d_{NN}(i,j)$ connectivities are indicative of beta sheet geometry. Sequential assignment and the associated secondary structure determination are represented by N-H strips corresponding to adjacent amino acids in the DNA-binding domain of heat shock transcription factor in Fig. 2.9).

4D $^{13}C/^{13}C$-edited NOESY-HMQC

Although the ^{15}N-separated NOESY-HMQC is a good starting point for structural analysis, for definition of the global fold, it is often necessary to define the position of side chains with respect to one another, particularly in alpha helical proteins since most of the contacts between elements of secondary structure are between protons in the sidechains. The most effective experiment for probing such sidechain-sidechain contacts is undoubtedly the 4D $^{13}C/^{13}C$-edited NOESY-HMQC (Clore, et al., 1991) (P.S. #15). This experiment identifies NOEs between protons attached to ^{13}C nuclei by correlating the proton frequencies with the frequencies of the attached ^{13}C nuclei. This is achieved by inserting an HMQC sequence before the proton evolution period in a regular NOESY experiment (Fig. 1.1D) and after the mixing time to ^{13}C frequency label both protons involved in the NOE. Correlation of the ^{1}H resonances to the frequencies of the attached ^{13}C nuclei results in a further reduction in overlap over the related 3D ^{13}C-separated

NOESY-HMQC experiment and is particularly useful in proteins with extensive ^{1}H degeneracy.

4D $^{13}C/^{15}N$-edited NOESY-HMQC

A second 4D NOESY experiment (4D $^{13}C/^{15}N$-edited NOESY-HMQC) (P.S. #17) uses the HMQC sequence to correlate the two protons involved in an NOE with the attached ^{13}C and ^{15}N nuclei respectively (Kay, et al., 1990). This experiment is useful in eliminating overlap of ^{13}C-attached protons which show NOEs to the same ^{15}N-attached protons. This is frequently a problem in helical regions were H_{α}-degeneracy prevents assignment of $d_{\alpha N}(i,i+1)$, $d_{\alpha N}(i,i+2)$, $d_{\alpha N}(i,i+3)$, and $d_{\alpha N}(i,i+4)$ NOEs. It can also be useful in eliminating degeneracy of aromatic protons showing NOEs to H_N nuclei, a problem encountered in proteins with a large number of aromatic sidechains defining the core of the protein (Chapter 2 and 3).

Generation of Distance Restraints

Once the crosspeaks in the multidimensional NOESY experiments have been assigned by frequency, they can be converted into distance constraints (Appendix E) which are used to define a family of structures which satisfy the constraints using one of several algorithms implemented by programs such as INSIGHT and X-PLOR.

Distance Geometry Methods

Distance geometry is a method of converting a list of pair-wise distance constraints between parts of a molecule into a consistent three dimensional coordinate set for the atoms in the set (Crippen & Havel, 1978; Havel & Wüthrich, 1984; Havel & Wüthrich, 1985; Sippl & Sheraga, 1986). Qualitatively, the method works as follows. The set of pair-wise distance constraints corresponds to an N x N dimensional space representation of the molecule. In such a space, all the pair-wise distance constraints can be satisfied using the following algorithm. Each distance constraint places an upper and lower bound on the distance between the two atoms involved. This can be represented by a number in the top

triangle of an N x N matrix where one reads the names of all the atoms along the rows and columns of the matrix. This set of ordered numbers is referred to as the upper bounds matrix. Similarly, one can use the lower triangle to represent all the lower bounds on the distance between each pair of atoms.

The next step is to generate a random number representing a first guess for the distance between each pair of atoms involved in a distance constraint with the number falling between the upper and lower bounds. The numbers are then adjusted to reflect the triangle inequality so that the distance $d_{ab} + d_{bc} \geq d_{ac}$. Higher order triangle inequalities are also forced to be satisfied. The N x N dimensional representation is then projected (in the linear algebra sense) into a three dimensional space using a least squares solution which reduces the number of violations of the original constraints (Crippen and Havel, 1978). This is the distance geometry solution to the input constraints.

Simulated Annealing Methods

Distance geometry solutions have a number of problems which have lead to the development of a second type of structure refinement known as simulate annealing . Simulated annealing uses a dynamics algorithm to simulate the motion of the atoms of the protein structure under energy constraints defined by two classes of functions (Clore, et al., 1985; Brünger, 1992). The first class is a set of energy functions which define normal local covalent protein geometry such as bond lengths, bond angles, peptide and aromatic planarity, and chirality. The second class of functions define an energy which represents the degree to which the structures satisfy experimental distance constraints determined from the NOE data and dihedral angle constraints.

$$E_{total} = E_{geometry} + E_{experimental} \qquad \textit{eqn 1.19}$$

The derivative of the energy function is the force field and this force field can be used for short time steps to calculate the trajectory of the atoms under the influence of this energy potential. Thus the motion of the atoms of the protein can be simulated under this combined force field. By use of an annealing protocol one can search the conformational space efficiently to find global minima with respect to the total energy function E_{total} (often referred to as the target function). Given sufficient input constraints, these minima should correspond to atomic coordinates which satisfy (within acceptable error margins) the experimentally derived distance constraints and the normally observed protein geometry represented by the $E_{geometry}$ energy term. Thus a family of solutions which are consistent with the NMR constraints can be determined reliably. In the annealing protocol, the temperature is raised during the dynamics simulation to search large regions of conformational space and then a slow cooling is applied to "zoom" in on regions of the energy function with low energy. Eventually the structures converge on a local (and hopefully global) minimum to the target function.

Although the Simulated Annealing algorithm can begin from random starting coordinates, this can lead to a small fraction of converged structures (with low E_{total}) because the protein gets stuck in local minima during cooling. Another strategy commonly employed is to use the distance geometry solution as a starting point for simulated annealing. This approach is referred to as the hybrid distance geometry/simulated annealing algorithm (DG/SA) and takes advantage of the speed of the DG algorithm to find starting coordinates in the neighborhood of the global minimum, and the efficiency of the SA algorithm to find global minima (Clore, et al., 1986; Brünger, 1992).

Chapter 2 Medium Resolution Solution Structure of HSF

Introduction

The Heat Shock Response

When an organism encounters elevated temperatures, a rapid and drastic shift in protein expression occurs resulting in the production of a new class of proteins known as the heat shock proteins. This response occurs in response to environmental shocks such as elevated temperatures, ethanol, heavy metals, or virus invasion (Morimoto, et al., 1992). This response is universal occurring in all cells from bacteria to plants to mammals including humans. The proteins expressed under heat shock conditions, known as heat shock proteins (HSPs), are necessary for the cell to survive the shock conditions. These proteins, some of which are also referred to as molecular chaperones, are highly conserved across organisms from prokaryotes to eukaryotes, with sequence conservations in some cases as high as 50%. The proteins function to "chaperone" the unfolded proteins through the shock conditions and return them to an active folded conformation when the stressful conditions have passed. They are also involved in transport and translocation of synthesized proteins in the cell (Craig, et al., 1993). Recently a number of studies have shown that the HSP, GroEL from *E. coli*, forms oligomeric complexes with denatured protein and that release and refolding of the bound protein is controlled by the binding of a second HSP, GroES, which enhances ATP hydrolysis by GroEL (Gray & Fersht, 1991; Bochkareva, et al., 1992). It is likely that many of the HSPs expressed in eukaryotes serve similar functions although with the diversity of heat shock gene families, other functions important in helping cells survive stress conditions cannot be ruled out. One of the fascinating aspects of the Heat Shock Response, or more generally the stress response, is that it is induced rapidly (within minutes) by such a diverse range of stimuli. The profile of HSP gene expression varies depending on the stimulus as well. These observations have

led to careful study of the transcriptional regulation of the heat shock response to shed light on the nature of the transduction of stress stimuli into gene expression and the differential control of HSP gene expression.

Regulation of Heat Shock in Eubacteria

The heat shock response is regulated differently in eubacteria than in eukaryotes. In bacteria, the sudden shift in protein expression is accomplished through the alternate sigma factor σ^{32} which binds to the apo form of the RNA polymerase in place of the standard sigma factor σ^{70}. The sigma factor, σ^{32}, redirects the holoenzyme to bind exclusively to promoters of heat shock genes. Sigma factors confer specificity for two regions of the promoter known as the -35 region and -10 region respectively. Consensus sequences for both regions have been determined for heat shock genes and "normal" genes and differ significantly (Fig. 2.1A). Although when isolated, σ^{70} does not bind to these regions, it can be cleaved into fragments which bind the -35 region and -10 region respectively (fig2.1B) (Dombroski, et al., 1992). The fragment of σ^{70} which binds to the -35 region contains a sequence predicted to form a helix-turn-helix structure which is a common motif for DNA recognition (Harrison & Aggarwal, 1990). This helix-turn-helix motif is also predicted to occur in other bacterial sigma factors including σ^{32} (Gardella, et al., 1989).

Regulation of Heat Shock in Eukaryotes

In eukaryotes, heat shock is regulated by the Heat Shock Transcription Factor (HSF) (for recent reviews see (Craig, et al., 1993; Lis & Wu, 1993; Morimoto, 1993)). Under heat shock conditions in higher eukaryotes this factor has been shown to trimerize and move into the nucleus binding to promoter elements upstream of the heat shock genes. In lower eukaryotes (yeasts), HSF is always in the trimeric state and bound to DNA. Yeast requires low level constitutive expression of heat shock proteins for growth (Jakobsen &

A

Promoter	-35 region		-10 region	+1
groE	TTTCCCCCTTGAA	GGGGCGAAGCCAT	CCCCATTTCTCTGGTCAC	
dnaK 1	TCTCCCCCTTGAT	GACGTGGTTTACGA	CCCCATTTAGTAG	TCAA
dnaK 2	TTGGGCAGTTGAA	ACCAGACGTTTCG	CCCCTATTACAGACTCAC	
C62.5	GCTCTCGCTTGAA	ATTATTCTCCCTTGT	CCCCATCTCTCCCACATC	
Consensus	-35 region		-10 region	
σ^{32}	T--C-C-CTTGAA		CCCCAT-T	
σ^{70}		TTGACA	TATAAT	

B

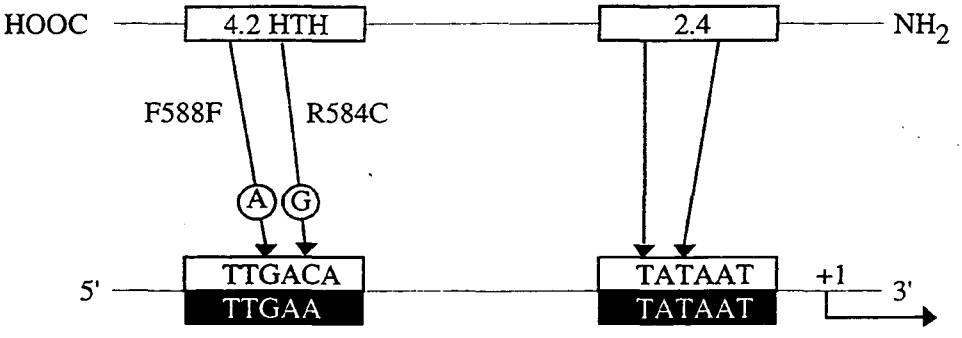


Figure 2.1 Sequences for heat shock promoters in *E. coli* and domain structure of σ^{70}. **A**. Alignment of promoter sequences from several heat shock genes in *E. coli* and the consensus sequences for heat shock gene promoters and the "standard" promoter. The transcription start site is marked +1. Note that the last three bases GAA in the -35 region are invariant in all the sequences examined and that they differ from the consensus sequence for the -35 region of normal genes (ending in GACA). (Taken from Table 16-1 of Watson et al., 1987). **B**. Domain arrangement for the region of σ^{70} which is involved in recognizing the -35 and -10 promoter sites of *E coli* genes. The DNA sequences recognized by the separated domains appear in a parallel arrangement below. The region labeled 4.2 HTH contains a helix-turn-helix sequence which has also been identified in σ^{32}. The HTH sequence of σ^{32} shows limited homology to the HTH region of HSF (fig 2.5). In Figure 2.1, the corresponding consensus sequences for σ^{70} and σ^{32} are indicated above and below the line representing the DNA respectively. Contacts between σ^{70} and the promoter DNA inferred from mutagenesis studies which create new promoter specificity are indicated by arrows with the new base that is recognized indicated inside a circle. The amino acid change of the mutations are indicated by one-letter amino acid code next to the arrows (Dombroski et al., 1992).

Pelham, 1988; Sorger & Pelham, 1988; Wiederrecht, et al., 1988; Park & Craig, 1989; Gross, et al., 1990; Chen & Pederson, 1993; Gallo, et al., 1993) . In both higher and lower eukaryotes, heat shock conditions induce a transient increase in production of heat shock proteins which continues if a high level of heat shock conditions are maintained (Nieto-Sotelo, et al., 1990; Sorger, 1990; Jakobsen & Pelham, 1991; Gallo, et al., 1993).

The promoter elements are known as heat shock elements (HSEs) and consist of multiple repeats of the sequence nGAAn (known as GAA boxes) occurring in head to tail orientation. The smallest HSE consists of three GAA boxes (Fig. 2.2A) and HSF binds to such a site as a homotrimer with approximately 10^{-12} M affinity. Typically heat shock gene promoters have more than three GAA boxes and HSF exhibits cooperativity in binding to adjacent HSEs (Xiao, et al., 1991). This is a hint at how the heat shock response can take over protein expression with such high efficiency.

Domain Structure of the Heat Shock Factor

Deletion analysis of the gene from HSF reveals at least four functional domains (Fig. 2.3A). There are two domains which affect transcriptional activation and two highly conserved domains involved in DNA-binding and trimerizarion respectively. Indeed, comparison of these last two regions reveals a high degree of sequence conservation in organisms as diverse as yeast, fruit flies and tomatoes (Fig. 2.3B). This high degree of sequence conservation emphasizes the universal importance of the heat shock factor and implies that what is learned from studying HSF from one organism will have general implications for the functioning of heat shock factors from all other organisms.

The trimerization region contains a leucine zipper sequence and forms a triply coiled coil of three parallel helices placing all three DNA-binding domains at the N-termini of the coiled coil (Peteranderl & Nelson, 1992). This suggests a mode of DNA-binding in which the trimeric symmetry of the HSF homotrimer is broken by interactions with DNA. An

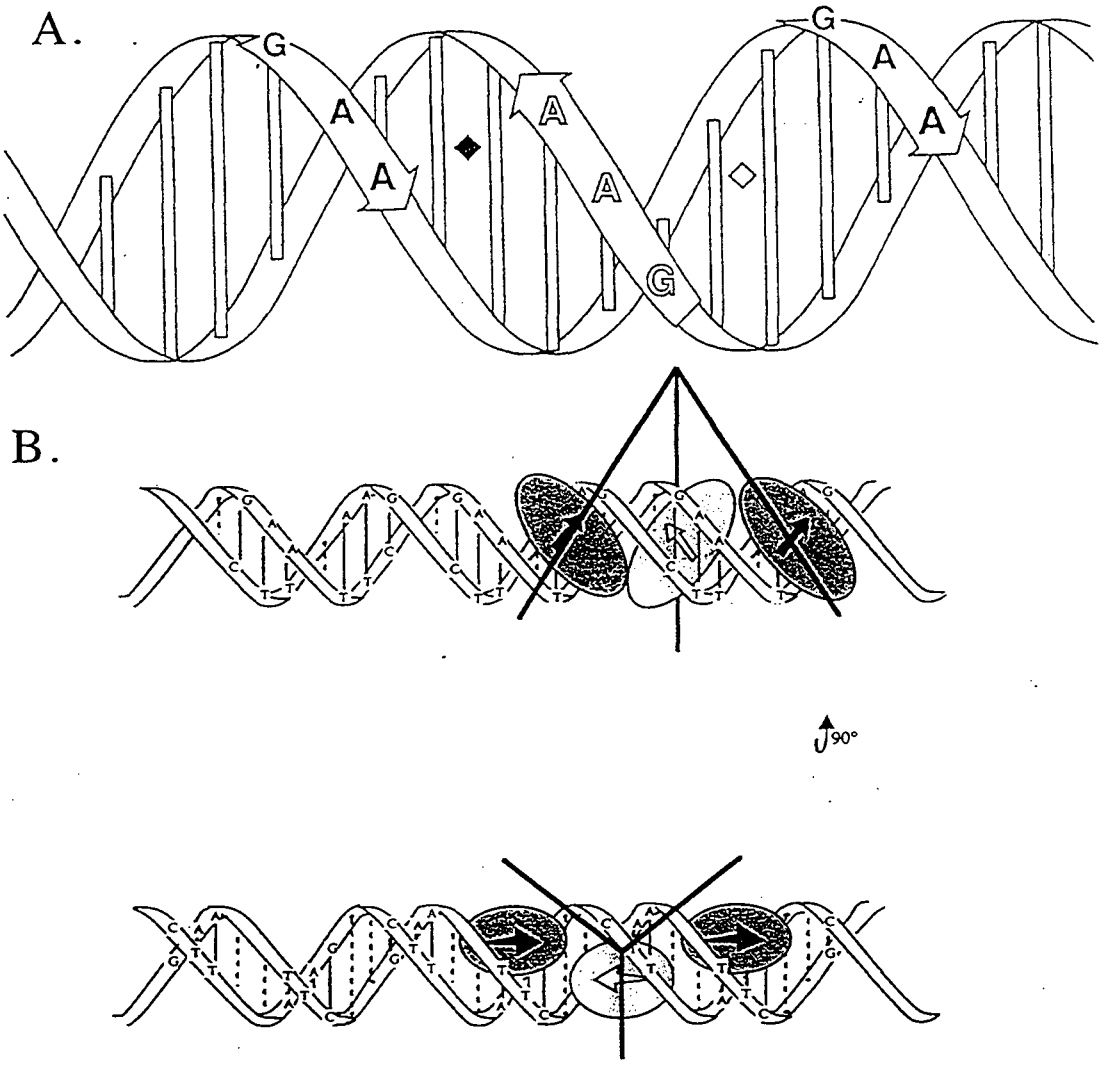


Figure 2.2 Minimal three GAA box Heat Shock Element
A. A duplex with the sequence 5'-NGAANNTTCNNGAAN-3' on one strand and the complementary sequence on the other strand contains three GAA boxes (indicated by arrows) which are arranged on two dyad axes of symmetry (small diamonds). **B.** This same symmetry must be imposed on the homotrimer upon binding so that three monomer units interact with the three GAA boxes. The symmetry of an unbound homotrimer is shown with dark bars. Bound HSF looses this symmetry and must take on the symmetry of the DNA duplex . The HSF trimer /DNA complex is shown with two views rotated 90° with respect to one another. Typical heat shock genes contain more than three GAA boxes at the promoter site and cooperative binding is exhibited by HSF in these cases suggesting adjacent trimers of HSF on the DNA form protein-protein interactions to help stabilize the complex and keep expression of heat shock genes turned on at high temperatures or other environmental stresses. Figure courtesy of Karen Flick.

HSE contains two dyad axes of symmetry between each pair of GAA boxes. This head to tail to head arrangement must be imposed on the three DNA-binding domains of HSF upon interaction of the homotrimer with the HSE (Fig. 2.2B).

The DNA-binding domain of HSF was mapped genetically to a sequence of roughly 130 amino acids (Wiederrecht, et al., 1988) which was subsequently reduced to an 89 amino acid sequence in yeast. This fragment was shown to be capable of binding HSE sites sequence specifically (Flick, et al., 1994). The sequence in this region is the most highly conserved across all HSFs. The DNA-binding domain of HSF shows no sequence homology to any known DNA-binding proteins whose structures have been determined. A limited homology to the proposed helix-turn-helix region of σ^{32} has been identified (Fig. 2.4A) with approximately 32% sequence identity between σ^{32} and the most conserved region of the HSF DBD.

The Importance of Structural Studies of HSF

The structure of this domain is of intense interest because HSF binds DNA in a highly unusual homotrimer mode and because its DNA-binding motif cannot be convincingly identified based on sequence homology. Because of the high degree of sequence conservation in this region among a diverse range of organisms, the structure should give direct insight into the recognition of HSEs by HSF in all eukaryotes and more generally add to the understanding of the role of HSF in the transduction of the heat shock response.

A number of structural studies of the DNA-binding domain of HSF are being pursued in different laboratories. The crystal structure of the DNA-binding domain of HSF from the yeast *Kluyveromyces lactis* was determined concurrently with this work in the laboratory of Professor Hillary Nelson. A protein consisting of the 89 amino acid DNA-binding fragment defined in yeast was expressed in *E. coli* and shown to bind DNA with approximately micromolar affinity. It has a melting temperature of 80° C, and is readily

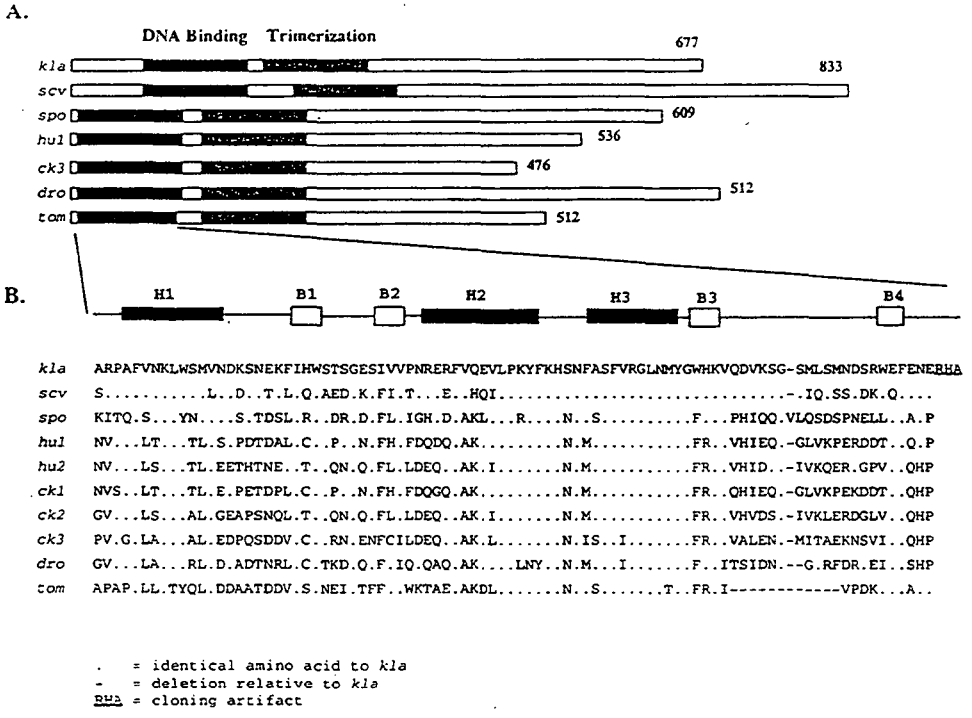


Figure 2.3 HSF domain structure, DBD sequence alignment and secondary structure **A.** Alignment of functional domains identified by deletion analysis and biochemical studies for several organisms (*Kla* = *K. lactis;* Scv = *S. cerevisiae; Hu1* = Human; *Spo* = *S. pombe; Ck3* = Chicken, HSF3; *Dro* = *D. melanogaster; Tom* = Tomato, Lp-8 HSF1). The two most highly conserved regions of HSF are the DNA-binding domain and adjacent trimerization domain. The yeasts *Kla* and *Scv* contain an N-terminal activation domain and all HSF proteins contain C-terminal activation domains. Since HSF is constitutively bound in yeast, much of the transcriptional regulation must come from interactions with the activation domains. **B.** Sequence alignment of the DNA-binding domains of HSF. Secondary structure elements (H=helices, B=sheet) described later in the text are indicated by solid bars for helices and open bars for sheets. The most highly conserved regions correspond to H2 and H3.

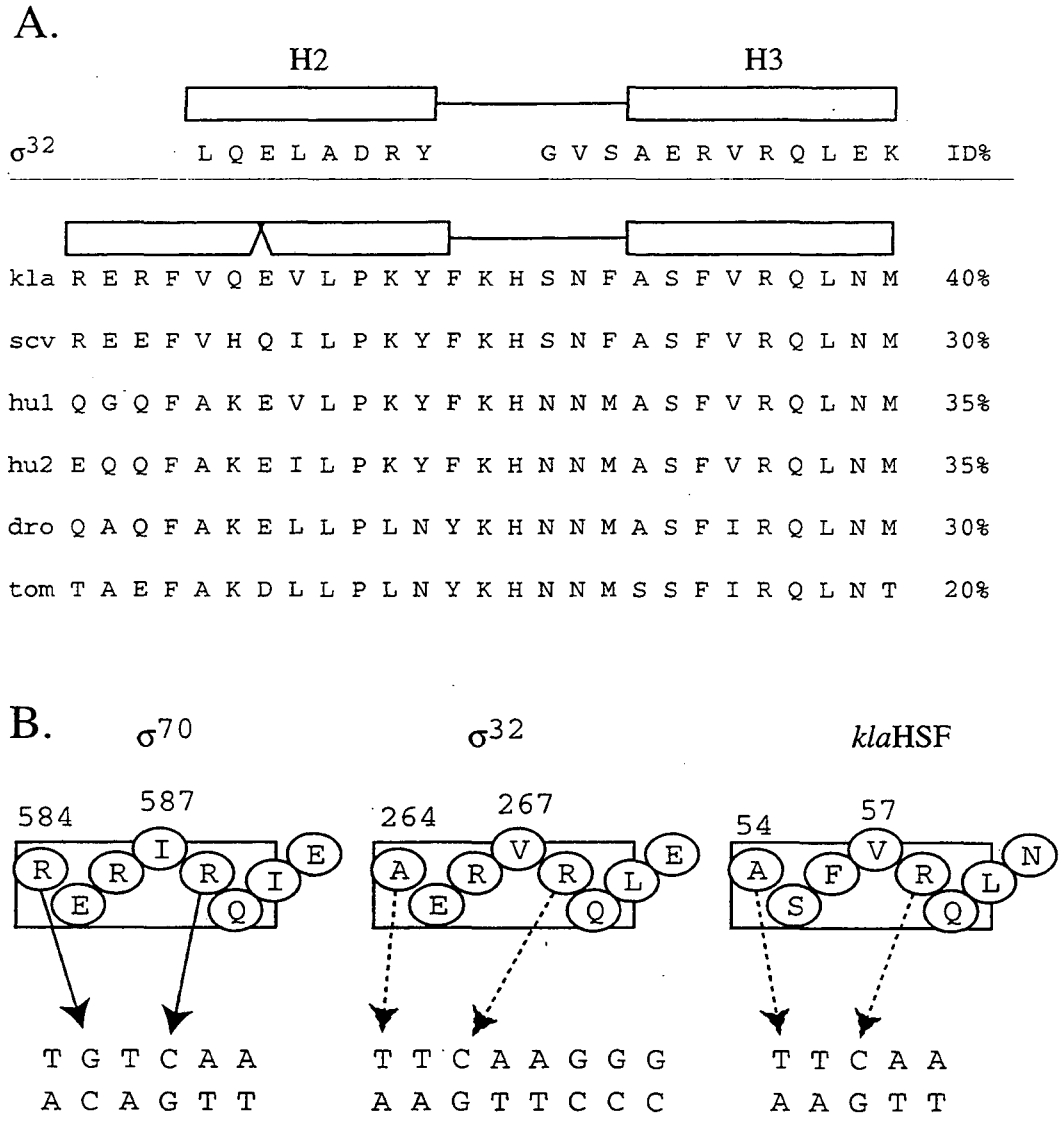


Figure 2.4 Sequence homology in the HTH of bacterial sigma factors with HSF **A.** Alignment of the sequence of residues L253-K272 in the proposed HTH of σ^{32} with the residues R36-M62 from the DNA-binding domain of the *K. lactis* HSF using the package GCG 7.3 (Devereux *et al.*, 1984) reveals a modest homology with 40% sequence identity for the residues which have partners and 65% sequence homology. HSF from other organisms also show modest homology in the helix-turn-helix region with an average for the shown sequences of 32% identity / 50% homology. **B.** Interactions seen for putative recognition helix of σ^{70} with the -35 promoter site, analogous interactions proposed for σ^{32} based on homology (Dombroski *et al.*, 1992) and interactions proposed here based on homology of HTH sequence in helix H3 and overlapping DNA-binding specificity (GAA).

soluble after the last step of purification. HSF has been crystallized in two forms (Harrison, 1994). It has all the hallmarks of a well-defined, independently folding protein domain. NMR structural studies were carried out on this protein to determine the solution structure and identify any differences from the crystal structure and to allow characterization of the dynamics. The NMR studies are also part of an ongoing effort to structurally and dynamically characterize interactions between HSF DNA-binding domains and DNA as well as protein-protein interactions between HSF DNA-binding domains.

Materials and Methods

The Sequence and Expression of the DNA-binding domain of HSF

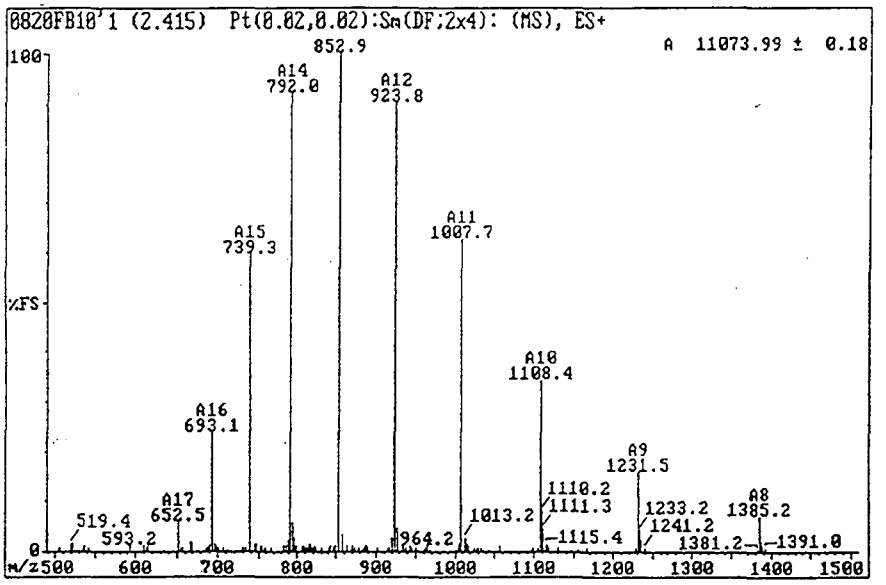
The HSF DNA binding domain from *K. lactis* (Residues 194-285) was expressed in *E. coli* using a T7 expression system (Studier, et al., 1990). The amino acid sequence encoded by the plasmid construct includes an N-terminal Met and the C-terminal sequence Arg-His-Ala. Samples labeled with ^{15}N and $^{13}C/^{15}N$ were prepared by growing bacteria on minimal media containing $^{15}NH_4Cl$ or $^{15}NH_4Cl/^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively. The protein overexpression and purification protocol are described elsewhere (Harrison, 1994). Approximately 20 mg of purified protein was obtained from each liter of cell growth. Purity was shown to be greater than 99% for all samples as determined by electrospray ionization mass spectrometry (Fig. 2.5). The molecular weights obtained indicated the absence of the N-terminal Met, as well as complete incorporation of the isotopic labels (Harrison & King, unpublished results).

Sample Preparation

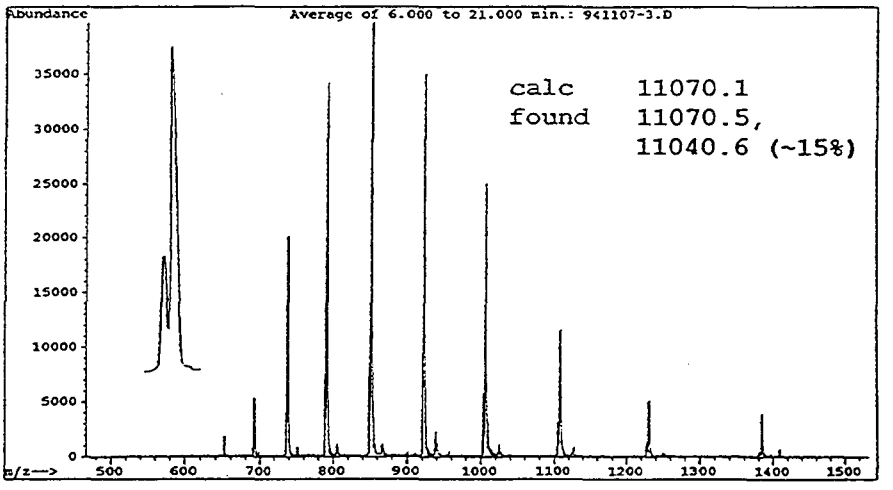
NMR samples in H_2O were prepared by dissolving 5-10 mg of lyophilized protein in 450 µl H_2O and 50 µl D_2O. NMR samples in D_2O were prepared by dissolving protein in 500 µl of 99.96% D_2O , relyophilizing the protein and redissolving it again in 99.96%

Figure 2.5 Electrospray Ionization Mass Spectrometry of isotope labeled HSF
A. ESI MASS-SPEC of uniformly ^{15}N labeled HSF. Calculated mass for fully ^{15}N labeled HSF is 11074.3. Measured mass derived from this spectrum is 11074.0 ± 0.2. Other mass species were of negligible amount. **B**.ESI-MASS-SPEC of uniformly ^{15}N (99%) labeled HSF from a separate preparation than **A**. Major species measured mass M was 11070.5 close to the calculated value. A second species (15% of sample) exhibited a measured mass of 11040.6 (M-30). This second species was traced to misincorporation of Lys for Arg at one position (R36) in the primary structure of HSF. The *K. lactis* sequence for the DBD of HSF contains three codons (AGA) (corresponding to R2, R36 and R83) which are rarely used to code for ARG in E. coli. but are quite common in yeast. Moreover E. coli commonly use AAA to code for LYS. Overexpressing bacteria are likely to find limiting supplies of charged tRNAs and the mistranslation frequency, particularly for rare codons in some cases increases dramatically. M-28 species are often observed in purified overexpressed protein and in this case it was suspected to correspond to ARG->Lys translation error at one site (since no M-2x28 or M-3x28 was observed).
N-terminal Edman sequencing of the protein was used to rule out R2 as the site of misincorporation. 0% LYS was observed at position 2 and 100% ARG. A Lys-C digest was used to generate peptide fragments corresponding to C-terminal cuts at each of the LYS sites as well as the new LYS site created by mistranslation which existed in 15% of the protein. Peptides from Lys-C digests were separated using analytical HPLC. ESI-MASS-SPEC was used to identify each of the predicted fragments from the native sequence. Lys-C digestion was found to go to completion in 2 hr at 55 C with a ratio of 1:100 Lys-C to protein in 100 mM TRIS buffer (pH 8.5 at 25C). In addition to the predicted fragments for the native sequence, a non-native fragment (M=1821.41) was identified which corresponded to the peptide FIHWSTSGESIVVPNK. The predicted mass for this peptide (uniformly ^{15}N labeled with 98% ^{15}N) is 1821.41 and it corresponds to the sequence of residues F21-R36 in the protein sequence. The ratio of integrated areas of the UV measured at 210nm for the peptide F21-K36 to that of peptide F21-R36 is 0.15. The ratio of ion signal intensity of the two peptides from ESI MASS-SPEC is 0.15. Thus all of the M-30 species can be accounted for by mistranslation of LYS for ARG at a single site (residue 36). To determine the level of translation at R83, the fragment corresponding to Lys-C digestion at residue 83 (W84-EFENERH-A92) was prepared using solid phase peptide synthesis. The retention time was determined on the same analytical HPLC used to separate Lys-C digests . No peak was detected at this retention time in repeated Lys-C digests thus ruling out mistranslation at R83 to a level below 1%. **C**. ESI-MASS-SPEC of uniformly $^{13}C/^{15}N$ labeled HSF. Calculated mass for fully $^{13}C(99\%)/^{15}N(99\%)$ labeled HSF is 11552.57. The observed mass for the only major species is 11553.4 in good agreement with the expected mass for full incorporation of the isotopes.

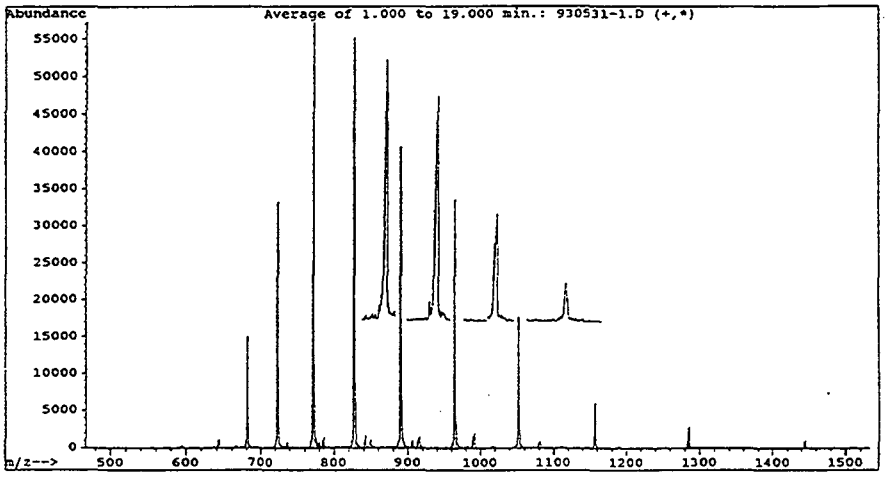
A.


0820FB10'1 (2.415) Pt(0.02,0.02):Sm(DF;2x4): (MS), ES+
A 11073.99 ± 0.18
100
%FS
0
m/z
852.9
A14 792.0
A12 923.8
A15 739.3
A11 1007.7
A10 1108.4
A16 693.1
A9 1231.5
A17 652.5
A8 1385.2
519.4
593.2
1013.2
964.2
1110.2
1111.3
1115.4
1233.2
1241.2
1381.2
1391.0
500 600 700 800 900 1000 1100 1200 1300 1400 1500

B.


Abundance
Average of 6.000 to 21.000 min.: 941107-3.D
calc 11070.1
found 11070.5,
11040.6 (~15%)
35000
30000
25000
20000
15000
10000
5000
0
m/z-->
500 600 700 800 900 1000 1100 1200 1300 1400 1500

C.


Abundance
Average of 1.000 to 19.000 min.: 930531-1.D (+,*)
55000
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
m/z-->
500 600 700 800 900 1000 1100 1200 1300 1400 1500

D_2O. All samples were buffered to pH 3.4 using 10 mM potassium phosphate buffer. Typical protein concentrations were 2 to 4 mM. Samples were stored in NMR tubes at 4°C.

NMR Spectroscopy

NMR experiments were performed on a Bruker AMX 600 spectrometer equipped with a 3-channel interface, external 50 W linear amplifier (ENI, Rochester, NY.) for ^{13}C frequencies , triple-resonance (^{1}H, ^{15}N, ^{13}C) carbon-optimized probe, and UXNMR software version 930301. All experiments were conducted at 25°C. Presaturation was accomplished using rf field strengths of 35 Hz for samples in D_2O and 140 Hz for samples in H_2O with the carrier placed on the water resonance (4.77 ppm). All NMR data were processed using the FELIX program (version 2.0) distributed by Biosym Technologies (San Diego, CA). All linear prediction was implemented using FELIX (version 2.0). Distance geometry (DG) and simulated annealing (SA) calculations were performed using the X-PLOR program version 3.0 (Brünger, 1992) . All calculations were performed on SGI workstations (Silicon Graphics, Mountain View, CA.).

^{1}H-^{1}H NMR

Two-dimensional ^{1}H-^{1}H spectra were collected with 512 t_1 and 1024 complex t_2 points using Time Proportional Phase Incrementation (TPPI) (Marion and Wüthrich, 1983) for quadrature detection in t_1 and with solvent presaturation. Spectral widths were 14.12 ppm in both dimensions. NOESY (Macura, et al., 1981) (P.S.# 3), DQF-COSY (Rance, et al., 1983) (P.S.# 1), and clean-TOCSY experiments (Griesinger, et al., 1988) (P.S.# 2) were performed in both H_2O and D_2O. 2D NOESY spectra were recorded with a mixing time of 100 ms during which the solvent was irradiated. 2D clean-TOCSY spectra (Griesinger, et al., 1988) were acquired with a 100 ms MLEV17 (Bax and Davis, 1985) mixing sequence. The ratio of clean-TOCSY delay Δ , to the spinlock 90° pulse length τ_{90} , was 2.5 with an average rf power for the spinlock of 7.5 kHz. All ^{1}H-^{1}H 2D spectra were zero filled to 1K by 1K points. In some cases linear prediction was applied in the

second dimension as implemented in FELIX version 2.0. Data were apodized in both dimensions using shifted sine-bell window functions with phase shifts ranging from 45° to 90° for different datasets.

Heteronuclear 2D NMR

Quadrature detection for all heteronuclear 2D NMR experiments was accomplished using States-TPPI (Marion, et al., 1989b). Unless otherwise stated, in all heteronuclear 2D experiments WALTZ16 modulation (Shaka, et al., 1983) was used for ^{15}N decoupling and GARP modulation (Shaka, et al., 1985) was used for ^{13}C decoupling. 2D ^{1}H-^{15}N HSQC spectra (Bodenhausen and Ruben, 1980; Bax, et al., 1990c) (P.S.#6) and 2D ^{1}H-^{15}N HMQC spectra (Mueller, 1979; Bax, et al., 1990c) (P.S.#5) were acquired on uniformly ^{15}N or $^{13}C/^{15}N$ labeled HSF in 90% H_2O/10% D_2O. The water signal was suppressed using presaturation in the HMQC experiments and both presaturation and a 1.8 ms spinlock pulse (Messerle, et al., 1989) were used in HSQC experiments. Typically the data were collected with spectral widths of 29.78 ppm in t_1 (512 complex points) and 14.12 ppm in t_2 (1024 complex points).

Amide exchange rates were probed by lyophilizing the protein from H_2O and redissolving the protein in D_2O (pH 2.75, uncorrected for ^{2}H isotope effects), and acquiring a series of 2D ^{1}H-^{15}N HMQC spectra at 13.1, 37.7, 76.8, 156, and 370 minutes after addition of D_2O. Very slowly exchanging amide signals were observed in spectra obtained from a sample dissolved in D_2O for several weeks. Spectra were zerofilled to 1K by 1K points and apodized using 60° skewed sine-bell window functions.

2D ^{1}H-^{13}C ct-HSQC spectra (P.S. #12) of the aliphatic region of HSF uniformly labeled with $^{13}C/^{15}N$ were recorded as described (Vuister and Bax, 1992). Solvent suppression was achieved using presaturation and a 1 ms spinlock pulse (Messerle, et al., 1989). Carbon decoupling during detection was accomplished using GARP (Shaka, et al., 1985) modulation of a 4.2 kHz rf field. A constant time delay of 26.6 ms was used which

resulted in oppositely phased signals for carbons attached to an even or odd number of ^{13}C neighbors. The data were collected with spectral widths of 33.13 ppm (^{13}C) in t_1 (128 complex points) and 14.12 ppm (^{1}H) in t_2 (512 complex points). Spectra were processed as described (Vuister and Bax, 1992) except that linear prediction, as implemented in FELIX (version 2.0), was used.

2D ^{1}H-^{13}C ct-HSQC (P.S.#12) and ct-HCCH-TOCSY (P.S.#19) spectra (Ikura, et al., 1991b) optimized for detection of aromatic resonances were collected with the ^{13}C carrier set to 125 ppm and INEPT transfer delays set to 1.3 ms. The spectra were collected with spectral widths of 26.29 ppm (^{13}C) in t_1 (56 complex points) and 11.73 ppm (^{1}H) in t_2 (512 complex points). The constant time delay was tuned to ^{13}C-^{13}C couplings of 70 Hz (14.3 ms) and no carbonyl pulses were applied. The ^{13}C mixing time in the ct-HCCH-TOCSY experiment was accomplished using a 14.34 ms DIPSI-2 sequence (Shaka, et al., 1988) with a 7.6 kHz spinlock field. Both datasets were apodized with a 15° skewed sine-bell in t_2 and a 35° skewed sine-bell in t_1. Linear prediction as implemented in FELIX (version 2.0) was applied in the t_1 time domain. The final data sets were 1024 x 1024 real points in size.

$^{3}J_{NH\alpha}$ coupling constants were measured semiquantitatively (either less than 6 Hz or greater than 8 Hz) from a ^{15}N HMQC-J spectrum (Kay & Bax, 1990) in which 650 complex points (t_{1max} = 214 ms) were collected. Exponential line broadening (7 Hz) and exponential line narrowing (-2Hz) were applied in the ^{1}H and ^{15}N dimensions, respectively. After zero-filling, the final digital resolution was 4 Hz (^{1}H) and 2 Hz (^{15}N). These data were used to estimate ϕ torsion angles for 36 residues.

$^{3}J_{NC\gamma}$ and $^{3}J_{C'C\gamma}$ coupling constants for Val, Ile, and Thr residues were determined from 2D ^{15}N and ^{13}C spin-echo difference ct-HSQC experiments as described (Grzesiek, et al., 1993; Vuister, et al., 1993) using a constant time period of 28.6 ms. These data were used to determine χ_1 torsion angle constraints for 7 of 9 Val, both Ile, and

the single Thr residue, and to stereospecifically assign the $^{13}C\gamma$ signals for 7 of the 9 Val residues.

^{15}N-separated 3D NMR

3D ^{15}N-separated TOCSY-HMQC (Driscoll, et al., 1990) (P.S. #7) and 3D ^{15}N-separated NOESY-HMQC experiments (Kay, et al., 1989; Marion, et al., 1989c) (P.S.#8) were acquired on uniformly ^{15}N-labeled samples of HSF in 90% H_2O/10% D_2O. Quadrature detection in the indirectly detected dimensions was accomplished using TPPI (Marion and Wüthrich, 1983). In both experiments GARP modulation (Shaka, et al., 1985) was used to decouple ^{15}N. Solvent suppression was implemented using presaturation. The TOCSY-HMQC experiment was performed using a 100ms clean TOCSY MLEV17 mixing sequence (Bax and Davis, 1985; Griesinger, et al., 1988). The ratio of the clean TOCSY delay , Δ , to the spinlock 90° pulse length, τ_{90} , was equal to 2.5 with an average spin-lock field strength of 6.7 kHz. The 3D-^{15}N-separated NOESY-HMQC experiment was performed using a 100 ms mixing time. Both the TOCSY and NOESY experiments were acquired with spectral widths of 14.12 ppm, 60.0 ppm, and 14.12 ppm in F1 (^1H), F2 (^{15}N), and F3 (^1H) respectively, and with 200 real points in t_1, 62 real points in t_2, and 1024 real points in t_3. A total of 16 scans were signal averaged for each real t_1, t_2 pair. Shifted sine-bell window functions were applied in all three dimensions and linear prediction (FELIX program, version 2.0) and zerofilling in t_1 and t_2 resulted in a final matrix for both experiments of 512 (^1H) x 512 (^1H) x 64 (^{15}N) points.

Triple Resonance NMR

CBCANH (Grzesiek and Bax, 1992a) (P.S.#9) and CBCA(CO)NH (Grzesiek and Bax, 1992b) (P.S. #10) triple resonance experiments were performed as described on uniformly $^{13}C/^{15}N$-labeled samples in 90% H_2O/10% D_2O. Presaturation and a 1.8 ms spinlock pulse (Messerle, et al., 1989) were used to suppress the solvent in the CBCANH experiment whereas only the 1.8 ms spinlock pulse was retained for solvent suppression in

the CBCA(CO)NH experiment. Both the CBCANH and CBCA(CO)NH spectra were acquired with spectral widths of 56.0, 24.9, and 14.12 ppm for F1 (^{13}C; 52 complex points), F2 (^{15}N; 32 complex points), and F3 (^{1}H 512 complex points) respectively, and with 32 scans per complex t_1, t_2 pair. The carriers were placed at 43.0 ppm, 117.9 ppm, and 4.77 ppm for the ^{13}C, ^{15}N, and ^{1}H channels, respectively. ^{15}N decoupling during acquisition was accomplished using WALTZ16 (Shaka, et al., 1983) modulation of a 1.6 kHz rf field. ^{1}H decoupling during the ^{13}C evolution period was achieved using DIPSI-2 modulation (Shaka, et al., 1988) of a 5.4 kHz rf field. Time domain convolution (Marion, et al., 1989a) was used to eliminate the residual solvent signal in both experiments and a 50° skewed sine-bell window function was applied in the t_3 dimension. Linear prediction (FELIX version 2.0) was applied in t_1 and t_2 followed by multiplication by unshifted cosine squared window functions. Both matrices were 512 (^{1}H) x 256 (^{13}C) x 64 (^{15}N) real points.

^{13}C-separated 3D experiments

3D-HCCH-TOCSY (Clore, et al., 1990; Bax, et al., 1990b) (P.S. #14) and 3D-^{13}C-separated NOESY-HMQC spectra (Ikura, et al., 1990) (P.S. #13) were recorded using uniformly $^{13}C/^{15}N$-labeled HSF samples in 99.96% D_2O. Quadrature detection was achieved using the States-TPPI method (Marion, et al., 1989b). In both experiments the ^{13}C carrier was set to 43 ppm and ^{13}C decoupling during acquisition was accomplished using GARP modulation (Shaka, et al., 1985) of a 4.1 kHz field. Solvent suppression was achieved in the HCCH-TOCSY experiment using presaturation and a 1.8 ms spinlock pulse. In the ^{13}C-separated NOESY-HMQC spectrum 0.5 ms and 1.8 ms spinlock pulses were used for solvent suppression (Messerle, et al., 1989) . Mixing in the ^{13}C HCCH-TOCSY experiment was accomplished using a DIPSI-2 sequence (Shaka, et al., 1988) with a mixing time of 22.6 ms and a spinlock field of 5.4 kHz. The mixing time in the ^{13}C-separated NOESY-HMQC experiment was 100 ms. Both experiments were collected

with 128 complex points in t_1 (^1H), 32 complex points in t_2 (^{13}C), and 256 complex points in t_3 (^1H) and with 32 scans per complex t_1, t_2 pair. Spectral widths for the HCCH-TOCSY spectrum were 8.01, 23.8, and 8.01 ppm in F1 (^1H), F2 (^{13}C), and F3 (^1H), respectively, and for the ^{13}C-separated NOESY-HMQC spectrum they were 10.93, 23.8, and 13.44 ppm in F1 (^1H), F2 (^{13}C), and F3 (^1H), respectively. The HCCH-TOCSY data were multiplied by 50° shifted sine-bell window functions in all three dimensions and linear prediction (FELIX version 2.0) was applied in t_1 resulting in a final matrix size of 512 (^1H) x 128 (^1H) x 64 (^{13}C) points. The ^{13}C-separated NOESY data were zerofilled and then apodized with a 90° skewed sine-bell squared window function in t_3. Data in the t_1 and t_2 dimensions were multiplied by a skewed sine-bell window function and linear prediction was applied in t_1 to arrive at a final matrix size of 256 (^1H) x 128 (^1H) x 64 (^{13}C) real points.

4D $^{13}C/^{13}C$-edited NOESY spectrum

A 4D $^{13}C/^{13}C$-edited NOESY spectrum was recorded as described (Clore, et al., 1991) (P.S. #15) on a $^{13}C/^{15}N$ labeled sample of HSF dissolved in 99.96% D2O with a mixing time of 100 ms. The acquired data matrix consisted of 8 (t_1, ^{13}C) x 64 (t_2, ^1H) x 8 (t_3, ^{13}C) x 256 (t_4, ^1H) complex points with spectral widths of 23.8 ppm in both ^{13}C dimensions and 9.15 ppm in both ^1H dimensions. The ^{13}C carrier was centered at 43 ppm during labeling of carbon magnetization and the ^1H carrier was positioned at 3.5 ppm except for solvent presaturation which was accomplished at 4.77 ppm. Data were processed as described (Clore, et al., 1991) except that linear prediction as implemented in FELIX (version 2.0) was used to extend each ^{13}C dimension from 8 to 12 points followed by zero filling to 32 complex points. The final data matrix consisted of 32 (F1) x 128 (F2) x 32 (F3) x 256 (F4) real points.

Chemical shift referencing

Chemical shifts were directly referenced to 3-(2,2,3,3-$^2H_4$) trimethylsilyl propionate (TSP) (^1H) and indirectly referenced to liquid ammonia (^{15}N) (Live, et al., 1984), and TSP (^{13}C) (Bax & Subramanian, 1986).

Structure Calculations

NOEs were classified qualitatively as strong (1.8-2.7 Å), medium (1.8-3.5 Å), and weak (1.8-5.0 Å) (Williamson, et al., 1985; Clore, et al., 1986). A correction was added to the upper limit for constraints involving methyl protons and nonstereospecifically assigned protons (Wüthrich, et al., 1983; Clore, et al., 1986). Phi torsion angles were constrained to -60° ± 30° for $^3J_{NH\alpha}$ values that were less than 6 Hz, and to -120° ± 40° for $^3J_{HN\alpha}$ values that were greater than 8 Hz (Pardi, et al., 1984). χ_1 angles were constrained to within ± 20° of the measured values (Kraulis, et al., 1989). Hydrogen bonds were identified on the basis of NOEs identified in regular secondary structures and low amide proton exchange rates (still present after 370 minutes of exchange). Constraints were applied between N and O atoms (2.8-3.3 Å) and between H_N and O atoms (1.8-2.3 Å) (Clore, et al., 1990).

Structures were calculated using a hybrid distance geometry/simulated annealing method (Nilges, et al., 1988) as contained in the X-PLOR program 3.0 (Brünger, 1992). Random starting structures were calculated using the X-PLOR routine *generate-template*. A subset of atoms from this extended conformation was used for embedding using the distance and torsion angle constraint matrix with the *dg_sub_embed* routine. The hybrid distance geometry/simulated annealing routine *dgsa* was applied to the full structure as described (Brünger, 1992). Final refinement was applied using the routine *refine*.

The initial set of structures, calculated on the basis of approximately 300 NOE and 24 hydrogen bond constraints (12 H-bonds), was used to resolve ambiguities in NOE assignments allowing new NOE constraints to be added in further rounds of refinement.

Thirty final structures were calculated and the 24 best were chosen for structural comparison. The chosen 24 structures had no distance violations greater than 0.5 Å and no torsion angle violations greater than 5°. The average total target function energy for the family of structures at the end of X-PLOR refinement was 150.9 ± 13.4 kcal/mol. The average NOE and dihedral violation energies were 25.9 ± 5.4 kcal/mol and 1.0 ± 1.6 kcal/mol respectively, using force constants of 50 kcal/mol $Å^2$ and 200 kcal/mol rad^2. The <RMSD>s between the distance constraints and calculated structures for the family of 24 structures was 0.028 ± 0.003 Å and the <RMSD> between the dihedral constraints and calculated structures was 0.35 ± 0.13°. The coordinates for these 24 structures have been deposited with the Brookhaven Protein Data Bank (file 2HSF). Superpositions and analysis of the structures was accomplished using X-PLOR and Insight II software.

Results

Resonance Assignments

Using the methods described below complete sequential assignments were obtained for 85 of 89 backbone N-H signals in the ^1H-^{15}N HSQC spectrum (Fig. 2.6). This includes all N-H signals except those of Ala-1, Gly-28, Asn-80 and Asp-81. Aside from Ala-1, tentative assignments could be made for these residues, although not confirmed due to overlap (Asn-80, Asp-81) or lack of NOEs (Gly-28). In addition all the εN-H pairs of the six Arg and ε1 N-H pairs of the four Trp residues were assigned based on characteristic Arg-$^{15}N\varepsilon$ (~85 ppm) and Trp-$NH_{\varepsilon 1}$ (~10ppm) shifts. Twelve pairs of signals expected from the 9 Asn and 3 Gln NH2 groups were also identified although not assigned because of severe overlap of their $^1H_\gamma$ and $^1H_\beta$ resonances. The completeness of the backbone N-H assignments (95 %) was achieved despite a significant number of degeneracies observed: 10 backbone amides within 0.3 ppm (^{15}N) and 0.02 ppm (^1H) of another resonance (Fig. 2.6).

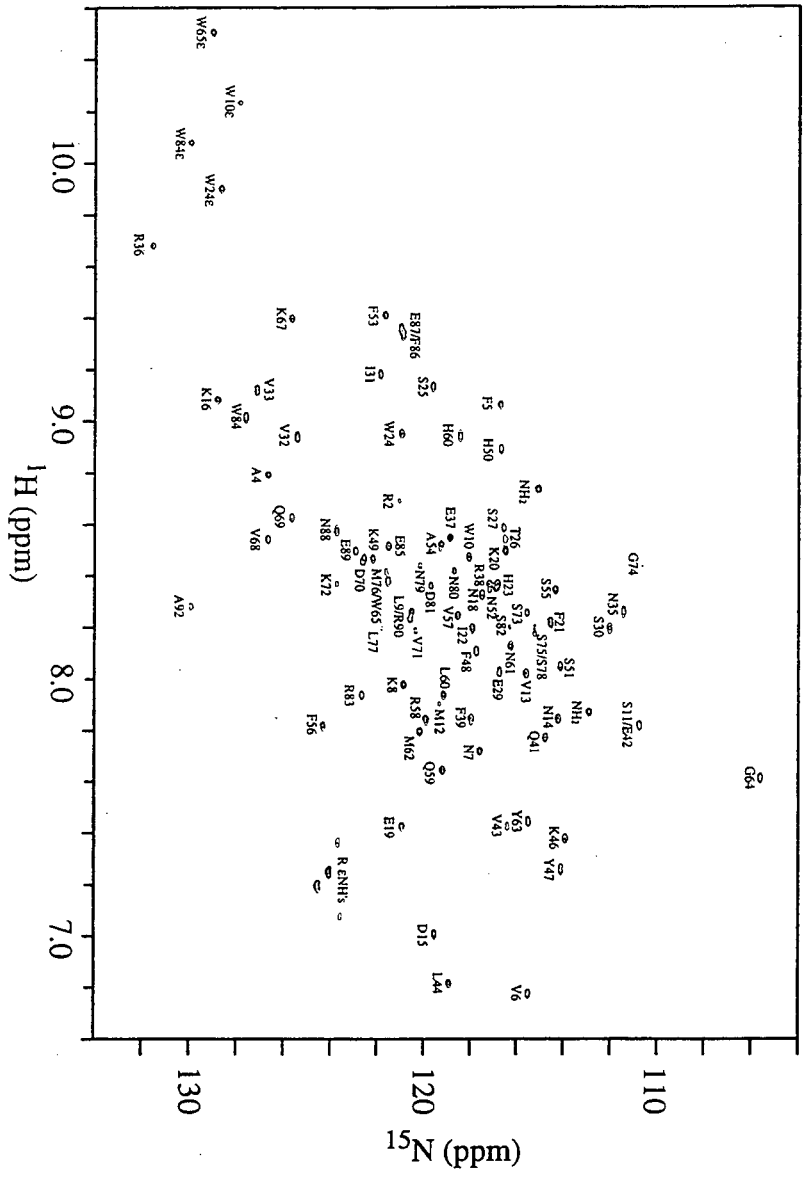


Figure 2.6 ^{1}H-^{15}N HMQC spectrum of uniformly ^{15}N-labeled HSF at 25 C, pH 3.4. Note the overall good dispersion of the spectrum, indicative of a structured protein. Several regions of significant degeneracy occur however in the central grouping of peaks. Overall 85 of the 89 expected backbone N-H resonances where assigned in this spectrum through methods described in the text. The downfield group of 4 resonances near 10 ppm in the proton dimension were assigned to the four εN-H signals from the Trp indole groups. A group of 6 peaks aliased from outside the ^{15}N spectral width (^{1}H frequencies of ~7.1 ppm) have been assigned to the six Arg εN-H groups. Figure courtesy of Ho Cho and Corey Liu.

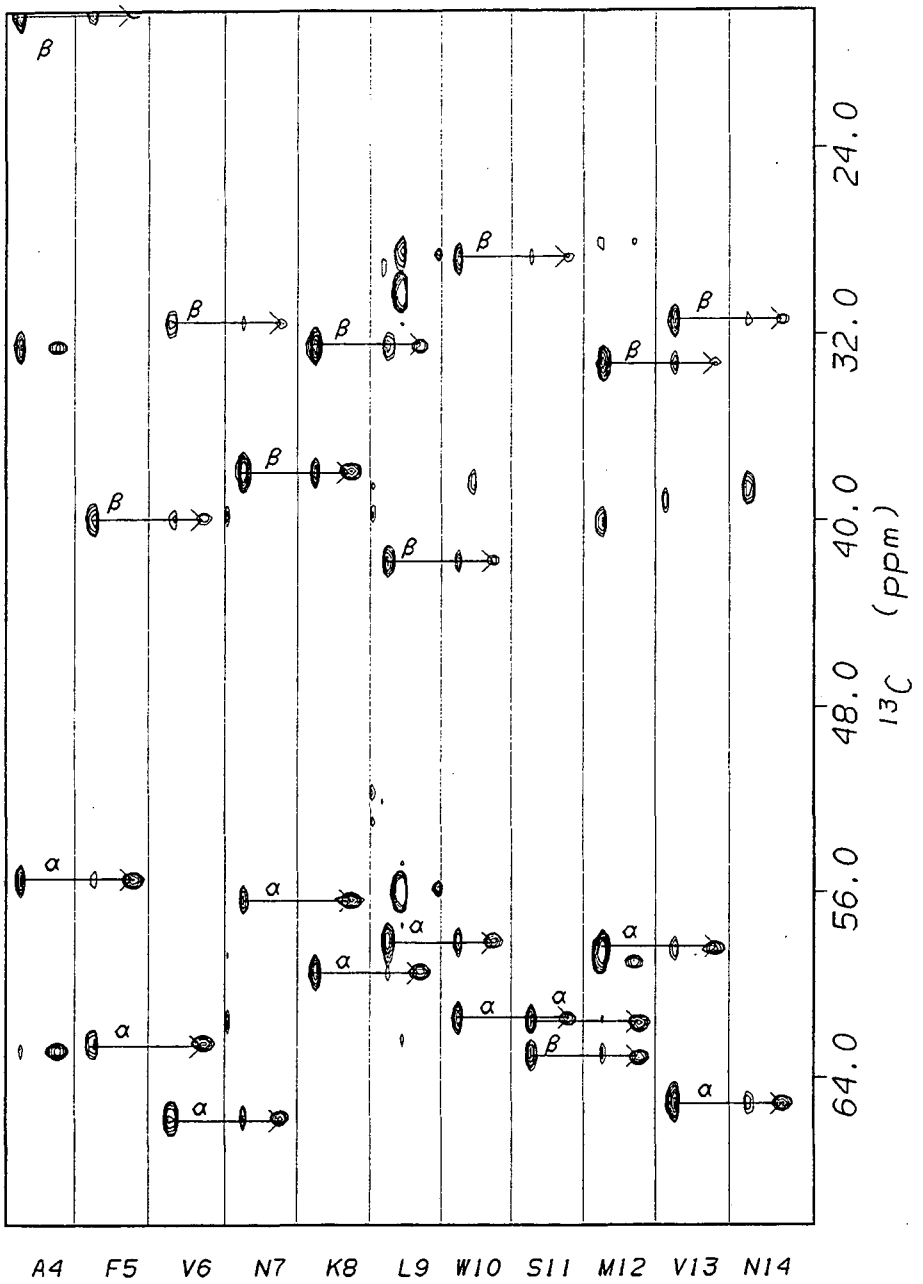


Figure 2.7 Sequential strips for residues Ala-4 to Asn-14 extracted from 3D triple-resonance CBCANH (left side of strips) and CBCA(CO)NH (right side of strips) spectra that were used to make sequential backbone assignments of HSF. Residues are denoted by one-letter amino acid code and residue number. Intraresidue C_α and C_β correlations (CBCANH) are labeled with Greek letters. Arrows link intraresidue C_α and C_β signals to the interresidue C_α and C_β signals observed in both CBACNH and CBCA(CO)NH strips from the following residue. Both intraresidue and sequential C_α cross peaks in the CBCANH spectrum are of opposite sign to other crosspeaks. ross peaks rendered with dashed lines are of opposite sign to those drawn with solid lines.

Assignments using Triple Resonance Data

The 3D triple-resonance CBCANH (Grzesiek and Bax, 1992a) and CBCA(CO)NH (Grzesiek and Bax, 1992b) experiments were used to sequentially assign the backbone ^{15}N, NH, C_α, and sidechain C_β signals. This method has recently been successfully applied to *Acanthamoeba* profilin-I (Archer, et al., 1993), human profilin (Metzler, et al., 1993) and human macrophage inflammatory protein-1β (hMIP-1β) (Lodi, et al., 1994). As can be seen in Fig. 2.7, the CBCANH experiment correlates both the C_β and C_α frequencies of a residue with its own ^{15}N and NH frequencies via the one-bond $^1J_{C\alpha N}$ coupling (~ 12 Hz), and with the ^{15}N and ^1H of the following residue via the two-bond $^2J_{C\alpha N}$ coupling (~ 7 Hz). In contrast, the CBCA(CO)NH experiment correlates the C_β and C_α frequencies of a residue to only the ^{15}N and ^1H of the following residue. Thus, through comparison of the two spectra, intra- and inter-residue correlations can be identified and used to define which residues are adjacent in the sequence. For example, for Trp-10 the inter-residue C_α and C_β correlations identified from the CBCA(CO)NH spectrum (C_α 58.0 ppm, C_β 47.1 ppm) correspond to the intraresidue C_α and C_β correlations for Leu-9 from the CBCANH spectrum indicating that these residues are adjacent in the sequence. Using this approach we were able to correlate C_α, C_β, ^{15}N, and NH signals for 65 of the 89 residues identified in the triple-resonance spectra and link fragments of sequential residues from 2 to 11 amino acids in length.

Spin system Identification

Through analysis of a ^{15}N-separated TOCSY-HMQC spectrum, we were able to correlate ^{15}N, HN, H_α, and H_β signals for approximately 70% of the residues. These data combined with the intra-residue C_α and C_β chemical shifts for each residue identified in the triple-resonance data provided a convenient starting point for assignment of the side chain ^1H and ^{13}C signals through analysis of a 3D HCCH-TOCSY spectrum (Clore, et al., 1990; Bax, et al., 1990b) as described previously (Pelton, et al., 1991; Ikura, et al., 1991a). and

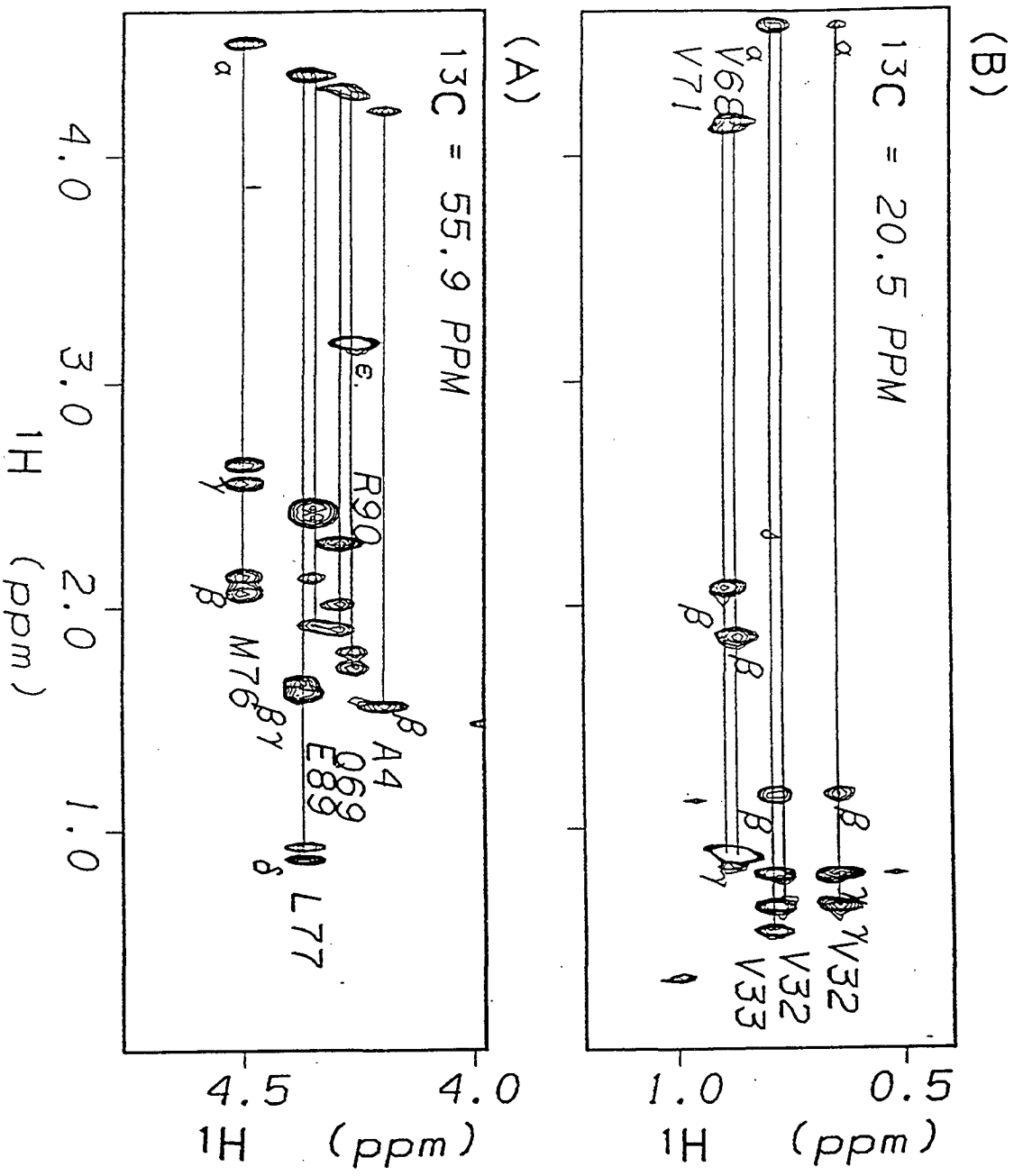


Figure 2.8 Expansion of representative planes from the 3D HCCH-TOCSY spectrum **A**. showing TOCSY correlations from labeled $^{13}C_{\alpha}$-$^{1}H_{\alpha}$ pairs (^{13}C = 55.9 ppm) and **B**. TOCSY correlations from valine $^{13}C_{\gamma}$-$^{1}H_{\gamma}$ pairs (^{13}C = 20.5 ppm).
Note the efficient transfer out to the ends of the Leucine H_{δ}'s from the H_{α}. The highly efficient magnetization transfer in the HCCH TOCSY through the large one-bond ^{13}C couplings in the aliphatic sidechains allowed complete sidechain ^{1}H-^{13}C assignments to be obtained.

in Chapter 1. Representative data are shown in Fig. 2.8. Fig. 2.8A shows $^{13}C_{\alpha}$-H_{α} ($^{13}C_{\alpha}$ 55.9 ppm) to sidechain ^{1}H correlations for several residues including Ala-4, Gln-69, Met-76, Leu-77, and Glu-89. Fig. 2.8B shows correlations from Val $^{13}C_{\gamma}$-H_{γ} pairs ($^{13}C_{\gamma}$ 20.5 ppm) to H_{β} and H_{α} resonances.

Once the ^{1}H spin systems were identified using the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ signals, the remaining sidechain ^{13}C spin systems were assigned by searching for the characteristic ^{1}H chemical shift pattern of a residue in the appropriate ^{13}C chemical shift range (plane) of the 3D spectrum. For example in Fig. 2.9 the ^{1}H correlations obtained from several sidechain Val $^{13}C_{\gamma}$-^{1}H pairs are shown. The $^{13}C_{\gamma}$-$^{1}H_{\alpha}$ signals in this plane were matched to $^{13}C_{\alpha}$-$^{1}H_{\alpha}$/$^{1}H_{\gamma}$ signals in the corresponding C_{α} plane thus assigning the Valine $^{13}C_{\gamma}$-$^{1}H_{\gamma}$ signals. The spin system assignments were used to register many of the fragments of sequential residues identified in the CBCANH and CBCA(CO)NH data with the sequence of HSF. In addition, analysis of the HCCH-TOCSY spectrum allowed us to identify C_{β} and C_{α} frequencies in cases where these signals were weak or absent in the CBCANH/CBCA(CO)NH data allowing for the extension of the fragments. Moreover by making use of type assignments and characteristic ^{13}C chemical shifts (Horvath & Lilley, 1978), complete side chain ^{1}H and ^{13}C assignments were made without use of a 3D HCCH-COSY spectrum (Bax, et al., 1990a) The CBCANH and CBCA(CO)NH data again supplied useful complementary information by eliminating ambiguities in distinguishing H_{β}-C_{β} and H_{γ}-C_{γ} pairs in the HCCH-TOCSY spectrum since C_{β} chemical shifts were already determined from the triple resonance experiments.

In addition to the HCCH-TOCSY spectrum, 2D ^{1}H DQF-COSY and TOCSY spectra and a 3D ^{15}N-separated TOCSY-HMQC spectrum were used to assign some spin systems using standard methods (Wüthrich, 1986). For example, the Arg spin systems were identified in the 3D ^{15}N-separated TOCSY-HMQC experiment by matching backbone NH and sidechain NH_{ε} cross peak patterns. Fifteen other residues were identified by

looking for spin system patterns in 2D DQF-COSY and TOCSY spectra and correlating them with the 3D ^{15}N-separated TOCSY-HMQC spectrum. Many of the remaining residues were in the large class of AMX spin systems which, excluding the distinctive serines, comprise 27 of the 89 spin systems observable in the ^{15}N-separated 3D experiments (there are three prolines in the sequence). Finally, the 4 Trp, 2 Tyr and 4 of the 7 Phe aromatic spin systems were assigned through analysis of 2D ct-HSQC and 2D ct-HCCH-TOCSY spectra (Ikura, et al., 1991b) which were optimized for aromatic spin systems. Subsequently, sequential assignments were obtained through identification of intraresidue NOEs between H_δ and both H_α and H_β and between H_δ and HN in 3D ^{13}C-separated NOESY-HMQC and 3D ^{15}N-separated NOESY-HMQC spectra, respectively. In total, using the methods described, 95% of the backbone and 80% of the sidechain ^1H and ^{13}C signals have been assigned. The assignments are summarized in Appendix A.

Secondary Structure

Upon completion of the resonance assignments, we undertook analysis of 2D ^1H NOESY and 3D ^{15}N-separated and ^{13}C-separated NOESY-HMQC spectra and a 4D $^{13}C/^{13}C$-edited NOESY spectrum to define the secondary structure of HSF. Characteristic patterns of sequential, medium-range, and long-range NOEs observed in these spectra can be used to identify elements of regular secondary structure (Wüthrich, 1986). Strips from the 3D ^{15}N-separated NOESY-HMQC spectrum of HSF are shown in Fig. 2.9 (and a complete plane also is shown in Fig. 1.6), and a complete analysis of the sequential and medium-range NOEs is presented in Fig. 2.10. In Fig. 2.9, the presence of strong NOEs between adjacent NH protons (d_{NN} NOEs) for residues Phe-56 to Tyr-63 (H3) indicate that this segment adopts a helical structure. The identification of $d_{\alpha N(i,i+3)}$ and $d_{\alpha\beta(i,i+3)}$ NOEs, slowly exchanging amide protons, small $^3J_{NH\alpha}$ coupling constants, and upfield shifted $^{13}C_\alpha$ signals (Spera & Bax, 1991; Wishart, et al., 1991) (Fig. 2.10) confirm that these residues form part of an α-helix. The segments Ala-4 - Asn-14 (H1) and Arg-36 -

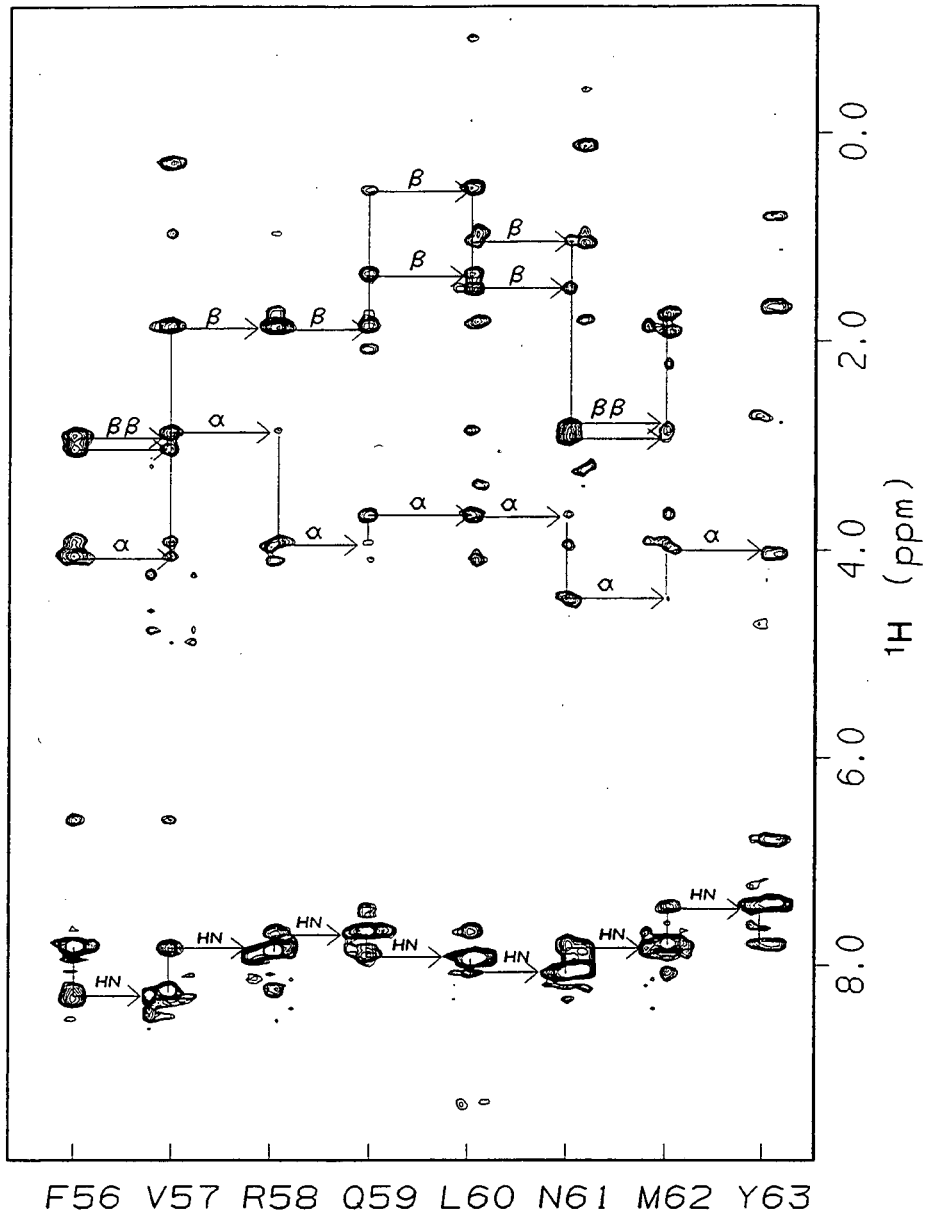


Figure 2.9 Sequential strips for residues Phe-56 to Tyr 63 extracted from a 3D ^{15}N-separated NOESY-HMQC spectrum of HSF. Residues are indicated by one-letter amino acid code and residue number. Intraresidue dαN and $d_{\beta N}$ NOEs are labeled with greek letters. Arrows indicate NOEs to sequential amide protons. The helical secondary structure of this region is indicated by the strong sequential d_{NN}, $d_{\beta N}$, and weak sequential $d_{\alpha N}$ NOEs.

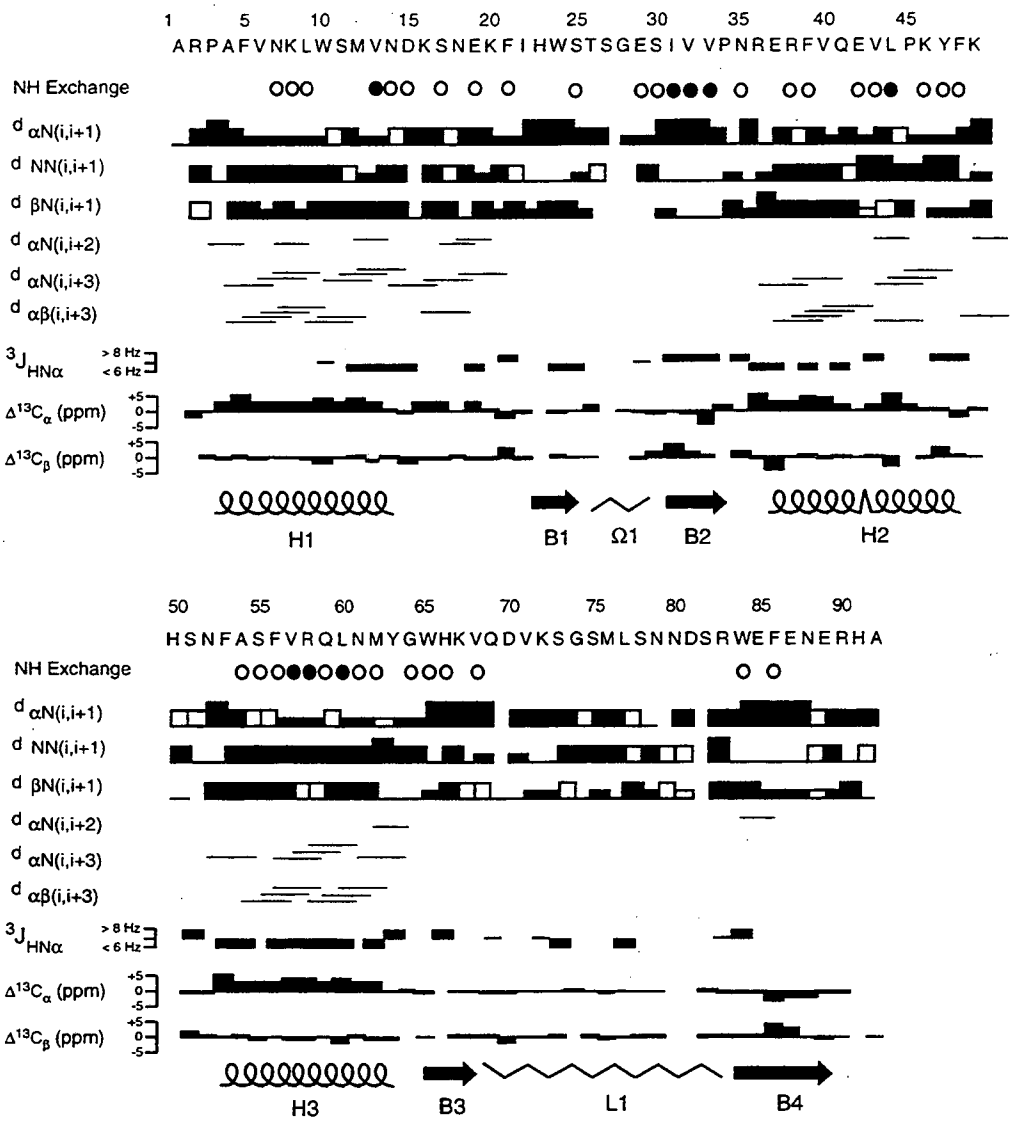


Figure 2.10 Secondary structure diagram of HSF
Summary of sequential and intermediate range NOE connectivities, amide hydrogen exchange rates, $^3J_{NH\alpha}$ coupling constants, and deviations of $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts from random coil values. The NOE correlations were determined from experiments described in the text. Sequential NOE intensities are denoted by bar height as strong, medium or weak. Open boxes indicate upper bounds on sequential NOEs in cases where overlap prevented accurate intensity determination. Lines indicate that no sequential NOE could be detected. For proline residues, NOE intensities to their $H\delta$ protons were substituted for NOE intensities to H_N protons. For medium range NOE connectivities, lines denote that an NOE was observed correlating the relevant protons. Open and filled circles indicate residues with amide H_N resonances that are not fully exchanged 370 minutes and several weeks after dissolution in D_2O (25°C, pH 2.75), respectively. Secondary structure deduced from the NOESY data are indicate below the $^{13}C\beta$ chemical shift data; helices by coils, sheets by arrows, random coil by wavy lines.

Phe-48 (H2) also show a series of d_{NN} NOEs indicative of helical secondary structure. For segment H1 the expected patterns of $d_{\alpha N(i,i+3)}$ and $d_{\alpha\beta(i,i+3)}$ NOEs and upfield shifted $^{13}C_{\alpha}$ signals confirm that this segment is α-helical.

For segment H2 the first residues (Glu-37 - Gln-41) show all the evidence for alpha helical conformation exhibited by segments H1 and H3. However the segment surrounding Pro-45 (Gln-41 - Phe-48) is characterized by several weak intensity $d_{\beta N}$ NOEs and far fewer $d_{\alpha N(i,i+3)}$ and $d_{\alpha\beta(i,i+3)}$ NOEs (Fig. 4). Furthermore, unlike segments H1 and H3 where $^{13}C_{\alpha}$ shifts are consistently upfield, several residues centered around Glu-42 show small or no upfield shifts and in addition, the $^{3}J_{NH\alpha}$ coupling constant of Val-43 is large (ϕ=-120 ± 68°) indicating a non-helical backbone structure. These data suggest that the central portion of helix H2, contains a distortion centered between residues Gln-41 and Val-43.

Between segments H1 and B1 is a short segment (Lys-16 - Lys-20) which shows medium d_{NN} and $d_{\beta N}$ NOEs, a number of $d_{\alpha N(i,i+2)}$ and $d_{\alpha N(i,i+3)}$ NOEs, a single $d_{\alpha\beta(i,i+3)}$ NOE and a number of upfield shifted $^{13}C_{\alpha}$ signals indicative of a helical /turn structure.

In addition to the three helices, four segments characterized by strong $d_{\alpha N}$ and weak d_{NN} NOEs were observed indicating that these residues adopt an extended conformation. As shown in Fig. 2.11, these segments form a four-stranded antiparallel β-sheet as evidenced by the long-range $d_{\alpha N}$, d_{NN} and $d_{\alpha\alpha}$ NOEs observed in 3D ^{15}N- and ^{13}C-separated NOESY-HMQC and a 4D $^{13}C/^{13}C$-edited NOESY spectrum. Another segment consisting of residues Gln-69 to Ser-83 (L1) also exhibited medium intensity $d_{\alpha N}$ NOEs. However, no medium-range or long-range NOEs were identified in the NOESY spectra. Furthermore, no protected amides were found, and the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ shifts for these residues show relatively small shifts (Fig. 2.10). Together, these data indicate that

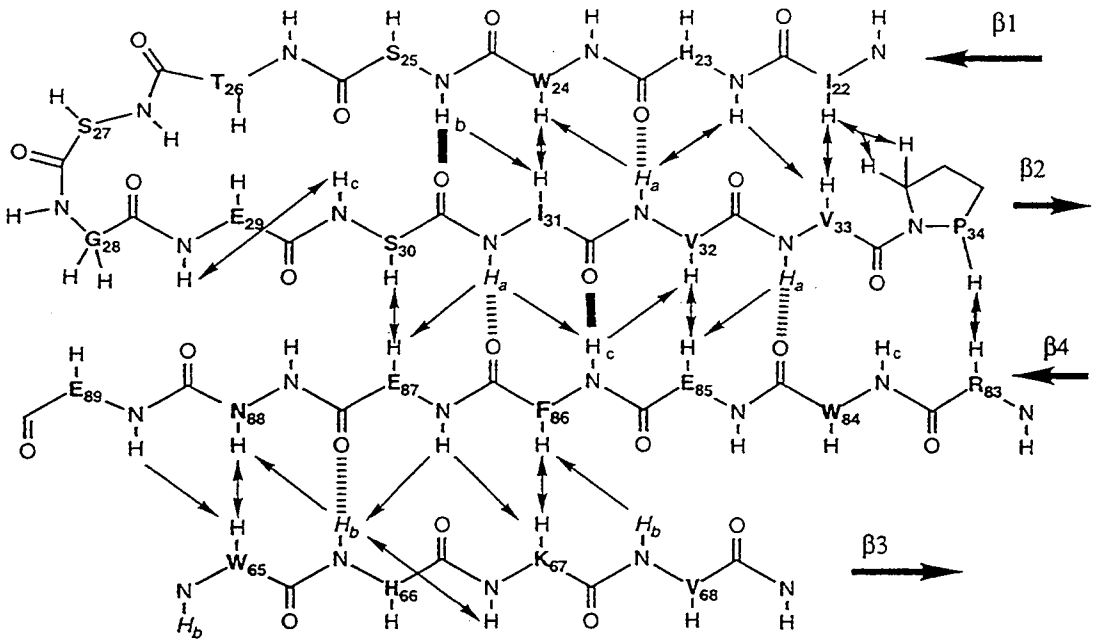


Figure 2.11 Schematic diagram of the β-sheet topology of HSF. Solid arrows denote long-range d_{NN}, $d_{\alpha N}$, and $d_{\alpha\alpha}$ NOEs. Dashed lines represent hydrogen bonds derived from the amide hydrogen exchange and NOE data. Thick arrows indicate the direction and location of the four β-strands of antiparallel β-sheet. H_a and H_b indicate amide hydrogens that were still present after dissolution in D_2O for two weeks and 370 minutes respectively. H_c indicates hydrogen bonds still present after 180 minutes which could not be detected after 370 minutes of exchange. Additional hydrogen bonds assigned at a latter stage of analysis (see Chapter 3) are indicated by dark bars and were not used in structure calculations described in this chapter.

Table 2.1 $^3J_{NC}$ and $^3J_{CC}$ Coupling Constants[a], Stereoassignments[b], and χ_1 Angles for Valine, Isoleucine and Threonine Residues of HSF.

Residue	Chemical Shift (ppm)	$^3J_{NC\gamma}$ (Hz)	$^3J_{C'C\gamma}$ (Hz)	Residue	Chemical Shift (ppm)	$^3J_{NC\gamma}$ (Hz)	$^3J_{C'C\gamma}$ (Hz)	χ_1 NMR (± 20°)	χ_1 X-ray (°)
V6 γ1	21.6(0.90)	>1.6[c]	<1.0	V6 γ2	23.0(0.80)	<1.0	~3[c]	180	180
V13 γ1	21.1(-0.41)	1.8 ± 0.1	<1.0	V13 γ2	22.2(0.13)	1.2 ± 0.1	3.4 ± 0.3	~180[d]	74
V32 γ1	20.5(0.60)	1.7 ± 0.1	~1.0	V32 γ2	20.8(0.74)	<1.0	~3[c]	180	-176
V33 γ1	21.1(0.76)	~1.8[c]	<1.0	V33 γ2	20.2(0.49)	<0.5	3.8 ± 0.4	180	177
V40 γ1	21.6(0.96)	nd[e]	<1.0	V40 γ2	24.7(1.11)	<1.0	~3.7[c]	180	165
V43 γ1	23.0(0.87)	~1.6[c]	<1.0	V43 γ1	23.5(0.84)	<1.0	~3[c]	180	177
V57 γ1	21.6(0.96)	~1.8[c]	<1.0	V57 γ2	23.9(0.27)	nd	3.5± 0.3	180	174
I22 γm	14.1(0.88)	~1.0[f]	<1.0					180	-179
I31 γm	17.5(0.34)	2.2 ± 0.1	<1.0					-60	-64
T26 γm	22.1(1.27)	1.6 ± 0.1	<1.0					-60	58

[a] Coupling constants were determined from spin-echo difference ct-HSQC spectra of HSF (Grzesiek et al., 1993; Vuister et al., 1993). The precision in the coupling constants was determined by calculating RMS differences in intensities of leucine and isoleucine Cδ methyl groups. The coupling constants have been corrected for a 5-7% ($^3J_{CC}$) and 7-10% ($^3J_{NC}$) underestimate due to systematic errors inherent in the method by multiplication of the measured values by 1.06 ($^3J_{CC}$) and 1.08 ($^3J_{NC}$). Systematic errors increase linearly with the size of the coupling constant.

[b] Chemical shifts of ^{13}C and ^1H (in parentheses) are referenced to TSP. Stereoassignments were not determined for Val-68 or Val-71 due to overlap.

[c] Estimate due to partial overlap

[d] Compared to the other valines, the larger than normal value for V13 γ2 $^3J_{NC}$ suggests some degree of rotamer averaging. This measurement was not included as a constraint in structure calculations.

[e] Not determined due to overlap.

[f] Estimate. Could not be integrated due to partial overlap with a t_1 streak.

this region does not conform to a regular pattern of secondary structure and is likely to be unstructured in solution.

Stereoassignments and $\chi 1$ angles for Val, Leu, and Ile residues

Stereoassignments and χ_1 angles for 7 of the 9 valines, both isoleucines, and the single threonine (Table 2.1) were determined on the basis of $^3J_{NC\gamma}$ and $^3J_{C'C\gamma}$ coupling constants (Grzesiek et al., 1993; Vuister et al., 1993). For the valine residues the relatively large $^3J_{NC\gamma 1}$ and $^3J_{C'C\gamma 2}$ coupling constants and the comparatively small $^3J_{NC\gamma 2}$ and $^3J_{C'C\gamma 1}$ coupling constants indicate that each adopts the *trans* configuration. The $^3J_{NC\gamma 2}$ coupling for Val-13 is slightly larger than for the other valines which may reflect some conformational averaging or skewing of χ_1 toward the *gauche*-+ conformation. By comparison, in peptides $^3J_{C'C\gamma}$ for valine typically falls in the range of 2 - 2.5 Hz, which is significantly different than the measured values of $^3J_{C'C\gamma 2}$ (3.4 Hz) and $^3J_{C'C\gamma 1}$ (<1 Hz) for Val-13. These data together with a value of 1.7 Hz for $^3J_{NC\gamma 1}$ show that although there may be some averaging or skewing toward *gauch*-+, Val-13 exists predominantly as the *trans* rotamer. Similar arguments were used to determine stereospecific assignments and χ_1 angles for Ile-22, Ile-31, and Thr-26. Where possible (including Val-13) intraresidue $d_{N\gamma m}$ and $d_{\alpha\gamma m}$ NOEs were used to confirm the χ_1 determinations (Zuiderweg, et al., 1985).

Tertiary Structure

To study the tertiary structure of HSF, distance and dihedral constraints determined from the NMR data were used in a hybrid distance geometry/simulated annealing protocol (Nilges, et al., 1988; Brünger, 1992). The final structures were calculated on the basis of 175 structurally significant intra-residue, 211 sequential, 93 medium-range, and 104 long-range distance constraints, 35 ϕ and 9 χ_1 dihedral restraints, and 30 hydrogen bond constraints (15 hydrogen bonds). The distribution of NOE constraints is shown in Fig.

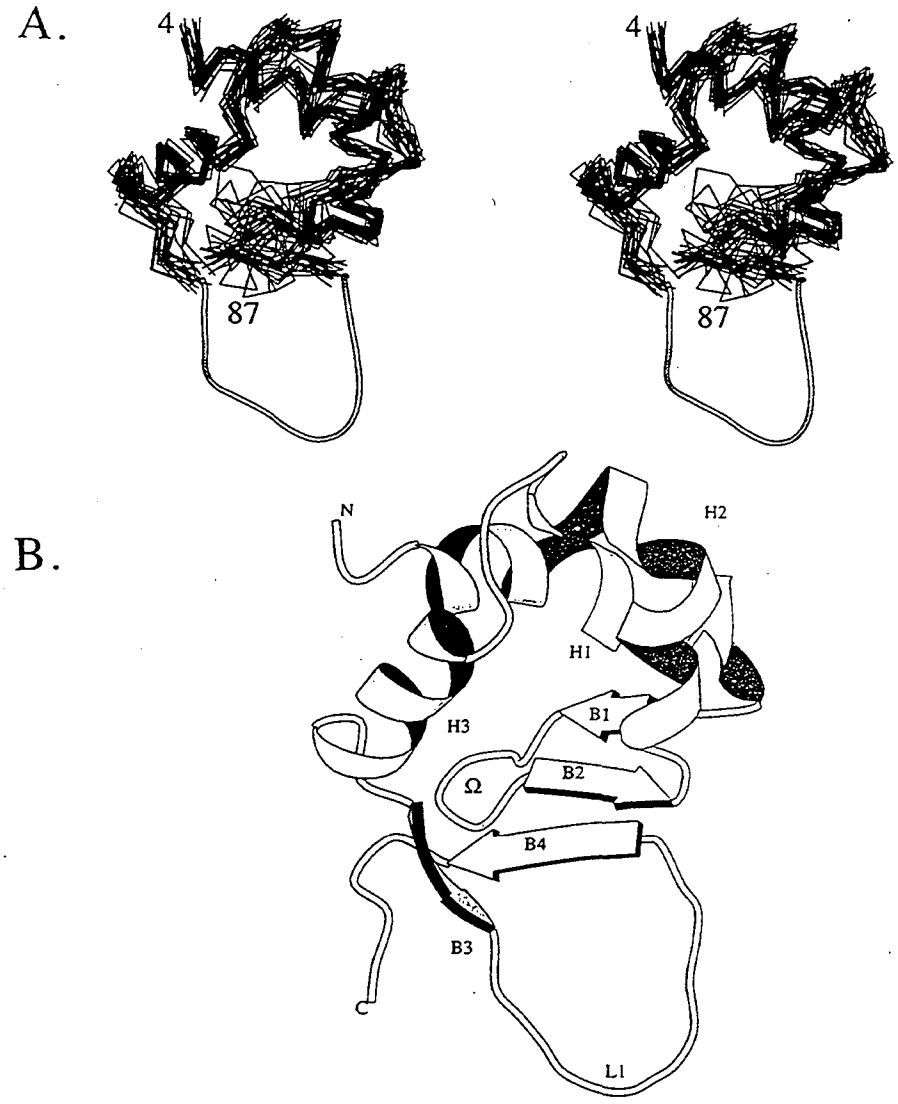


Figure 2.12 C_α coordinates of the Heat Shock Factor Solution Structures.
A. Stereoview of the superposed coordinates for 24 accepted structures of HSF. Shown are C_α coordinates for residues 4-68 and 83-87 along with a tube indicating the position of the C_α atoms of the poorly defined loop (L1) for one of the 24 structures, For the alignment, only backbone coordinates of residues in well-defined segments of structure (4-25, 31-68, 84-86) were used. RMSD for these backbone coordinates is 1.03 ± 0.19 Å. Ala-4 and Glu-87 are labeled to indicate the termini of the chain trace. **B.** Ribbon diagram of HSF showing the relative orientations of the three α-helices and the four-stranded antiparallel β-sheet. The structure with the lowest target function energy at the end of the last X-PLOR refinement is represented. The helices and sheets and two loops are labeled as defined in Fig 2.10. Note the pronounced kink in helix H2 due to the proline insertion, and the rotation of β-strand B3 out of the plane of rest of the β-sheet. the N- and C- termini are also labeled. The figure was created with MOLSCRIPT software (Kraulis, 1991).

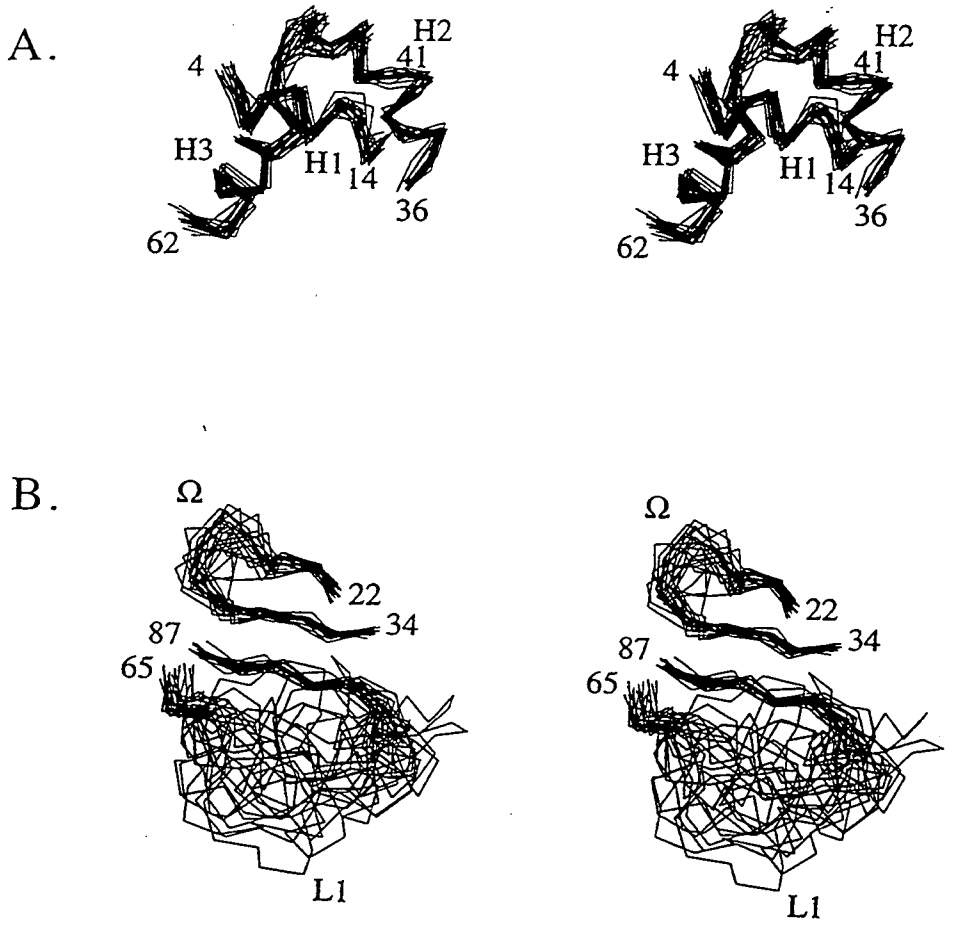


Figure 2.13 Superposed C_α coordinates for regions of secondary structure in HSF. **A.** Stereoview of the superposed coordinates of the three-helix bundle for the 24 accepted structures. For the alignment, the backbone coordinates of the residues in the three helices and the turn connecting helices H2 and H3 (residues 4-14, 36-62) were used. RMSD of these backbone atoms relative to those of the mean for superposing this region was 0.81 ± 0.15 Å. The helices are labeled as defined in 4.11. The N- and C-termini and Gln-41 are labeled by residue number. The central bulge is clearly visible in helix H2 near Gln-41. **B.** Stereoview of the superimposed C_α coordinates of the β-sheet and connecting loops for the 24 accepted structures. Structures were aligned with the mean structure of this region using backbone coordinates for the four strands (residues 22-24, 31-34, 65-68 and 84-86) (RMSD for these coordinates is 0.99 ± 0.27 Å). The omega-loop connecting β-strands B3 and B4 is clearly somewhat ill-defined and the extended loop connecting β-strands B3 and B4 is highly undefined. The N- and C-terminal amino acids of the four termini are labeled by residue number.

2.14A. The H-bond assignments accounted for 15 of the 34 most slowly exchanging amides. Four H-bonds were assigned in the β-sheet and four and seven H-bonds respectively were assigned in helices H1 and H3. Because of the distortion identified in helix H2, which may cause non-regular H-bonding patterns, no H-bonds were assigned for this helix. Of the 30 structures calculated, 24 had acceptably low energies and no large distance (>5 Å) or torsion angle violations (>5°) (see materials and methods).

The best fit superposition of the 24 structures is shown in Fig. 2.12A. The root mean square deviation relative to the mean <RMSD> for the backbone positions of the protein core (residues 4-14, 31-68 and 84-86) was 1.03 ± 0.19 Å (Fig. 2.13A). This value should be compared to <RMSD>s of 0.4 to 0.5 Å for highly refined structures (Powers, et al., 1993; Lodi, et al., 1994). The present resolution of these structures is sufficient, however, to define the backbone coordinates of HSF (Fig. 2.12B) and improved with the addition of further long range NOEs, stereoassignment of the β-protons and additional χ_1 dihedral constraints pursued in a later stage of structure determination (Chapter 3).

The protein topology is as follows: helix H1; a helical/turn region (residues Lys-16 - Lys-20) ; β-strand B1, a short Ω–loop (residues Ser-25 - Ser - 30); β-strand B2; helix H2; helix H3; β-strand B3; extended loop L1, and β-strand B4. As can be seen in Fig. 2.12B, the β-sheet forms a platform against which the C-termini of helices H1 and H3 and the N-terminus of helix H2 rest. The three amphipathic helices together form a three-helix bundle with the hydrophobic faces packing against one another and against the hydrophobic face of the antiparallel β-sheet . The β-strand B3 is significantly rotated away from the plane formed by β-strands B1, B2, and B4. In addition, the loop L1 extends away from the outside face of the β-sheet.

Superpositions for the three helix bundle (<RMSD> of 0.81 ± 0.15 Å) and the four-stranded antiparallel β-sheet (<RMSD> of 0.99 ± 0.27 Å) are shown in Figures

2.13A and 2.13B, respectively. The distortion in H2, identified previously, is clearly evident near Gln-41. It can also be seen that H2 and H3 are connected by a short segment such that the axis of the C-terminal portion of H2 lies almost perpendicular to the axis of H3. It is clear from Figures 2.13A and 2.13B that the three helix bundle and the β-sheet are well defined and that the Ω–loop connecting strands B1 and B2, and the extended loop L1 connecting B3 and B4, are less well defined. This is confirmed by examining the <RMSD> of backbone coordinates and the ϕ angular order parameters (S(ϕ)) (Hyberts, et al., 1992) as a function of sequence (Fig. 2.14B & 2.14C). The N- and C- termini, the Ω-loop, and the extended loop (L1) all show <RMSD>s greater than 2 Å and S(ϕ) values less than 0.6. These regions were excluded from the overall superposition. Segments which showed moderate <RMSD>s and S(ϕ) values which were included in the superposition were the helical/turn region between H1 and B1 (Lys-16 - Lys-20) (<RMSD> of 1.12 ± 0.19 Å and average S(ϕ) of 0.85 ± 0.17)), the turn between H2 and H3 (Lys-49 - Phe-53) (<RMSD> of 1.05 ± 0.27 Å and S(ϕ) of 0.73 ± 0.27), and the distorted β-strand B3 (Trp-65 - Val-68) which shows the largest <RMSD>s (1.88 ± 0.60 Å) and smallest S(ϕ) (0.63 ± 0.24). Indeed when this strand is excluded from the superposition of the β-sheet backbone coordinates the <RMSD> is reduced from 0.99 ± 0.27 Å to 0.54 ± 0.11 Å . This region shows a greater range of conformations in the family of structures (Fig. 2.13B) because there are fewer long range NOE distance constraints between strand B3 and other regions of the structure (Fig. 2.14A) as compared to the other β-strands.

Discussion

Overall fold and Identification of the helix-turn-helix motif

Using a combination of 3D triple-resonance CBCANH and CBCA(CO)NH (Grzesiek and Bax, 1992a; Grzesiek and Bax, 1992b), double-resonance HCCH-TOCSY (Clore, et al., 1990; Bax, et al., 1990b) , 3D ^{15}N TOCSY-HMQC (Driscoll, et al., 1990), and various

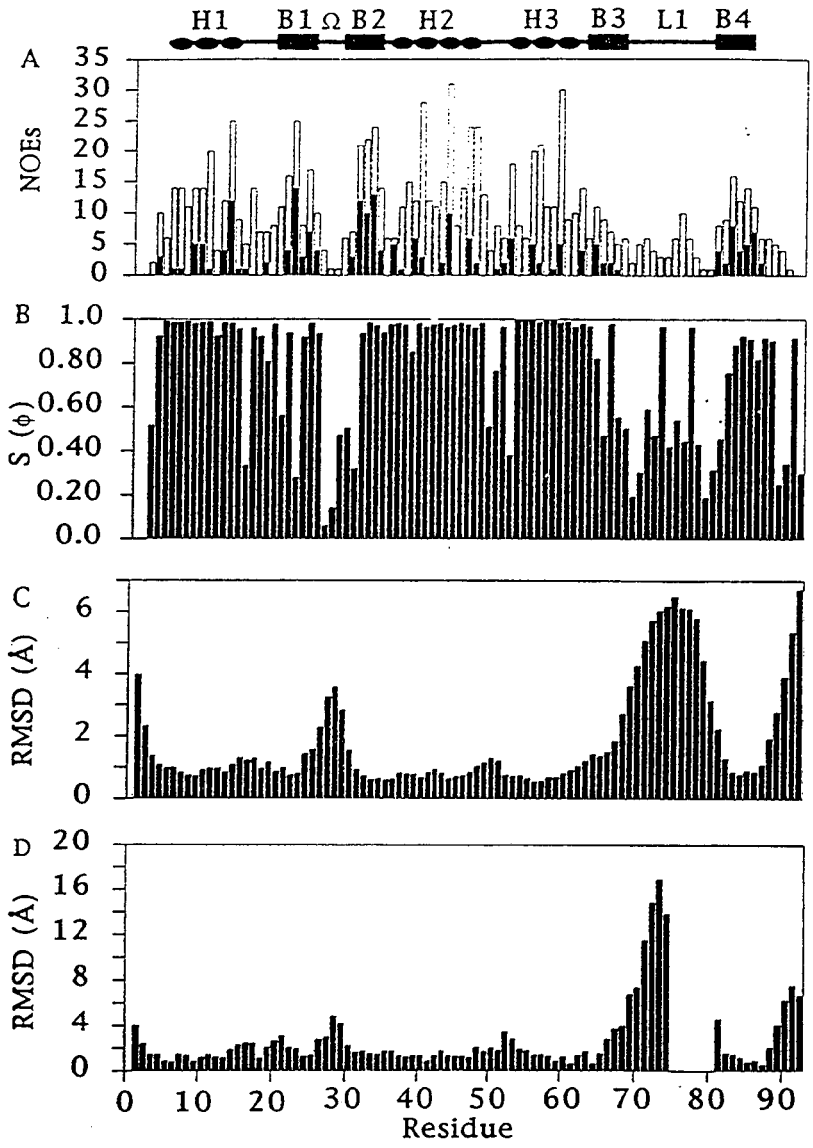


Figure 2.14. NOEs, S(ϕ), and RMSDs by residue. **A**. Distribution of NOE constraints for HSF. The height of each bar represents the total number of NOEs identified for that residue. Shaded bars represent long-range NOEs. For each constraint, the originating and destination residues were each counted once. **B**. The angular order parameter S^{angle} for ϕ dihedral angles as a function of residue for the 24 accepted structures. An S value of 0.9 corresponds to a standard deviation in the angle of ±24° (Hyberts et al, 1992). **C**. RMS differences of backbone (N, CA, C, O) coordinates for the family of accepted structures relative to the average structure. The superposition included residues 4-24, 31-68, and 84-86. The omega-loop (25-30) and the extended loop L1 (residues 69-83) show large deviations and are poorly defined by the NMR constraints. The β-strand B3 (65-68) also shows larger deviations than other regions of secondary structure. **D**. RMS differences of backbone coordinates for the family of 24 accepted structures relative to the 1.8 Å X-ray structure. Superposition as for Fig 4.14C.

2D experiments, nearly complete ^{1}H, ^{15}N and ^{13}C assignments were obtained for the 92 residue DNA-binding domain of HSF from *K. lactis* (Appendix A). The CBCANH, CBCA(CO)NH and HCCH-TOCSY experiments rely primarily on strong one and two-bond coupling constants which allow for assignment of the NMR signals without reference to the secondary structure of the protein as in more traditional methods (Wüthrich, 1986). In addition, for HSF, this combination of experiments provided for the relatively efficient assignment of the NMR signals with a minimal use of spectrometer time (all of the above mentioned 3D experiments were collected in approximately nine days).

Analysis of 2D NOESY, 3D ^{15}N- and ^{13}C-separated NOESY-HMQC spectra, a 4D $^{13}C/^{13}C$-edited NOESY spectrum, and D_2O exchange spectra, yielded a total of 583 NOE distance constraints and 15 H-bonds (30 H-bond constraints) which were used to characterize the secondary and tertiary structure of HSF. Analysis of HMQC-J (Kay and Bax, 1990) and spin-echo difference ct-HSQC spectra (Grzesiek, et al., 1993; Vuister, et al., 1993) also provided important ϕ and χ_1 torsion angle restraints as well as stereospecific assignments for many of the C_γ methyl groups (Table 2.1). Together, the distance and dihedral constraints were sufficient to define the conformation of HSF using a hybrid distance geometry-simulated annealing protocol (Brünger, 1992).

Unlike other DNA-binding proteins characterized to date which typically bind as monomers or dimers, HSF binds to DNA as a homotrimer. Specifically, a C-terminal triple-stranded coiled coil (Sorger & Nelson, 1989; Peteranderl and Nelson, 1992) directs three N-terminal DNA binding units consisting of approximately 90 residues (Flick, et al., 1994) to a series of 5 base-pair inverted repeats of the form nGAAn (Amin, et al., 1988; Xiao & Lis, 1988). Most of the residues which are conserved among all known HSFs (Harrison, et al., 1994) are located in helices H2 and H3 (Fig. 2.3B) and include all of the surface sidechains of helix H3. The high degree of conservation of the residues in these helices combined with their arrangement in the structure (Fig. 2.12) suggest that these

helices constitute a DNA-binding region and that HSF is a member of the helix-turn-helix (HTH) family of DNA-binding proteins. Similar suggestions have been made based on the location of secondary structure elements in *D. melanogaster* HSF (Vuister, et al., 1994b) and a crystal structure of *K. lactis* HSF (Harrison, et al., 1994). Moreover the NMR data show that the secondary structure and topology (α1–β1–β2–α2–α3–β3–β4) of HSF and CAP are almost identical. Except for an additional β-strand (B1) this structure also resembles that of the HTH eukaryotic transcription factor hepatocyte nuclear factor/*forkhead* (HNF-3γ) (Clark, et al., 1993) as suggested previously (Harrison, et al., 1994; Vuister, et al., 1994b).

Unusual Structural Features of HSF

Although the HTH is a familiar DNA-binding motif, two novel features were found in HSF which add to the considerable variations of the HTH observed in other proteins (Harrison and Aggarwal, 1990; Steitz, 1990; Brennan, 1993; Feng, et al., 1994). One such feature of HSF is the irregular structure in the central portion of helix H2 as evidenced by $^{3}J_{NH\alpha}$ coupling constants, $^{13}C_{\alpha}$ chemical shifts and NOE data (Fig. 2.10), and the presence of a conserved proline (Pro-45) toward the C-terminus of H2. The structure calculations show a discontinuity in helix H2 between residues Glu-42 and Pro-45 (Fig. 2.12). Typically, the presence of a proline residue in the central portion of a helix produces a slight bend but does not significantly distort the ϕ/ψ angles of residues on either side of the bend (Barlow & Thornton, 1988). The distortion observed in HSF appears to be atypical with distortions on either side of the proline. This distortion is investigated further in the high resolution structure of HSF described in Chapter 3. Another unusual feature is the ill-defined loop (L1) between residues Gln-69 and Lys-83 that connects β-strands 3 and 4 (Fig. 2.13B). This region, lacking in CAP, is reminiscent of loop L1 that connects β-strands 2 and 3 in HNF-3γ (Clark, et al., 1993) which makes important contacts to DNA in the HNF-3γ -DNA complex.

An interesting feature of HSF is that helix H3 terminates with a Shellman-like CAP sequence (Shellman, 1980; Aurora, et al., 1994) [C3(Leu-60)-C2-C1-Ccap-C'-C"-C"'(Trp-65)] except that the C1 position is occupied by an apolar residue (Met-62). This termination sequence is characterized by an apolar sidechain contact between C3 and C" (confirmed by sidechain NOEs between Leu-60 and Leu-65), two H-bonds from C1 to C' and C2 to C" (confirmed by slowed amide proton exchange of Gly-64 (fully exchanged in just under 370 minutes) and Trp-65), and positive ϕ angle for the Gly (average ϕ in NMR family of +88°, S(ϕ) of 0.8).

Hydrophobic Sidechain Packing

Sidechains whose packing is crucial to preserving the overall fold of HSF should correspond to conserved hydrophobic residues in the sequence. The interior face of the β-sheet is made up of hydrophobic residues, most of which are conserved, while the exterior face is hydrophilic and poorly conserved. The orientation of H3 with respect to the β-sheet is defined by the contacts between their hydrophobic surfaces. This arrangement must be crucial to the functioning of the domain judging from the high degree of sequence conservation at the interface. The N-terminus of helix H1, whose axis is at right angles to that of H3, makes numerous contacts with β-strand B1 and to a lesser extent with β-strand B2. Helix H1 also makes many contacts to the other two helices. The orientation of helix H2 is determined primarily by sidechain contacts with H3 across the turn region of the HTH and a number of contacts to H1 and B4. More detailed investigation of specific sidechain-sidechain interactions is given in Chapter 3.

Proposed DNA Interactions and Relationship with bacterial sigma factors

Considering the DNA contacts seen in other HTH proteins complexed with DNA, it is very likely that the majority of DNA contacts in the HSF-DNA complex will be made by the exposed residues of helix H3 (the recognition helix) such as Ala-54, Ser-55, Arg-58,

Gln-59, Asn-61, and Met-62. The positions of several conserved positively charged residues in the solution structure suggest that the sidechains of Lys-8, Lys-46, and Lys-49 located above the N-terminus of H3 and Lys-67 below the C-terminus of H3 are important to the functioning of the domain. Because they are highly conserved (Fig. 2.3B), positively charged, and located so close to the recognition helix (H3) it is likely that these sidechains make important contacts to the phosphate backbone.

Although the evolutionary link between bacteria and yeast is very distant, it is interesting to note that the putative helix-turn-helix of the *E. coli* alternative sigma factor for heat shock (σ^{32}) shows 40% sequence identity (65% sequence similarity) with the HTH region of HSF (Fig. 2.4A) (based on BESTFIT alignment (Genetics Computer Group, 1991) of residues 253 to 272 of *E. coli* sigma-32 with residues 36 to 62 of *K. lactis* HSF). Comparing HSF sequences reported recently for several organisms (Harrison, et al., 1994) using the same alignment gives an average of 32% sequence identity/ 50% sequence similarity. Moreover, the -35 promoter consensus of σ^{32}, TNtCNcCTTGAA (Cowing, et al., 1985), contains the sequence GAA (Fig. 2.1A), and genetic evidence derived from σ^{70} (a homolog of σ^{32}) implicate Arg-268 and Ala-264 in the putative recognition helix of σ^{32} as making contacts with the G and the last A respectively of the GAA sequence (Fig. 2.4B) (Gardella, et al., 1989; Siegele, et al., 1989; Dombroski, et al., 1992). The two homologous residues in HSF, Arg-58 and Ala-54 are on the exposed face of the recognition helix H3 and may make sequence specific contacts to the GAA sequences within the HSEs in a fashion analogous to that proposed for σ^{32} (Fig. 2.4B). Confirmation of this hypothesis must await further genetic and biophysical characterization of the HSF-DNA complex.

Comparison to other Structural Studies

After the signal assignments for HSF had been completed and the secondary structure had been defined, a 1.8 Å X-ray derived model of the protein was solved and

made available (Harrison, et al., 1994). The X-ray data were collected from crystals grown in 0.2 M Ammonium Acetate/0.1 M Acetate pH 4.6/30% PEG 4000 at 8°C (Harrison, 1994). It should be noted that through the use of the heteronuclear experiments described herein it was possible to make assignments and determine the overall fold of the protein without reference to the crystal structure. The topology of the X-ray model is similar to the model derived from the NMR data. Both structures contain a four-stranded antiparallel β-sheet and a three-helix bundle. An irregular segment defined as an α-helical bulge was identified in helix H2 and is consistent with the NMR results. The loop (L1) with ill-defined structure connecting β-strands B3 and B4 (Fig. 2.12A and Fig. 2.13B) is consistent with the results of the X-ray model, which shows larger B-factors in this region than in the rest of the structure, and absence of electron density for residues 76 to 79 (Harrison, et al., 1994).

The solution structure of the DBD of HSF from *Drosophila melanogaster* (dHSF) (residues G43-K148) determined by multidimensional NMR in solution (Vuister, et al., 1994b) agrees with all the salient features identified in *K. lactis* HSF with a 3-helix bundle sandwitched against a small four stranded antiparallel β-sheet. The distortion in helix H2 is present and the extended loop shows a similarly poorly defined conformation. One additional feature is found at the C-terminus. The construct used for structural studies of dHSF extends further toward the C-terminus of the HSF sequence (an extra 15 residues) than that used for the structural work conducted on *K. lactis*. This C-terminal sequence forms a poorly ordered helical turn region which makes contacts to the backside of the β–sheet. Further structural comparison with dHSF is made in Chapter 3.

RMSD comparison to X-ray model

At the current stage of refinement, it is appropriate to compare the overall fold of the family of NMR structures with the X-ray structure and to compare in more detail the distance constraints used in the X-PLOR calculations with those predicted from the crystal

structure. Superposition of the backbone coordinates for all residues except the Ω-loop (25-30), the loop L1 (69-83) and the first three and last 6 residues onto the X-ray model reveals good overall agreement with an RMSD of 1.77 ± 0.19 Å (Fig. 2.14D). Comparison of the β-sheet coordinates reveals similarly good agreement (RMSD of 1.65 ± 0.24 Å). If β-strand B3 is excluded (residues 65-68) the agreement is significantly better (RMSD of 0.85 ± 0.14 Å) indicating that the orientation of B3 deviates from the x-ray model. As stated previously this β-strand is not precisely defined (Fig. 2.13B) due to the small number of cross strand constraints connecting it to the rest of the structure (Fig. 2.14A). The three helix bundle also shows good agreement with the crystal structure with an RMSD of 1.50 ± 0.18 Å for backbone coordinates. The similarity of the NMR and X-ray results is significant given that the solvent content of the crystals used in the X-ray analysis was unusually low (~30%). High resolution solution structural studies of HSF (described in Chapter 3) indicate that this agreement is even closer than indicated here.

Backbone NOE comparison to X-ray model

To identify possible differences between the solution and crystal structures, we compared the backbone distances derived from NOEs and ϕ angles derived from $^3J_{HNH\alpha}$ coupling constants with the corresponding distances and angles in the X-ray model. The comparison revealed that the constraints derived from the majority of backbone NOEs and all of the measured ϕ dihedral constraints were in agreement with the x-ray model. A few differences in measured and predicted backbone NOE intensities were noted which could be grouped into two classes: those where the observed NOE was weaker than expected compared to the crystal structure, and those where the NOE was stronger than expected. In the first class we found five examples; $d_{\alpha N}(5,6)$, $d_{\alpha N}(9,10)$, $d_{\alpha N}(40,41)$, $d_{\alpha N}(42,43)$, and $d_{NN}(15,16)$. For the four $d_{\alpha N}$ NOEs, a weak NOE was observed while a medium intensity NOE was expected (3.4-3.5 Å in the X-ray model) while for $d_{NN}(15,16)$ no NOE was observed although a weak NOE was expected (4.3 Å in the x-ray model). There are

many reasons why an NOE might be weaker, including exchange broadening or H_α saturation. In the present cases, the differences could not be explained by these reasons since other NOEs to these amide protons were observed and because these alpha protons all resonate far from H_2O (4.77 ppm). Hence, the data suggest true structural differences. In the second class, two differences were observed, namely $d_{\beta N}(13,14)$ and $d_{NN}(66,67)$. In both cases the observed NOE is of medium intensity but was predicted to be weak based on the X-ray coordinates (4.1Å, and 4.3 Å). These disagreements between the NMR data and x-ray results may reflect differences in the two structures although spin diffusion can not be ruled out. Taken together, the differences observed for the backbone are relatively small, although it is interesting to note that the locations of the "perturbed" sequential NOEs correspond roughly to locations of protein-protein contacts in the crystal. Asn-7 is involved in such contacts as are Asn-14, Gln-41, Glu-42 and His-66. We have also observed several sidechain NOEs with altered intensities involving residues in these regions of HSF. These apparent differences are analyzed further in the determination of the high resolution solution structure of HSF described in Chapter 3.

Comparison of $\chi 1$ angles to X-ray model

Out of the 10 measured χ_1 angles (Table 2.1), two show significant differences with those observed in the X-ray structure. The NMR data for Val-13 and Thr-26 indicate χ_1 values of 180° (with evidence for some rotamer averaging near 180°) and -60° degrees respectively, while in the X-ray structure the measured angles were 74° (Val-13) and 58° (Thr-26). Val-13 is located at the C-terminus of H1. The methyl groups point toward the interior of the protein and show NOEs to residues in β-strands B1 and B2. Several differences also exist between sidechain NMR distance constraints that involve Val-13 and residues close to it. One previously mentioned is the sequential sidechain-backbone NOE $d_{\beta N}(13,14)$ which the X-ray structure predicts to be weak but shows a medium intensity NOE. This result is consistent with the change in the sidechain conformation of Val-13 in

solution, placing Val-13 H_β closer to Asn-14 H_N than predicted from the χ_1 angle observed in the x-ray structure. The structural difference in this region is probably related to differences in sidechain packing rather than global backbone conformation, as evidenced by the close agreement of the backbone coordinates for the family of NMR structures with the crystal structure. These differences are examined further in Chapter 3. Thr-26 is located in the Ω-loop (Ser-25 - Ser-30), which is on the surface of the structure. This region shows no long range NOEs to other parts of the structure except for the two terminal residues of the Ω-loop (Ser-25 and Ser-30) which show NOEs to the neighboring β-strands. Consequently it appears less well defined in the NMR family of structures (Fig. 2.13B). In the crystal structure on the other hand, the Ω-loop is involved in extensive protein-protein interactions and has B-factors that are similar to the rest of the structure. It is possible that Thr-26 prefers one rotamer state in the solution structure due to local packing of its γ methyl group with other atoms in the Ω-loop or hydrogen bond interactions with the sidechain hydroxyl group and that this interaction is perturbed in the crystal by protein-protein contacts.

Summary of comparison to X-ray model

In conclusion, the backbone geometry of the x-ray structure is very similar to the solution structure although subtle differences appear to exist some of which probably relate to protein-protein contacts in the crystal. These differences are located at the end of helix H1 ($d_{\alpha N}(5,6)$, $d_{\alpha N}(9,10)$, $d_{\beta N}(13,14)$ and $d_{NN}(15,16)$ NOEs, and Val-13 χ_1), in the Ω-loop (Thr-26 χ_1), at the distortion of helix H2 ($d_{\alpha N}(40,41)$ and $d_{\alpha N}(42,43)$) and in the β-strand B3 $d_{NN}(66,67)$. To make a more detailed comparison and characterize the nature of these changes, particularly with respect to sidechain packing and distortions in the DNA-binding surface if they exist, a high resolution solution structure was determined which includes stereoassignments of the β-protons and a greater number of distance and χ_1 constraints (Chapter 3).

Conclusion

The NMR signal assignments and the solution structure of HSF presented here provide the basis for the design of new biochemical and biophysical experiments necessary for an understanding of HSF-DNA and HSF-HSF interactions and the effects of crystal packing on the protein structure. For example, the heteronuclear ^{15}N and ^{13}C assignments can be used to probe the dynamics of this DNA-binding domain (Chapter 3) and should also aid studies of larger HSF-DNA complexes through the use of 3D and 4D NMR experiments. The solution structure, and in particular the identification of a HTH motif, should also help direct mutagenesis studies aimed at the identification of residues important for complex formation. Moreover, further refinement of the solution structure should allow for the identification of any structural changes that occur in HSF upon DNA-binding.

Chapter 3 High resolution Solution Structure and Dynamics of HSF

Introduction

In chapter 2 the determination of the solution structure of the DBD of HSF from the yeast *Kluyveromyces lactis* at intermediate resolution was described. The resultant family of structures had a backbone <rmsd> of 1.03 ± 0.19 Å for ordered regions and allowed the global fold of the protein to be described. There have been two other structures of this domain reported; a crystal structure of the *K. lactis* DBD at 1.8 Å (Harrison, et al., 1994) and a solution structure of the DBD from *D. Melanogaster* (0.82 Å <rmsd> for backbone coordinates in ordered regions) (Vuister, et al., 1994a). These structural studies show that HSF DBD consists of a four stranded antiparallel beta sheet with a three-helix bundle packed against it. The topology (H1-B1-B2-H2-H3-B3-B4) and structure place HSF in the recently identified winged-helix class of DNA-binding proteins (Clark, et al., 1993). This group of proteins recognize their DNA target sites primarily using the helix-turn-helix (HTH) motif and consistent with this binding mode, the most conserved region of HSF DBD is helix H3 which corresponds to the recognition helix of the HTH motif. The structural studies of HSF revealed a number of unusual features to this HTH protein which add to the considerable variations already accumulated for this binding motif (Harrison and Aggarwal, 1990; Steitz, 1990; Brennan, 1993; Feng, et al., 1994). Helix H2 contains a completely conserved proline near its C-terminus which causes a kink in the helix axis and prevented HSF from being identified as a HTH protein based on sequence alignment. In addition, the interhelical angle between H2 and H3 is smaller than observed for most HTH proteins. Finally β-strands B3 and B4 are connected by an extended loop, analogous to one wing of the winged helix in HNF3-γ,which appears to be unstructured in solution based on an absense of long range NOEs. Further structural refinement of the DBD from *K. lactis* has been pursued to define the details of the structure at high resolution. This high

resolution solution structure determination will be described in this chapter along with the conclusions drawn from it.

An increase in the resolution of the HSF solution structure was accomplished using a number of heteronuclei-based experiments for stereoassignment of Hβ and methyl pairs along with a 4D $^{13}C/^{15}N$-edited NOESY-HMQC to define more NOE-based distance restraints, as well as further analysis of experiments described in Chapter 2 (Damberger, et al., 1994). The present family of structures fit all experimental constraints extremely well and have a backbone <rmsd> for structured regions of 0.81 ± 0.16 Å (0.61± 0.14 Å for regions of secondary structure). In addition due to the increased number and precision of the constraints (1073 total constraints) the sidechains in the interior of the protein are also well ordered (0.84 ± 0.15 Å <rmsd>). Greater resolution of sidechain interactions allowed the identification of interactions in the region near the proline insertion in helix H2. These packing interactions appear to be responsible for the strong sequence conservation of residues in this region of the protein interior. Dynamics data indicate that the extended loop region previously shown to have disordered conformations is dynamically disordered in solution. Structural comparisons are also made to the DBD from *D. Melanogaster* and to the X-ray structure of the DBD from *K. lactis* .

Materials and Methods

NMR spectroscopy

Uniformly ^{15}N and $^{13}C/^{15}N$ -labeled samples of HSF prepared as described in Chapter 2 were used for all experiments described in this Chapter. In all cases, the protein was dissolved in 10 mM phosphate buffer (pH 3.4). All spectra were recorded at 25° C on an AMX 600 NMR spectrometer equipped with a 3 channel interface and a ^{13}C-optimized $^{1}H,^{13}C,^{15}N$ triple-resonance probe or a tunable inverse probe. ^{1}H and ^{13}C signals are

referenced to TSP (Live, et al., 1984), and ^{15}N signals are referenced to liquid ammonia (Live, et al., 1984).

Stereoassignment experiments for Hβ protons

A 3D HNHB spectrum was collected as described (Archer, et al., 1991) (P.S. #27) on uniformly ^{15}N labeled HSF in 90% H_2O/10% D_2O. The acquired data matrix consisted of 60 (t_1, ^1H) x 30 (t_2,^{15}N) x 256 (t_3, ^1H) complex points with spectral widths of 8.33, 24.9 and 8.5 ppm in the three respective dimensions. The ^{15}N and ^1H carrier frequencies were 120.0 ppm and 4.77 ppm respectively. Solvent suppresion was achieved using ^1H saturation and ^{15}N decoupling during acquisition was implemented with the GARP decoupling sequence (Shaka, et al., 1985). Data was processed as described (Archer, et al., 1991) resulting in a final data matrix consisting of 128 (F1) x 64 (F2) x 512 (F3) real points. In addition, a two dimensional ^1H-^1H HNHB spectrum was collected with spectral widths of 14.12 ppm in both dimensions (512 x 512 complex points).

A 3D ^{13}C-separated ROESY-HMQC experiment was collected as described (Clore, et al., 1991) (P.S. #29) on a uniformly ^{13}C/^{15}N labeled sample of HSF in 99.96% D_2O except that ^1H 180° refocussing pulses applied during ^{13}C evolution were replaced with compensated 180° pulses (Freeman, 1988a). The spinlock was applied as a single 40 ms pulse with field strength of 6.3 kHz and was positioned at 6.0 ppm (to reduce Hartman-Hahn matching artifacts between Hα and Hβ protons). Solvent suppresion was achieved using presaturation at the HDO frequency (4.77 ppm). GARP modulation of a 3.8 kHz rf field was used for ^{13}C decoupling during acquisition. The acquired data matrix consisted of 124 (t_1, ^1H) x 28 (t_2,^{13}C) x 256 (t_3, ^1H) complex points with spectral widths of 10.93 (^1H), 23.8 (^{13}C) and 15.14 (^1H) ppm in the three respective dimensions. The ^{13}C and ^1H carrier frequencies were set to 43.0 ppm and 6.00 ppm respectively. The data were

processed as described (Clore, et al., 1991) . The final data matrix contained 256 (F1) x 64 (F2) x 512 (F3) real points.

A 3D ^{15}N-separated TOCSY-HMQC spectrum was collected as described (Clore, et al., 1991) (P.S. #28) on a uniformly $^{13}C/^{15}N$ labeled sample of HSF in 90% H_2O/10% D_2O. Mixing was achieved using DIPSI-2 modulation of a 10.4 kHz rf field applied for 38 ms. The acquired data matrix contained 128 (t_1, ^1H) x 32 (t_2,^{15}N) x 256 (t_3, ^1H) complex points with spectral widths of 11.00 (^1H), 24.9 (^{15}N) and 14.12 (^1H) ppm. The ^{15}N and ^1H carrier frequencies were set to 120.0 ppm and 4.77 ppm respectively. The data were processed as described (Clore, et al., 1991) . The final data matrix consisted of 256 (F1) x 64 (F2) x 512 (F3) real points.

Parameters and conditions used for the DQF-COSY experiment and the 3D ^{15}N-separated NOESY-HMQC experiment (100 ms mixing time) were described in Chapter 2.

Stereoassignment experiments for Hδ protons

$^3J_{CC}$ and $^3J_{CH}$ scalar couplings for leucine residues were determined from semiquantitative analysis of peak intensities in 2D LRCC (Bax, et al., 1992) (P.S. #24) and LRCH (Vuister & Bax, 1993) (P.S. #25, #26) experiments recorded on uniformly $^{13}C/^{15}N$ labeled HSF in D_2O as described. For both experiments the ^1H-^{13}C dephasing and rephasing delays were set to 1.7 ms and ^{13}C-^{13}C dephasing and rephasing delays were set to either 29.4 ms (LRCC) or 29.5 ms (LRCH). High power ^{13}C pulses were applied at a field strength of 13 kHz and ^{13}C decoupling was achieved through GARP modulation of a 4 kHz rf field. For the LRCC experiment, aliphatic resonances and carbonyl signals were selectively excited with 11.7 kHz and 1.7 kHz rf fields, respectively. In both experiments, the ^1H carrier was set to 4.77 ppm and the ^{13}C carrier was set to either 46 ppm (LRCC) or 43 ppm (LRCH). Spectral widths in the LRCC experiment were set to 12.2 ppm (^1H) and 85 ppm (^{13}C). A total of 512 (^1H) and 102 (^{13}C) complex

points were collected. Spectral widths in the two ^1H dimensions of the 2D LRCH spectrum were set to 12.6 ppm (t_2) and 11.5 ppm (t_1). A total of 256 (t_2) and 83 (t_1) complex points were collected. Both spectra were processed as described (Bax, et al., 1992; Vuister and Bax, 1993). The final data matrix for the LRCC experiment consisted of 1024 (^1H) and 512 (^{13}C) real points. The final data matrix for the LRCH experiment consisted of 512 (t_1) and 256 (t_2) real points.

4D $^{13}C/^{15}N$ edited NOESY-HMQC experiment

A 4D $^{13}C/^{15}N$-edited NOESY spectrum was recorded as described (Kay, et al., 1990) (P.S. #16) on a $^{13}C/^{15}N$ labeled sample of HSF dissolved in 90% H_2O/10% D_2O with a mixing time of 100 ms. The acquired data matrix consisted of 8 (t_1, ^{13}C) x 64 (t_2, ^1H) x 8 (t_3, ^{15}N) x 256 (t_4, ^1H) complex points with spectral widths of 23.8, 12.43, 16.9, and 12.43 ppm in the respective dimensions. The ^{13}C, ^{15}N, ^1H carriers were centered at 43.0, 117.8, and 4.77 ppm respectively. Solvent suppresion was accomplished with on resonance presaturation. Data were processed as described (Kay, et al., 1990) resulting in a final data matrix consisting of 32 (F1) x 128 (F2) x 32 (F3) x 256 (F4) real points.

Heteronuclear NOE experiment

2D ^1H-^{15}N Heteronuclear NOE spectra were acquired as described (Barbato, et al., 1992) (P.S. #21, #22) on uniformly ^{15}N labeled HSF in 90% H_2O/10% D_2O except that the water signal was suppressed using presaturation in the reference experiments. It should be noted that although it has been reported that saturation transfer can affect the results of NOE experiments by altering measured signal intensities in reference spectra (Li & Montelione, 1994) , this problem was not of concern here since spectra were collected at low pH (3.4) and consequently the exchange lifetime is much longer than the ^{15}N T_1

relaxation times. Data were collected with spectral widths of 20.0 ppm in t_1 (230 complex points) and 14.12 ppm in t_2 (1024 complex points) with the ^{15}N and ^{1}H carrier placed at 120.0 and 4.77 ppm respectively. Data were processed as described (Barbato, et al., 1992)

Hydrogen bond assignments

In addition to the hydrogen bond assignments made previously (Damberger, et al., 1994), new hydrogen bonds were assigned based on slowed amide exchange (amide proton exchanges in ≥ 370 min) and the occurrence of good hydrogen bonding geometry in over 80% of the converged structures in a given round of refinement . Good hydrogen bonding geometry was defined as follows: N-O distance ≤ 3.3 Å and N-H···O angle ≥ 120°. These hydrogen bonds were identified using a program written in C which read in the coordinate files for the converged (accept_) structures and compiled the frequency of occurrence of a given hydrogen bond pair. The program was repeated using a Unix script (hbondrev2.com) for each potential H-bond acceptor of a given slowly exchanging amide until all the amides were cycled through. The resultant output was diverted to a file for manual inspection (See Appendix D for code and output examples). New hydrogen bonds were assigned only if a given amide had one clear preference for the hydrogen bond acceptor.

Results

Determination of stereoassignments

Initial $\chi 1$ angle restraints and $H\beta$ stereoassignments were made via estimates of $^{3}J_{NH\beta 2}$, $^{3}J_{NH\beta 3}$, $^{3}J_{H\alpha H\beta 2}$, $^{3}J_{H\alpha H\beta 3}$ spin-spin coupling constants obtained from qualitative analysis of $H\alpha$-$H\beta$ cross peak intensities in a DQF-COSY spectrum and H_N-$H\beta$ cross peak intensities in a 3D HNHB spectrum. For example, in the DQF-COSY spectrum (Fig. 3.1) W24, Q59 and L44 each show one strong and one weak $H\alpha$-$H\beta$ correlation indicative of an ordered sidechain with one large and one small $^{3}J_{H\alpha H\beta}$

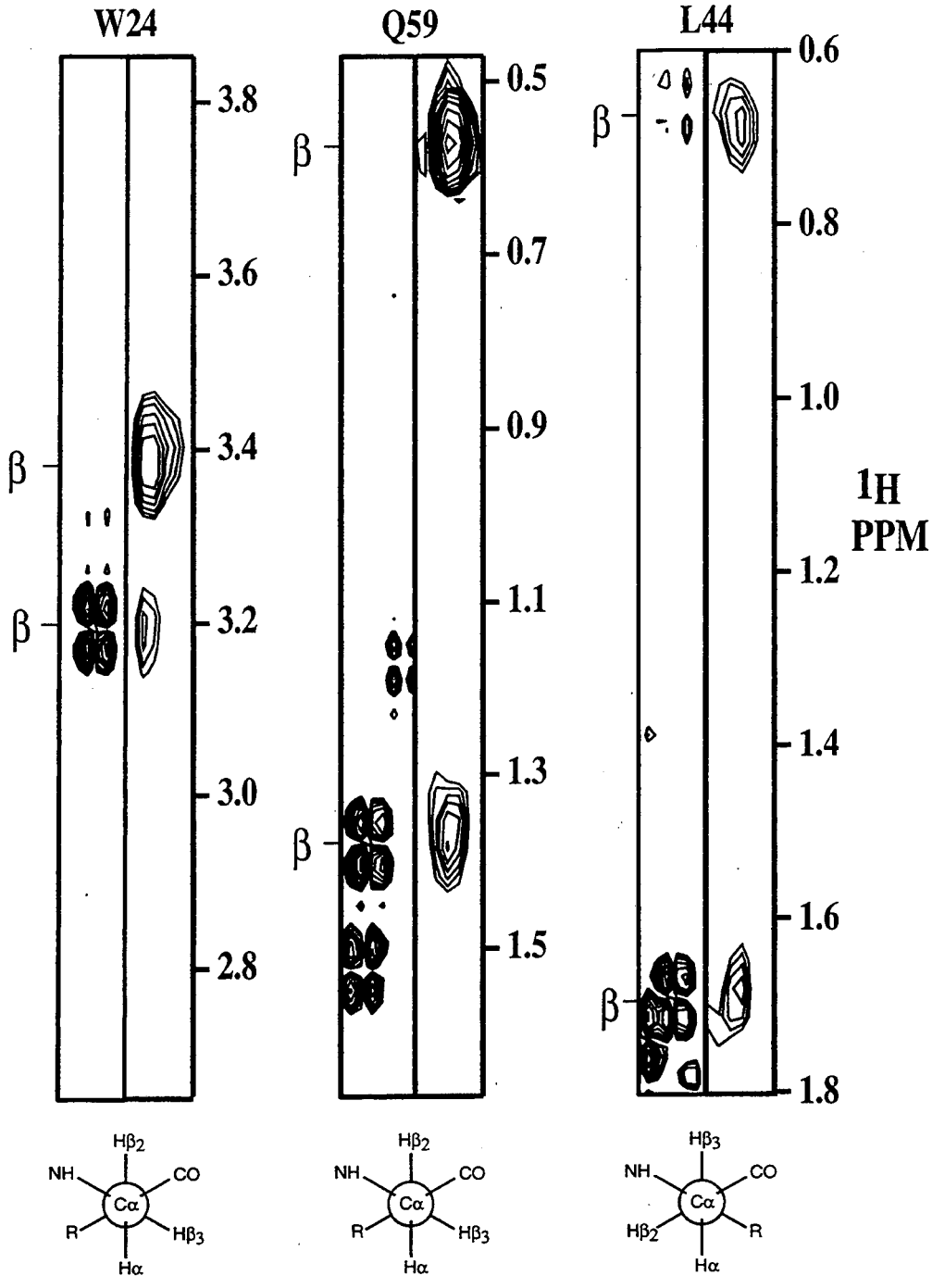


Figure 3.1 Strips from stereoassignment experiments. In each case the left strip is taken from DQF-COSY data; right strip from 3D HNHB spectrum. Strips corresponding to the regions containing the $H\beta$-$H\alpha$ and $H\beta$-H_N signals (for DQF-COSY and HNHB respectively) are shown for W24, Q59, and L44. Rotomer states consistent with crosspeak intensities observed in the two spectra are given below each pair of strips.

coupling. Similarly, in the HNHB spectrum, W24 and Q59 show one strong and one weak correlation indicative of one large and one small $^3J_{NH\beta}$ coupling. Comparison of these qualitative estimates of $^3JNH\beta$ and $^3J_{H\alpha H\beta}$ with those expected for the three $\chi 1$ rotomer states (Fig. 3.1) indicates that W24 and Q59 adopt the gauche- conformation with $\chi 1 = -60°$. For L44, both H_N-$H\beta$ correlations are of moderate intensity indicating small $^3J_{NH\beta}$ values. Together with the DQF-COSY results, these data indicate that L44 adopts the *trans* configuration ($\chi 1 = 180°$).

Determination of $\chi 1$ angles from this set of spectra is complicated by several factors. Amide exchange can affect the intensities of H_N-$H\beta$ cross peaks for different residues to varying degrees making classification of peak intensities difficult. The data were collected near the pH at which exchange is a minimum (pH 3.4), reducing adverse affects of exchange. Estimates of amide exchange rates determined previously were also referenced to insure accurate interpretations of cross peak intensities. Variations in the linewidths of $H\beta$ pairs can also distort the relative intensities of cross peaks in DQF-COSY spectra through cancellation of the antiphase multiplets. To avoid these artifacts, DQF-COSY data were used only when the linewidths of $H\beta$ pairs were found to be similar in 3D HCCH-TOCSY spectra .

Because of the complications noted above, the importance of distinguishing between ordered and non-ordered sidechain conformations, and problems of spectral overlap, $\chi 1$ angle restraints were confirmed using estimates of $^3J_{H\alpha H\beta}$ coupling constants obtained from a ^{15}N - separated TOCSY-HMQC spectrum (38 ms mixing time), both $d_{\alpha N}$ and $d_{\beta N}$ NOEs in a 3D ^{15}N-separated NOESY-HMQC (100ms) spectrum, and a 3D ^{13}C-separated ROESY-HMQC spectrum (40 ms) as outlined by Powers et al. (1993). Using these data $\chi 1$ angle restraints and $H\beta$ stereoassignments were determined for 21 of 25 residues with buried sidechains. In chapter 2, $^3J_{NC}$ and $^3J_{CC'}$ coupling constants were used to measure $\chi 1$ values and $H_{\gamma m}$ stereoassignments for 7 of the 9 Val residues, both Ile

residues, and the single Thr residue. Including the earlier results, a total of 25 χ_1 constraints (Appendix E) were used in the structure calculations described in this chapter. The peak intensities used to derive stereoassignments of $H\beta$ protons are listed in Appendix B along with obtained stereoassignments and χ_1 values.

Stereoassignment of $H\delta$ protons

$\chi 2$ angle restraints and $C\delta$ - methyl stereoassignments for three of the four leucine residues were obtained via semiquantitative estimates of $^3J_{C\delta C\alpha}$ and $^3J_{C\delta H\beta}$ coupling constants obtained from 2D LRCC (Bax, et al., 1992) and 2D LRCH (Vuister and Bax, 1993) spectra. For example, in the LRCC spectrum (Fig. 3.2) L9, L44, and L60 each show one strong (~ 3 Hz) and one weak $C\delta$-$C\alpha$ (< 1Hz) correlation, indicating that each residue adopts either the gauche+ or trans conformation. In the LRCH spectrum (Fig. 3.3) one methyl group for L9 and L60 shows a strong correlation to one $H\beta$ proton (L9 $H\beta 3$ and L60 $H\beta 2$). Together with the LRCC results, these data show that the $\chi 2$ angles of L9 and L60 adopt gauche+ and trans conformations, respectively and yield stereospecific assignments for the $C\delta$ resonances. Because of overlap in the 2D LRCH spectrum, the sidechain conformation of L44 was determined on the basis of LRCC data (Bax, et al., 1992) and $H\alpha$-$H\gamma$ and $H\alpha$-$H\delta$ cross peaks in 3D ^{13}C - separated NOESY and ROESY data as described (Powers, et al., 1993). The cross peak patterns for the remaining leucine (L77) showed cross peak patterns and $^3J_{CC}$ and $^3J_{CH}$ couplings characteristic of free rotation which is consistent with its location in the flexible loop L1 (see below). For the two isoleucine residues (I22 and I31) the strong $C\gamma 2$ - $C\alpha$ cross peaks observed in the LRCC spectrum (Fig. 3.2) indicate large $3JC\gamma 2C\alpha$ couplings which is only possible when $\chi 2$ adopts the trans conformation.

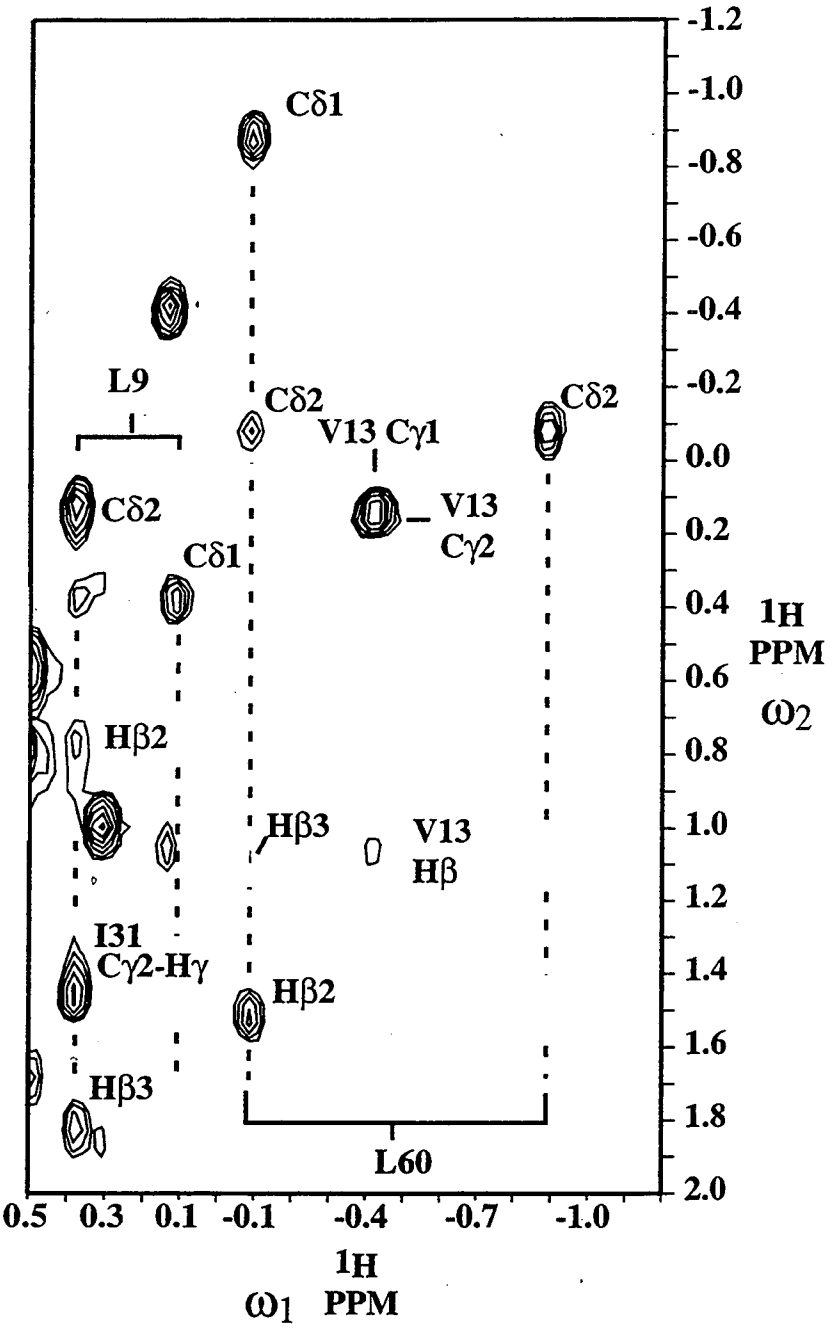


Figure 3.2 Portion from the 2D LRCH spectrum. Signals are shown for correlations between Hβ and Hδ nuclei modulated in intensity by the 3-bond coupling between the Hβ and Cδ nuclei. Signals are labeled by the assignment of the Hβ nucleus in ω_1 and the ^{13}C nucleus mediating the observed coupling intensity in ω_2.

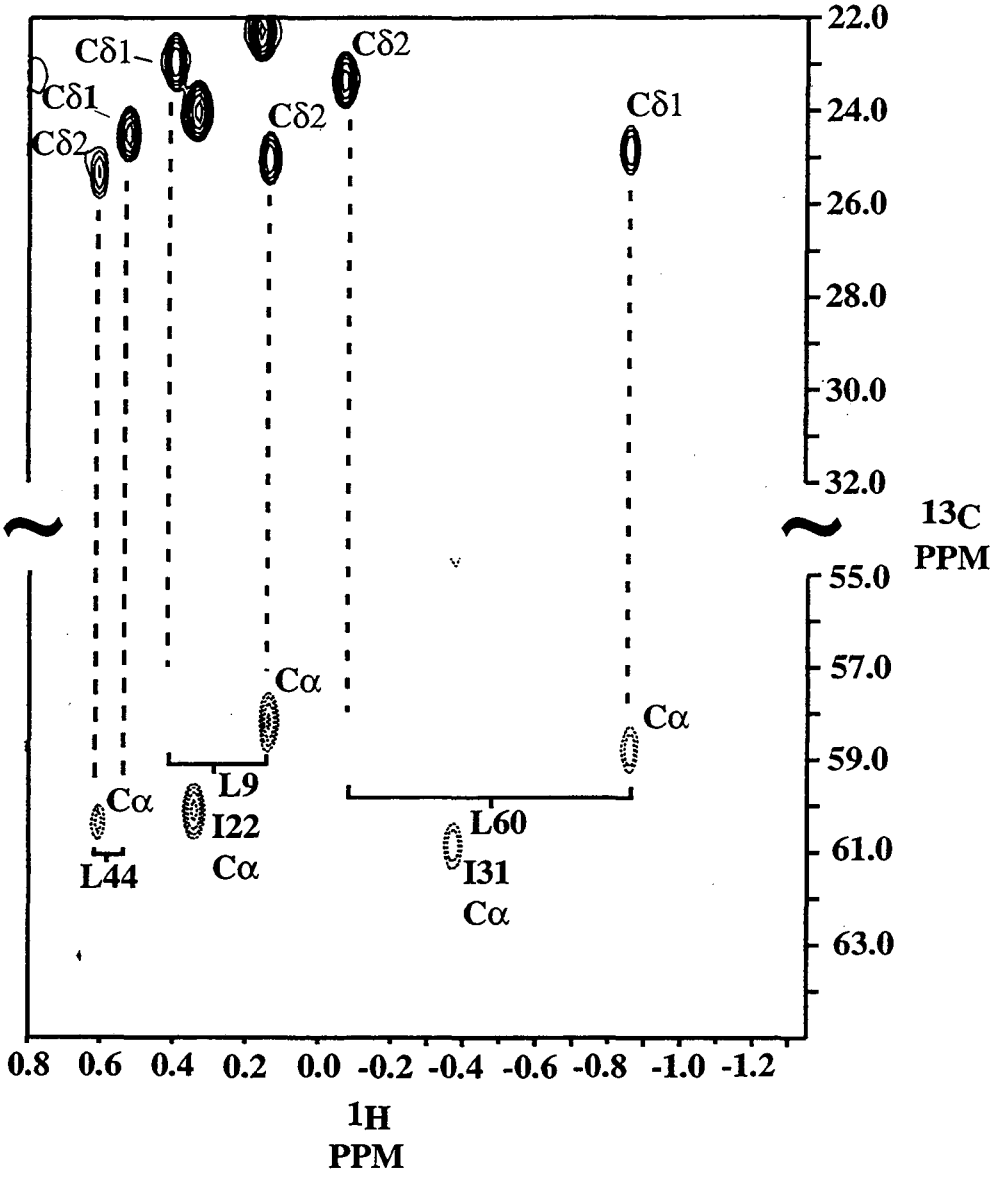


Figure 3.3 Extract from the 2D LRCC experiment. Two regions corresponding to H_α-C_α and H_α-C_δ correlations are shown. H_δ-C_α peaks are modulated in intensity by the magnitude of the C_α-C_δ 3-bond coupling constants, which give information about the stereoassignment of the two $(CH3)_\delta$ groups of leucine sidechains as well as the chi 2 angles of Ile sidechains.

Analysis of $^{13}C/^{13}C$ and $^{13}C/^{15}N$ edited NOESY-HMQC experiments

A list of all peaks in the $^{13}C/^{15}N$ separated NOESY-HMQC was generated using the automated peak-picking program CAPP (Garret, et al., 1991) (with manual editing using the PIPP program (Garret, et al., 1991) afterwards to remove peaks which were not of convincing intensity or were due to artifacts such as streaking or troughs from apodization. Software routines described in Appendix D were used to generate all possible NOE assignments based on matching chemical shifts of the crosspeaks to the chemical shift assignments (Appendix A). NOEs were assigned based on consistency with the structure family described in Chapter 2. A similar analysis was repeated for the $^{13}C/^{13}C$ separated NOESY-HMQC peak pick data to obtain further NOEs to use as distance constraints. These newly assigned NOEs were converted into structural constraints and included in structure refinement using the X-PLOR program (Brünger, 1982) to generate a new set of structures as previously described in chapter 2. The new family of structures was then used to identify additional NOEs from the peak picks and the process of structure refinement was iterated until no new NOEs were identified. This process in combination with the new stereoassignments allowed the identification of 384 new distance constraints along with the 20 new dihedral constraints for a total of 1073 NMR based constraints to define the structure (Appendix E).

Hydrogen Bond restraints

Of the 40 slowly exchanging amides (fully exchanged in D_2O in $\geq$ 370 min), 21 were identified in hydrogen bonds to specific backbone carbonyls (Appendix E and Appendix D - hbondrev2.com). All but two of these H-bonds are in regions of regular secondary structure identified in Chapter 2. A new H-bond, F21(HN)-(O)N18, occurs in the helical turn region between H1 and B1 (N14-F21). This region shows extensive amide

protection and a number of intermediate range NOEs characteristic of helical secondary structure (Fig. 2.10). In addition a new H-bond closes the Ω-loop, S25(HN)-(O)S30, and explains the slowed amide exchange of S25.

The new H-bonds identified in the β-sheet result in a complete network of H-bonds connecting all three strands of sheet, B1, B2 and B4. No new H-bonds were identified between the β-sheet and β-strand B3. The β-strand, B3, shows sparse NOE connectivity to the remainder of the sheet and appears rotated out of the plane defined by the other three strands. The slowly exchanging amide of V68 showed no evidence in the converged structure family of a preference for H-bonding with the carbonyl of E85 (the predicted partner based on the β-sheet topology) with 0% satisfying H-bond criteria. Because of the poorly defined backbone conformation in this region, it is difficult to make specific statements about the H-bond pairing of V68. In the crystal structure (Harrison, et al., 1994), the HN of V68 does not have a hydrogen bonding partner. The closest obvious candidate acceptor, O of E85, is 3.75 Å away. V68 HN is not able to participate in a hydrogen bond with E85 O because of the distortion in the sheet characterized by twisting of β-strand B3 out of the plane of the rest of the strands.

Helices H1 and H3 exhibit normal α-helix-type i,i+4 H-bonding patterns . Helix H2 which contains the proline insertion shows less regular H-bonding pattern, exhibiting an i,i+3 H-bond in the middle of the helix spanning P45, K46(HN)-(O)V43. Although not previously reported (Harrison, 1994) this hydrogen bond also appears to occur in the crystal structure (H-N···O angle 142°, H_N-O distance 2.08 Å). This makes up for one of the three i,i+4 hydrogen bonds which cannot be accomodated by the unusual geometry associated with P45 (namely P45(HN)-(O)Q41, K46(HN)-(O)E42, and Y47(HN)-(O)V43). dHSF does not appear to have this hydrogen bond (homologous residues L90(HN)-(O)L87) (Protein Data Bank Entry 1HKT).

Table 3.1 Structural statistics

	<SA>	$(SA)_r$	X-ray
rms deviations from experimental distance restraints (Å)			
all (967)	0.00994 ± 0.00012	0.01136	0.23099[a]
intraresiduc (li-jl=0) (438)	0.00749 ± 0.00249	0.00801	0.02032[b]
sequential (li-jl=1) (257)	0.01080 ± 0.00349	0.01916	0.22025[c]
short range (1<li-jl<6) (133)	0.01106 ± 0.00309	0.00431	0.41157[d]
long range (li-jl>5) (146)	0.00906 ± 0.00655	0.00189	0.31979[e]
H-bonds (42)	0.01433 ± 0.00284	0.00968	0.25214[f]
rmsd from experimental dihedral restraints (deg)	0.186 ± 0.110	0.107	12.59[g]
Lennard-Jones Energy[h] (kcal/mol)	23.8 ± 61.8	313.8	-649.0
X-PLOR energy F_{tot} (kcal/mol)	101.4 ± 8.5	102.1	92491.2[i]
F_{NOE} (kcal/mol)	5.17 ±2.35	3.04	2601.1
F_{tor} (kcal/mol)	0.18 ± 0.21	0.03	618.4
F_{repel} (kcal/mol)	6.80 ± 2.38	13.62	144.4
Deviations from idealized geometry			
bonds (Å)	0.002 ± 0.000	0.0014	
angles (deg)	0.427 ± 0.011	0.413	
impropers (deg)	0.267 ± 0.011	0.243	

[a]37 violations (>0.3Å) of 967 constraints.

[b]All violations are under 0.3 Å.

[c]Most significant violations are Leu 9 Hδ2 to Trp 10 Hζ3 (-1.8 Å) and Glu 37 Hγ* to Arg 38 Hα (-1.9 Å).

[d]Most significant violation is Val 6 Hγ1 to Trp 10 Hε1 (-3.5 Å)

[e]All violations under 1.5 Å.

[f]Most significant violation is 44 H_N to 40 O (-1.36 Å)

[g]Most significant violation is Thr 26 χ_1 (NMR = -60°, X-ray = +59.7°).

[h]Lennard-Jones energies were calculated using the CHARMm empirical energy function (Brooks et al., 1983) (files toph19.pro and param19.pro) within the program X-PLOR 3.1 (Brünger, 1986). 1-4 nonbonded interactions were excluded by setting the parameter NBXMod = 4. For NMR structures, a psf file was generated using toph19.pro and param19.pro. The vector statement: vector do (name = "HN") (name H) appears in the file used to calculate Lennard Jones energies along with an hbuild statement. This allowed computation of energies using the param19.pro forcefield in spite of the absense of non-polar hydrogen definitions. Positive Lennard-Jones energies were observed for many of the structures calculated using the protocals described in this chapter. Refinement performed after the completion with the final Van der Waals radii set to 0.80 times their normal values in the CHARMm empirical energy function resulted in structures with consistently negative Lennard-Jones energies. The default value for final Van der Waals radii in NMR refinement using the files dgsa and refine is 0.75 and was used for all calculations done in this dissertation. See Appendix D for further details (X-PLOR Macros for structure analysis).

[i]X-PLOR energies for the X-ray structure were calculated with the parallhdg.pro forcefield used during refinement. Van der Waals radii were set to 0.75 times their normal values in the CHARMm empirical energy function. Nonpolar hydrogens were built using the hbuild command within X-PLOR 3.1.

Table 3.2 Atomic rms differences[a]

	<SA> to SA[j]	<SA> to SA_r[k]	<SA> to X-ray[l]
ordered regions[b]	0.81 ± 0.16	0.91 ± 0.18	1.30 ± 0.11
2° structure[c]	0.61 ± 0.14	0.70 ± 0.14	1.05 ± 0.09
α-helices[d]	0.53 ± 0.13	0.58 ± 0.14	0.97 ± 0.11
Helix-turn-Helix[e]	0.50 ± 0.15	0.54 ± 0.14	0.81 ± 0.11
β-sheet[f]	0.57 ± 0.15	0.68 ± 0.17	0.99 ± 0.12
beta sheet-B3[g]	0.38 ± 0.07	0.54 ± 0.10	0.65 ± 0.08
20% buried[h]	0.84 ± 0.15	0.97 ± 0.18	1.43 ± 0.17
40% buried[i]	1.12 ± 0.18	1.28 ± 0.21	1.86 ± 0.14

[a] rms differences computed using X-PLOR 3.1 with avecycle.com described in Appendix E

[b] ordered regions consist of N,Cα,C',O for residues 4-24, 31-68, 84-86.

[c] secondary structured consisting of atoms N,Cα,C',O in residues 4-14, 22-24, 31-34, 36-48, 54-63, 65-68, 84-86.

[d] α-helix bundle consisting of atoms N,Cα,C',O for residues 4-14, 36-63.

[e] Helix-turn-Helix consisting of atoms N,Cα,C',O for residues 36-48, 54-63.

[f] β-sheet consisting of atoms N,Cα,C',O for residues 22-24, 31-34, 65-68, 84-86.

[g] β-sheet-B3 consisting of atoms N,Cα,C',O for residues 22-24, 31-34, 65-68, 84-86.

[h] 20% buried consists of heavy atoms in residues 20% buried in the crystal structure: 5, 9, 12-13, 18, 21-22, 25, 30-33, 39, 43-44, 47-48, 53, 56-57, 60, 64-65, 84, 86.

[i] 40% buried consists of heavy atoms in residues 40% buried in the crystal structure: 3, 5-6, 9, 12-13, 18, 21-25, 30-36, 39-40, 42-44, 47-48, 51, 53, 56-57, 59-60, 63-67, 82, 84-86, 88.

[j] <SA> represents the family of 30 accepted solution structures of *K. lactis* HSF DBD.

[k] SA represents the average coordinates for superposition of atoms described in the first column.

[l] X-ray represents the 1.8 Å crystal structure of *K. lactis* HSF DBD (Harrison et. al., 1994).

Structure calculations with new restraints

In the final round of structure refinement, 40 structures were generated of which 33 converged structures were chosen for structural comparison. These structures had target energies significantly lower than the structures described in Chapter 2. The previously reported family of structures had an average total energy of 150 ± 13.4 kcal/mol and average NOE and dihedral angle violation energies of 25.9 ± 5.4 kcal/mol and 0.96 kcal/mol. This constrasts with the present family of high resolution structures which have average total energies of 101 ± 9 kcal/mol and average NOE and dihedral angle violation energies of 5.2 ± 2.4 kcal/mol and 0.2 ± 0.2 kcal/mol respectively (Table 3.1). These energies are lower despite the nearly two-fold increase in the number of constraints (from 583 to 1073) emphasizing that the new restraints form a consistant set which results in converged lower violation energy structures. The structures show the expected distribution of phi-psi angles in a Ramachandran plot with most of the residues deviating from expected values occuring in unstructured regions.

The family of NMR structures show a high degree of precision in both the backbone (Fig. 3.4) and sidechain positions (Fig. 3.5) with <rmsd> for backbone coordinates of 0.81 ± 0.16 Å in structured regions ($S(\phi) = 0.90 \pm 0.18$) and 0.84 ± 0.15 Å <rmsd> for heavy atoms of buried sidechains (Fig. 3.5) ($S(\chi 1) = 0.84 \pm 0.26$, $S(\chi 2) = 0.87 \pm 0.22$). The sidechain of F53, which is buried in the crystal structure, shows an unusually high degree of disorder in the solution structure family ($S(\chi 1) = 0.54$). This is due to a lack of NOE constraints to the aromatic ring protons of this residue which are degenerate in both ^{13}C and ^{1}H chemical shifts (Appendix A) with those of the resonances of the neighboring residue F56. This degeneracy prevented the identification of NOEs between the two residues and from these residues to their mutual neighbors. Examining secondary structured regions, the precision of the structure family is even higher with 0.61

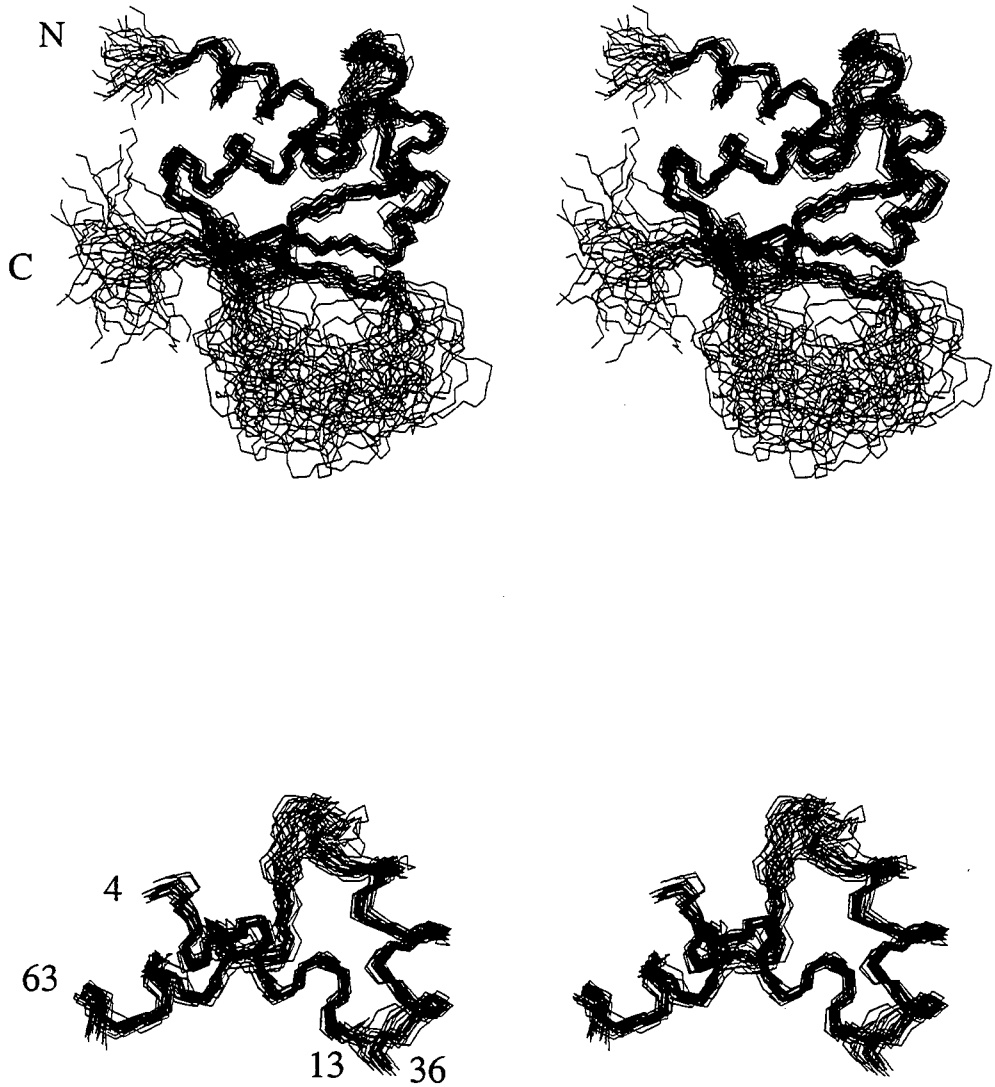


Figure 3.4 Backbone superposition of coordinates for HSF
A. Superposition of backbone atoms of all residues onto the mean coordinates for structured regions. The <rmsd> for these coordinates was 0.81 ± 0.16 Å. N- and C-termini are labeled with the letters N and C respectively.
B. Superposition of the backbone coordinates for the three helix bundle comprising residues 4-14, 36-48, and 54-63. These coordinates gave an <rmsd> of 0.53 ± 0.13 Å. Coordinates shown include residues 4-14 and 36-63. The termini are labeled by residue number. The distortion in helix H2 associated with the proline insertion is clearly visible.

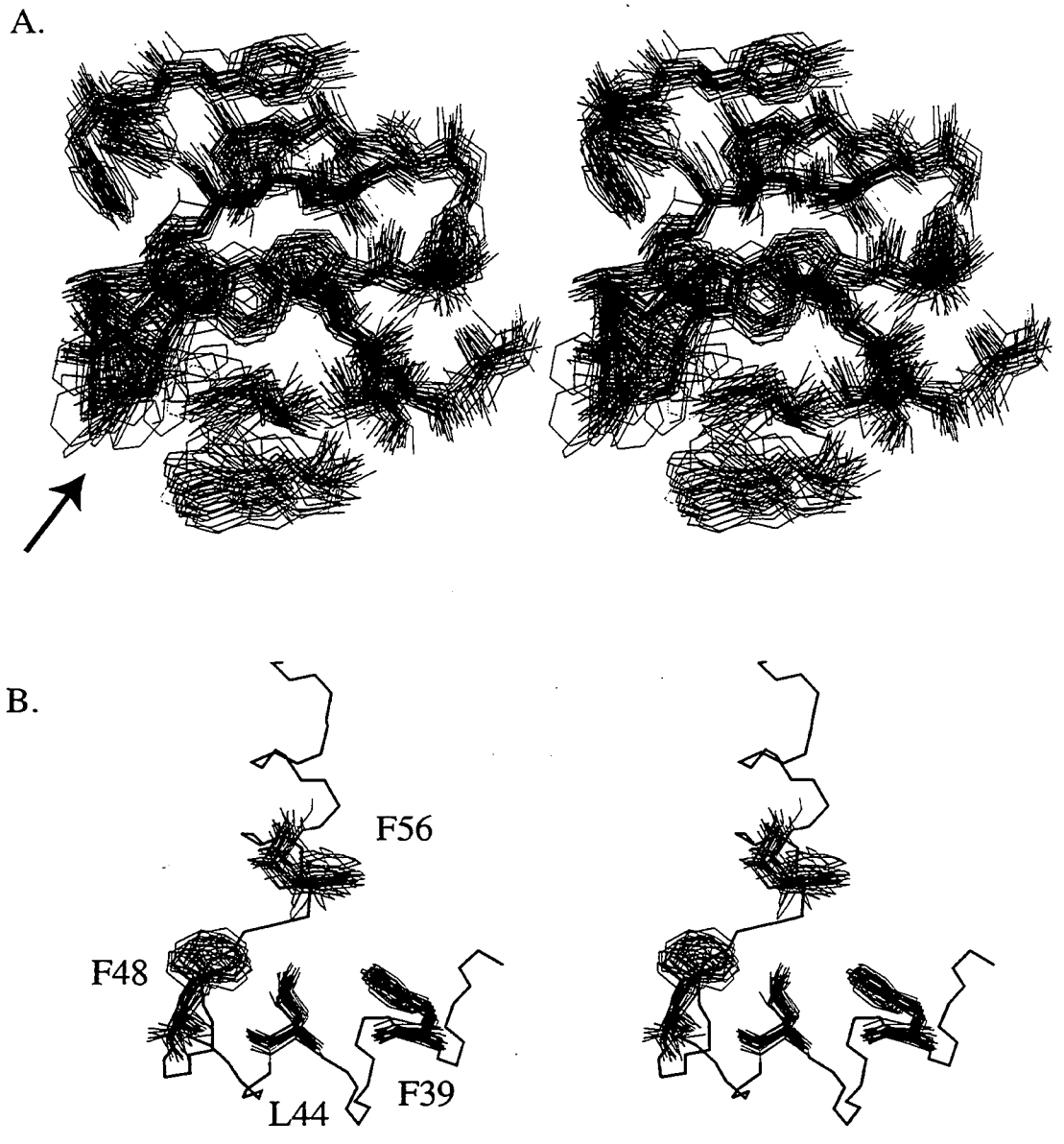


Figure 3.5 Core residue superpositions for HSF. **A.** Superposition of all the heavy atoms for the 25 buried residues (<20% exposed in the crystal structure). Note that the sidechains are relatively well ordered with an rmsd for this superposition of 0.84 ± 0.15 Å. The location of F53 is indicated by an arrow. **B.** Superposition of a cluster of highly conserved residues in helices H2 and H3 (rmsd of 0.58 ± 0.16 Å). The residues are identified with their one-letter amino acid code and location in the sequence. The backbone coordinates for helices H2 and H3 for one of the nmr structures is shown for reference. Note the underwinding and kink in helix H2 which place F39 near L44.

± 0.14 Å <rmsd> for the backbone coordinates of all secondary structured regions. The structure can be described as "high resolution" since the three helix bundle and beta sheet show backbone <rmsd>s of 0.53 ± 0.13 Å and 0.57 ± 0.15 Å respectively. Structural statistics for the family of structures are summarized in Table 3.1.

The larger <rmsd>s for larger segments of the structure can be explained by examining the distribution of NOEs (Fig. 3.6A and Appendix F) which show that there are far more NOEs constraining the secondary structured regions than non-secondary structured regions included in the rmsd calculations. Two regions which contribute significantly to the increased <rmsd>s for all structured regions are the turn region between helices H2 and H3 and the end of β-strand B3 (Fig. 3.6B). Inspection of the backbone phi angle distribution reveals that the local backbone conformation is not well defined in these regions either ($S(\phi) = 0.63 \pm 0.33$ for K49-F53 and $S(\phi) = 0.73 \pm 0.19$ for W65-V68) (Fig. 3.6C).

The degree to which sidechain conformations are well defined in the regions of secondary structure is revealed by the angular order parameters for $\chi 1$ and $\chi 2$ (Fig. 3.6DE and Appendix F) which have average values of 0.83 ± 0.24 and 0.72 ± 0.27 respectively. Particularly well defined are the sidechain conformations of helices H2 and H3 ($S(\chi 1)=0.88 \pm 0.17$). This high degree of ordering in these two secondary structure elements correlates with the high degree of sequence conservation in this region and its crucial role in the function of the DNA-binding domain of HSF.

Dynamics from Heteronuclear NOEs

Heteronuclear NOEs could be measured for 79 of 88 assigned backbone amide signals (Appendix G). The heteronuclear NOE is sensitive to rapid reorientation of the N-H bond vector from local dynamics and will exhibit a reduced or negative NOE for dynamically disordered regions of the structure (Barbato, et al., 1992). Rigid regions of

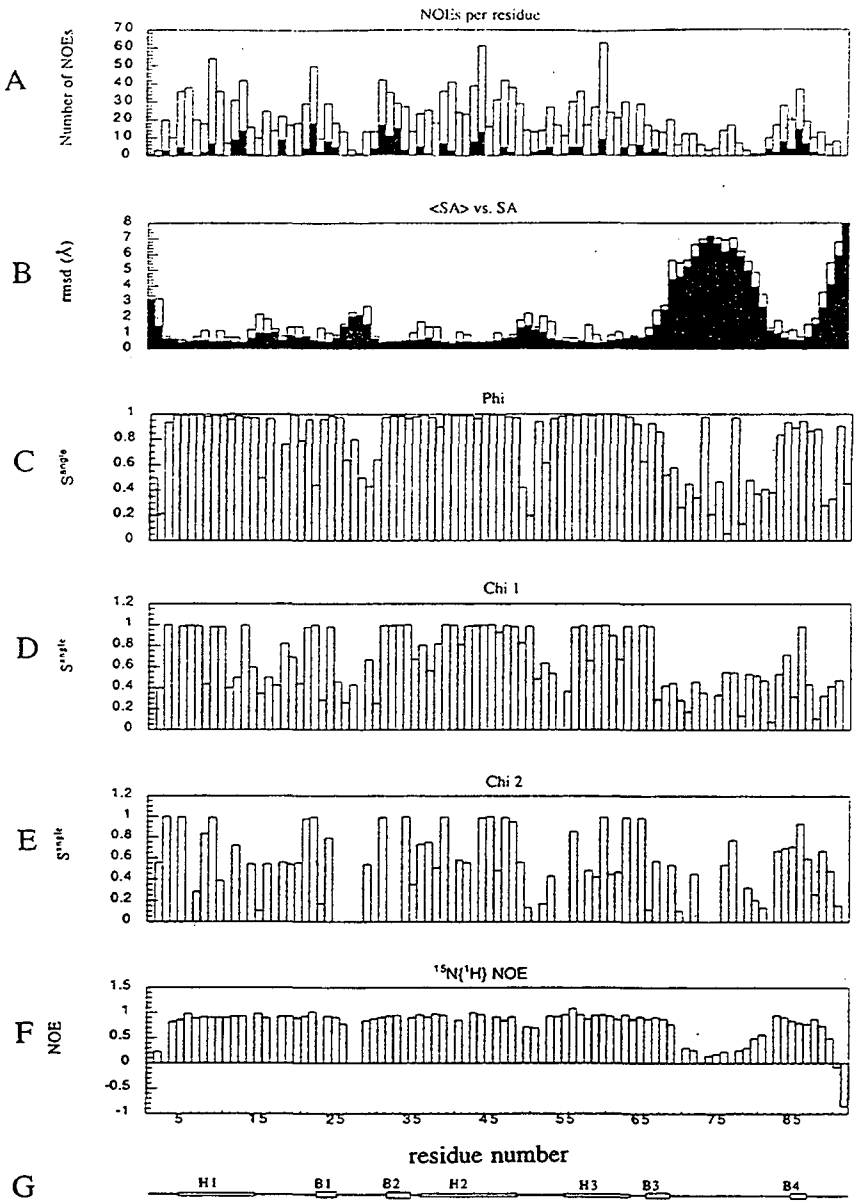


Figure 3.6 Number of NOEs, rmsds, angular order parameters and $^{15}N\{^1H\}$ NOEs versus residue number. **A**. Total NOEs to each residue shown in open bars (each NOE is counted twice, once for each partner. Solid bars indicate total number of long range NOEs to each residue. **B**. <rmsd> between the family of structures <SA> and the average structure SA based on an alignment of backbone coordinates in structured regions (residues 4-24, 31-68, 84-86). <rmsd> for backbone coordinates (N, C_α, C', O) (solidbars) and all heavy atoms (open bars) are shown. **C**. Angular order parameters for backbone phi angles. **D**. Angular order parameters for chi 1 angles. **E**. Angular order parameters for chi 2 angles. **F**. $^{15}N\{^1H\}$ NOEs for each residue. **G**. Representation of secondary structure elements in HSF; helices by narrow bars, and β-sheet by wide bars.

structure will exhibit large positive NOE intensities which correspond to the slower global tumbling of the molecule in solution. The structurally ordered regions of HSF exhibit large positive NOEs (Fig. 3.6F) (0.92 ± 0.07) with the notable exception of the turn region between helices H2 and H3 which exhibits significantly reduced NOEs (0.72 for H50 and 0.70 for S51) and appears to undergo restricted motion on fast timescales. It is unclear why the calculated heteronuclear NOEs are consistently larger than expected for a rigidly defined domain of this size (~0.85). Qualitative trends in fast timescale dynamics can none-the-less be identified.

Examining the loop regions, we find that the short loop connecting beta strands B1 and B2 (Ω1) shows significantly more fast timescale dynamics than the secondary structural elements (0.78 for T26 and 0.84 for E29) and the extended loop (L1) connecting beta strands B3 and B4 shows small NOEs indicative of rapid motion (0.35 ± 0.20 for Q69-D81). This is consistent with the absense of long range NOEs in this region. The helical turn region between helix H1 and β-strand B1 which has poorly defined secondary structure none-the-less shows large positive NOEs indicative of little motion on fast time scales (0.93 ± 0.04). The fourth beta strand B4 shows a rigid conformation on fast timescales extending N-terminally to R83 and C-terminally to N88. This correlates with the S(ϕ) values in this region (0.89 ± 0.04) which remain high from R83 to N88 (Fig. 3.6 and Appendix F) indicating that the structure is well-defined somewhat beyond the region described as β-strand B4 in Chapter 2. The N-terminal Alanine N-H shows a reduced NOE as do the the C-terminal residues with Ala-92 exhibiting a large negative NOE indicative of a completely flexible terminus.

Discussion

Comparison to crystal structure

Comparison of the high resolution NMR structure of *K. lactis* HSF to that of the 1.8 Å crystal structure reveals a remarkable degree of agreement between the two. The crystal structure shows no violations greater than 2 Å to the NMR distance restraints (with one exception detailed below) and has a backbone <rmsd> against the family of high resolution solution structures of only 1.31 ± 0.11 Å. This is down from the 1.77 ± 0.20 Å <rmsd> observed for comparison of the same coordinates for the medium resolution family described in Chapter 2. This improved agreement between the crystal structure and the NMR structure probably reflects the increase in precision of the NMR structures with a decrease in backbone <rmsd> for structured regions from the average coordinates of 1.03 ± 0.19 Å to 0.81 ± 0.15 Å. In addition the number of significant distance violations has decreased possibly reflecting an improvement in the accuracy of the restraints obtained through the higher dimensional experiments (which reduce the likelyhood of overlap and the associated errors in estimation of crosspeak intensity).

The crystal structure significantly violates the constraint corresponding to an NOE between V6 Hγ1 methyl and W10 Hε1. This NOE is observed to have a medium intensity in the 4D 13C/15N NOESY plane and is in a region of the spectrum which makes its assignment completely unambiguous. The 1.8 Å crystal structure shows a distance of 8.5 Å between these protons which would not give rise to an NOE. Hence there clearly is a local change in the relative positioning of the sidechains of V6 and W10. Examining the solution structure, W10 participates in the formation of a hydrophobic patch on the surface of the protein formed by two other trpytophans W24 and W65. In the crystal structure there are protein-protein contacts in this region which may be the cause of the reorientation of the

sidechains. Aligning the coordinates of the *K. lactis* HSF DBD from the NMR structure with those of D. melanogaster HSF reveals that the sidechain of W10 (W54 in D. *melanogaster*) has significantly different conformations in these structures. Where as in the *K. lactis* structure the sidechain is folded flat against the protein surface, in *D. melanogaster* it is perpendicular to the surface. This difference is probably due to the additional C-terminal fragment included in the *D. melanogaster* construct which forms a poorly ordered helical turn region (P139-Q144) that packs up against the hydrophobic surface formed by W54 and the surrounding hydrophobic residues (Vuister, et al., 1994a). This contact may be significant for the interaction between the DBD and the adjacent linker region and C-terminal activation domain.

Comparison to homologous protein structures

Aligning the backbone coordinates for the solution structures of *D. melanogaster* HSF (dHSF) and *K. Lactis* HSF (kHSF) revealed a remarkable correspondence between these proteins which are from such distinctly unrelated organisms (backbone rmsd for structured regions between homologous residues of the SA_r structures of 1.77 Å). The unusual variations of the HTH motif in kHSF are reproduced in dHSF with an extended and apparently unstructured loop in the residues corresponding to L1 (*D. melanogaster* HSF T113-R124) and the distorted helix containing the proline insertion, which packs against the recognition helix and stabilizes the folded native protein. The agreement between the two structures in this region is quite close (backbone rmsd for HTH 1.3Å) as might be expected for such a highly conserved protein, since this part of the molecule is crucial for DNA recognition.

The extensive sequence homology in the HTH region, particularly for the hydrophobic residues, are likely to have their origins in the preservation of the packing arrangement which stabilizes the protein fold and reproduces the recognition helix geometry

necessary for specific interactions with GAA boxes of HSE's. This is reflected in the structures of HSF from *D. melanogaster* and *K.lactis* with an rmsd for the hydrophobic conserved residues in the core of 2.78 Å for all heavy atoms (coordinates of minimized average structures of dHSF and kHSF). Closer examination of the strictly conserved residues in the HTH region reveals a structural basis for the strong homology in this region, and helps to explain how this unusual kinked-helix-turn-helix motif is stabilized.

Residues F39, L44 and F56 are completely conserved in all HSF's and the position of F48 only tolerates the conservative substitution of a tyrosine (in dHSF the equivalent residue is found to be a tyrosine). These residues form a cluster of highly ordered hydrophobic side chains, wedged into the turn region between H2 and H3 (Fig. 3.5B). The three aromatic rings surround the $H\delta$ methyl groups of L44 which makes contacts with the backbone atoms of the facing turn region between the two helices. This arrangement of aromatic groups around the Leu $H\delta$ methyls is reproduced in the dHSF structure (Fig. 3.7) except that the two δ-methyl groups of the Leucine have switched positions (rmsd for the analgous heavy atoms of these residues in the SA_r structures of dHSF and kHSF are 2.10 Å).

An interesting feature of this arrangement is that the side chain of F39 reaches forward to make contact with L44 which would not be possible if H2 had a normal helical geometry, since this would place it on the opposite face of the helix. Thus it is likely that this hydrophobic cluster is integral to stabilizing the helical kink induced by the proline insertion and is required to define the relative orientations of the helices necessary for the DNA interaction as well as overall protein stability. Note also that there are several conserved polar and exposed residues in the turn region after H2 whose positions would be changed if H2 had a normal geometry.

Additionally, in the crystal structure, residue N52 appears to be engaged in an N cap hydrogen bond (Harper & Rose, 1993; Lyu & Wemmer, 1993; Seale, et al., 1994)

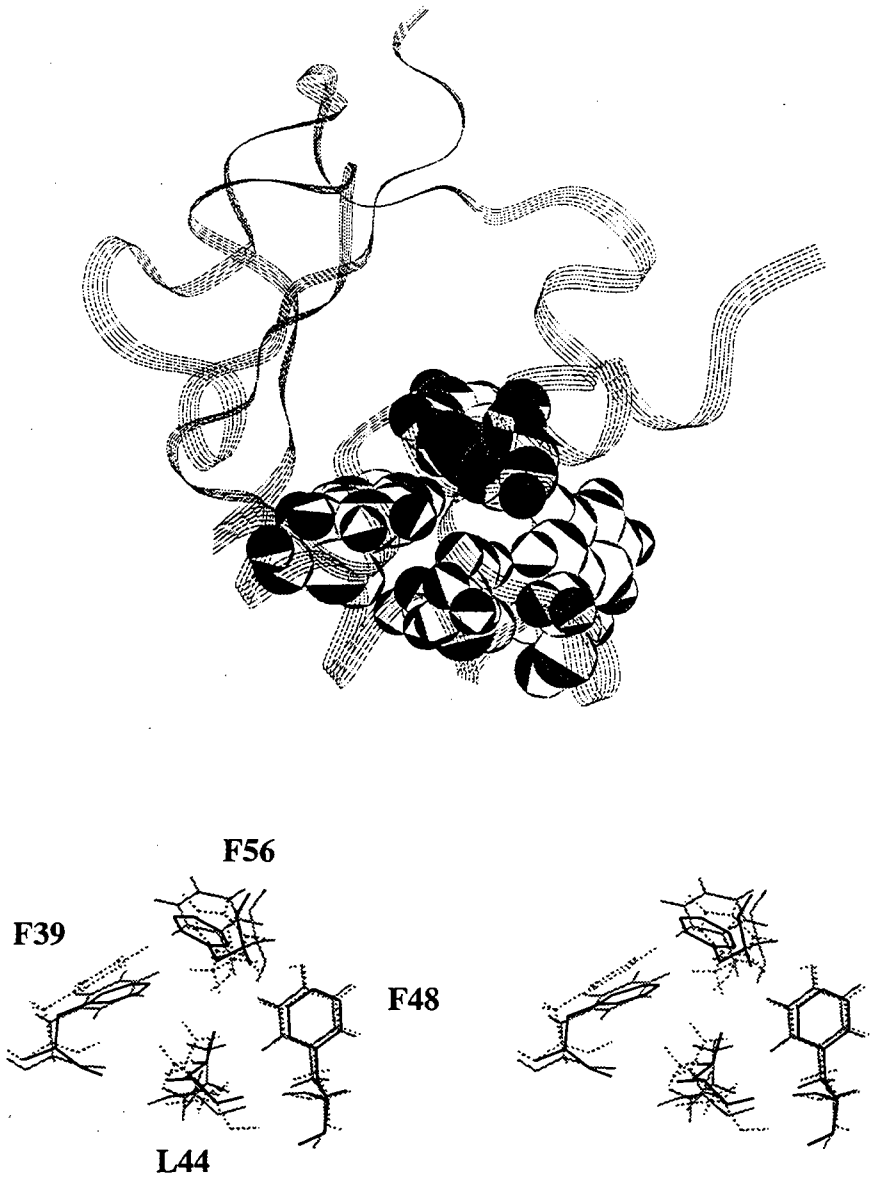


Figure 3.7 **A**. CPK representation of F39,L44,F48,F56 and ribbon diagram of a member of the family of *K. lactis* HSF structures. **B**. Stick figure of conserved residues in cluster mentioned in text in the same orientation as **A**. In black without Hydrogens is the 1.8 Å X-ray structure. Dashed lines indicate the *D. melanogaster* HSF solution structure (residues corresponding to the labels are F83, L88,Y92,F100). The *K. lactis* HSF solution structure $(SA)_r$ is rendered in solid grey lines. rmsd for the coordinates shown between the $(SA)_r$ structure and the crystal structure is 1.1 Å. Aligning the *D. melanogaster* HSF coordinates shown against $(SA)_r$ resulted in an rmsd of 2.1 Å.

with its side chain carbonyl oxygen accepting a hydrogen bond from the backbone amide of A54. This N-cap was not observed in our solution structures. Slowed amide exchange is observed for the A54 (Fig. 2.10) which should not be involved in backbone hydrogen bonds, since it is the first residue in the helix. This stabilizing cap interaction is also likely to be perturbed in an unkinked α-helix H2, thus the local packing interactions of the cluster shown in fig. 3.7, participate in the global stabilization of this unusual HTH motif.

Role of the Extended Loop and Dynamics

One of the unusual features of the DBD of HSF is the extended loop connecting β-strands B3 and B4. This loop is quite variable in sequence among different organisms and appears to have been dispensed with completely in tomato plants (Fig. 2.1). However, for organisms that have retained the loop, there is a completely conserved glycine (position 267 in *K. lactis*). In addition two yeast sequences (*K. lactis* and *Saccharomyces cerevisiae*) exhibit a region of complete conservation extending from β-strand B4 to the conserved glycine. Both yeasts constitutively express HSF and require its low level expression for normal growth (Jakobsen & Pelham, 1988). Moreover, *S. cerevisiae* expressing HSF with an Ala mutation at the conserved Gly position are not viable (Hubl, et al., 1994). A second point mutant of HSF at a conserved Lys at the end of B3 also results in nonviable yeast. Hence there appear to be HSF functions associated with the N-terminal portion of the extended loop L1. The loop in the analogous position in HNF3-γ forms one wing of the winged helix motif and makes contacts with the DNA which help to orient the recognition helix for proper readout of the DNA sequence. It is not clear what role the extended loop plays in yeast although it is tempting to speculate that it may be involved in contacts between adjacent protein subunits in the trimeric complex or possibly contacts to the DNA.

In the previously described structural studies of HSF very little information was obtained about the conformations of the loop region. The crystal structure of *K. lactis* DBD

shows missing electron density for residues 76-79 and extensive contacts to other proteins in the crystal for the remaining residues in the loop. Both *Drosophila* and *K. lactis* solution structures of the DBD show poorly defined conformations for L1 due to the absence of long range NOE distance constraints (Damberger, et al., 1994; Vuister, et al., 1994a). Small positive to negative heteronuclear NOE values for N-H signals in this loop (Fig. 3.6F) offer direct evidence for the rapid motional averaging occuring in this region in solution and explain the absence of defined conformations as determined by NMR structural analysis. In addition, $\chi 1$ angles of residues in the loop region appear to undergo averaging as evidenced by equally intense H_N-$H\beta$ and $H\alpha$-$H\beta$ signals for both $H\beta$ protons in a residue in the stereoassignment experiments (Appendix B).

If the extended loop, L1, plays a role in protein-protein contacts in the DNA complex it must undergo an order/disorder transition upon forming interactions with other monomers. The dynamic motion observed in L1 does not appear to be a consequence of truncating HSF to just the DBD and disrupting normal interdomain contacts since longer constructs containing the adjacent N-terminal activation domain exhibit resonances assignable to L1 which also display heteronuclear NOEs indicative of rapid timescale motions (Cho et al., unpublished results). It will be interesting to see what role this loop plays in the homotrimeric protein DNA complex in yeast and whether it becomes ordered upon DNA-binding of the DBD + trimerization domains. Small flexible loops are observable in spectra of large protein aggregates (Landry, et al., 1993) so that it will be possible to say if the loop becomes involved in interactions within the homotrimer or between adjacent homotrimers in larger cooperative complexes (Xiao, et al., 1991). Such large high affinity complexes with multiple HSEs are at the heart of the functioning of HSF in a biological context since they are what allow HSF to cause drastic changes in protein expression required under heat shock conditions in the cell.

Bibliography

Amin J, Ananthan J , Voellmy R. 1988. Key features of heat shock regulatory elements. *J Mol Biol 8*:3761-3769.

Archer SJ, Ikura M, Sporn MB, Torchia DA , Bax A. 1991. An Alternative 3D Technique for Correlating Backbone ^{15}N with Side Chain Hβ Resonances in Larger Proteins. *J Magn Reson 95*:636-641.

Archer SJ, Vinson VK, Pollard TD , Torchia DA. 1993. Secondary structure and topology of *Acanthamoeba* profilin I as determined by heteronuclear nuclear magnetic resonance spectroscopy. *Biochemistry 32*:6680-6687.

Aurora R, Srinivasan R , Rose GD. 1994. Rules for α-helix termination by glycine. *Science 264*:1126-1130.

Barbato G, Ikura M, Kay LE, Pastor RW , Bax A. 1992. Backbone Dynamics of Calmodulin Studied by ^{15}N Relaxation Using Inverse Detected Two-Dimensional NMR Spectroscopy: The Central Helix is Flexible. *Biochemistry 31*:5269-5278.

Barlow DJ , Thornton JM. 1988. Helix geometry in proteins. *J Mol Biol 201*:601-619.

Bax A, Clore GM, Driscoll PC, Gronenborn AM, Ikura M , Kay LE. 1990a. Practical aspects of proton carbon carbon proton 3-dimensional correlation spectroscopy of ^{13}C-labeled proteins. *J Magn Reson 87*:620-627.

Bax A, Clore GM , Gronenborn AM. 1990b. ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new 3-dimensional approach for assigning ^{1}H And ^{13}C spectra of ^{13}C-enriched proteins. *J Magn Reson 88*:425-431.

Bax A , Davis G. 1985. MLEV-17-based two dimensional homonuclear magnetization transfer spectroscopy. *J Magn Reson 65*:355-360.

Bax A, Ikura M, Kay LE, Torchia DA , Tschudin R. 1990c. Comparison Of Different Modes Of 2-Dimensional Reverse-Correlation NMR For The Study Of Proteins. *J Magn Reson 86*:304-318.

Bax A, Max D , Zax D. 1992. Measurement of Long-Range ^{13}C-^{13}C J Couplings in 20-kDa Protein-Peptide Complex. *J Am Chem Soc 114*:6923-6925.

Bax A , Subramanian S. 1986. Sensitivity enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J Magn Reson 67*:565-569.

Bochkareva ES, Lissin NM, Flynn GC, Rothman JE , Girshovich AS. 1992. Positive cooperativity in the functioning of molecular chaperone GroEL. *J Biol Chem 267*:6796-6800.

Bodenhausen G , Ruben DJ. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett 69*:185-189.

Braunschweiler L , Ernst RR. 1983. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J Magn Reson 53*:521-528.

Brennan RG. 1993. The winged helix motif: Another helix-turn-helix takeoff. *Cell 74*:773-776.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S , Karplus M. 1983. CHARMm: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem 4*:187-217.

Brünger AT. 1992. *X-PLOR Manual Version 3.0: A System for Crystallography and NMR*. New Haven, Connecticut: Yale University Press.

Bystrov VF. 1976. *Prog NMR Spectrosc 10*:41-81.

Cavanagh J, Chazin W , Rance M. 1990. The Time Dependence of Coherence Transfer in Homonuclear Isotropic Mixing Experiments. *J Magn Reson 87*:110-131.

Clark KL, Halay ED, Lai E , Burley SK. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature 364*:412-420.

Clore GM, Bax A, Driscoll PC, Wingfield PT , Gronenborn AM. 1990. Assignment of the side-chain ^{1}H and ^{13}C resonances of interleukin-1β using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry 29*:8172-8184.

Clore GM, Bax A , Gronenborn AM. 1991. Stereospecific assignment of β-methylene protons in larger proteins using 3D ^{15}N-separated Hartman-Hahn and ^{13}C-separated rotating frame Overhauser spectroscopy. *J Biomol NMR 1*:13-22.

Clore GM, Gronenborn AM, Brünger AT , Karplus M. 1985. *J Mol Biol 185*:435-455.

Clore GM, Kay LE, Bax A , Gronenborn AM. 1991. Four-dimensional $^{13}C/^{13}C$-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin 1β. *Biochemistry 30*:12-18.

Clore GM, Nilges M, Sukumaran DK, Brünger AT, Karplus M , Gronenborn AM. 1986. The three-dimensional structure of α1-purothionin in solution: Combined use of nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *EMBO J 5*:2729-2735.

Cowing DW, Bardwell JC, Craig EA, Woolford C, Hendrix RW , Gross CA. 1985. Consensus sequence for *Escherichia coli* heat shock gene promoters. *Proc Natl Acad Sci USA 82*:2679-2683.

Craig EA, Gambill BD , Nelson RJ. 1993. Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiological Rev. 57*:402-414.

Crippen GM , Havel TF. 1978. Stable Calculation of Coordinates from Distance Information. *Acta Crystallogr A34*:282-284.

Damberger FF, Pelton JG, Harrison CJ, Nelson HCM , Wemmer DE. 1994. Solution structure of the DNA-binding domain of the heat shock transcription factor determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Protein Science 3*:1806-1821.

Dombroski AJ, Walter WA, Record MTJ, Siegele DA , Gross CA. 1992. Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell 70*:501-512.

Driscoll PC, Clore GM, Marion D, Wingfield PT , Gronenborn AM. 1990. Complete resonance assignment for the polypeptide backbone of interleukin 1β using three-dimensional heteronuclear NMR spectroscopy. *Biochemistry 29*:3542-3556.

Ernst RR, Bodenhausen G , Wokaun A. 1990. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Oxford University Press.

Feng JA, Johnson RC , Dickerson RE. 1994. Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science 263*:348-355.

Fesik SW, Eaton HL, Olejniczak ET , Zuiderweg ERP. 1990. 2D and 3D NMR Spectroscopy Employing ^{13}C-^{13}C Magnetization Transfer by Isotropic Mixing. Spin System Identification in Large Proteins. *J Am Chem Soc 112*:886-888.

Flick K, Gonzalez LJ, Harrison CJ , Nelson HCM. 1994. Yeast heat shock factor contains a flexible linker between the DNA-binding and Trimerization domains; Implications for DNA-binding by trimeric proteins. *J Biol Chem 269*:12475-12481.

Freeman RA. 1988a. In: ed. *A Handbook of Nuclear Magnetic Resonance*. New York: Wiley. pp 46.

Freeman RA. 1988b. In: ed. *A Handbook of Nuclear Magnetic Resonance*. New York: Wiley. pp 14-16.

Gallo GJ, Prentice H , Kingston RE. 1993. Heat shock factor is required for growth at normal temperatures in the fission yeast *Schizosaccharomyces pombe*. *Mol Cell Biol 13*:749-761.

Gardella T, Moyle H , Susskind MM. 1989. A mutant *Escherichia coli* sigma 70 subunit of RNA polymerase with altered promoter specificity. *J Mol Biol 206*:579-590.

Garret DS, Powers R, Gronenborn AM , Clore GMA. 1991. A common sense approach to peak picking in two- three- and four- dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* 214-220.

Gray TE , Fersht AR. 1991. Cooperativity in ATP hydrolysis by GroEL is increased by GroES. *FEBS Lett 292*:254-258.

Griesinger C, Otting G, Wüthrich K , Ernst RR. 1988. Clean TOCSY for ^{1}H spins system identification in macromolecules. *J Magn Reson 110*:7870-7820.

Grzesiek S , Bax A. 1992a. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J Magn Reson 99*:201-207.

Grzesiek S , Bax A. 1992b. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc 114*:6291-6293.

Grzesiek S , Bax A. 1993. Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J Biomol NMR 3*:185-204.

Grzesiek S, Vuister GW , Bax A. 1993. A simple and sensitive experiment for measurement of J_{CC} couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. *J Biomol NMR 3*:487-493.

Harper ET , Rose GD. 1993. Helix Stop Signals in Proteins and Peptides: The Capping Box. *Biochemistry 32*:7605-7609.

Harrison CJ. 1994. The structure of the DNA-binding domain from *Kluyveromyces lactis* heat shock transcription factor [thesis]. Berkeley, California: University of California.

Harrison CJ, Bohm AA , Nelson HCM. 1994. Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science 263*:224-227.

Harrison SC , Aggarwal AK. 1990. DNA recognition by proteins with the helix-turn-helix motif. *Annu Rev Biochem 59*:933-969.

Havel TF , Wüthrich K. 1984. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance

measurements of intramolecular ^{1}H-^{1}H proximities in solution. *Bull Math Biol 46*:673-698.

Havel TF , Wüthrich K. 1985. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J Mol Biol 182*:281-294.

Horvath OW , Lilley DM. 1978. ^{13}C NMR of peptides. *Prog NMR Spectrosc 12*:1-40.

Hubl ST, Owens JC , Nelson HCM. 1994. Mutational analysis of the DNA-binding domain of yeast heat shock transcription factor. *Struct Biol 1*:615-620.

Hyberts SG, Goldberg MS, Havel TF , Wagner G. 1992. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci 1*:736-751.

Ikura M, Kay LE , Bax A. 1990. A Novel Approach for Sequential Assignment of ^{1}H, ^{13}C, and ^{15}N Spectra of Larger Proteins: Heteronuclear Triple Resonance Three-Dimensional NMR Spectroscopy. Application to Calmodulin. *Biochemistry 29*:4659-4667.

Ikura M, Kay LE , Bax A. 1991b. Improved three dimensional ^{1}H-^{13}C-^{1}H spectroscopy of a ^{13}C-labeled protein using constant-time evolution. *J Biomol NMR 1*:299-304.

Ikura M, Kay LE, Tschudin R , Bax A. 1990. 3-dimensional NOESY-HMQC spectroscopy of a ^{13}C-labeled protein. *J Magn Reson 86*:204-209.

Ikura M, Spera S, Barbato G, Kay LE, Krinks M , Bax A. 1991a. Secondary structure and side-chain ^{1}H and ^{13}C resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy. *Biochemistry 30*:9216-9228.

Jakobsen BK , Pelham HR. 1988. Constitutive binding of yeast heat shock factor to DNA in vivo. *Mol Cell Biol 8*:5040-5042.

Jakobsen BK , Pelham HR. 1991. A conserved heptapeptide restrains the activity of the yeast heat shock transcription factor. *EMBO J 10*:369-375.

Karplus MJ. 1959. *J Phys Chem 30*:11-15.

Kay LE , Bax A. 1990. New methods for the measurement of NH-C-α-H coupling constants in ^{15}N-labeled proteins. *J Magn Reson 86*:110-126.

Kay LE, Clore GM, Bax A , Gronenborn AM. 1990. Four-Dimensional Heteronuclear Triple-Resonance NMR Spectroscopy of Interleukin-1β in Solution. *Science 249*:411-414.

Kay LE, Marion D , Bax A. 1989. Practical aspects of 3D heteronuclear NMR of proteins. *J Magn Reson 84*:72-84.

Kraulis J, Clore GM, Nilges M, Jones TA, Pettersson G, Knowles J , Gronenborn AM. 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei.* A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry 28*:7241-7257.

Landry SJ, Zeilstra-Ryalls J, Fayet O, Georgopoulos C , Gierasch L. 1993. Characterization of a functionally important mobile domain of GroES. *Nature 364*:255-258.

Li Y-C , Montelione GT. 1994. Overcoming Solvent-Transfer Artifacts in Protein NMR at Neutral pH. Application of Pulsed Field Gradients in Measurements of ^{1}H-^{15}N Overhauser Effects. *J Magn Reson B 105*:45-51.

Lis J , Wu C. 1993. Protein traffic on the heat shock promoter: parking, stalling, and trucking along. *Cell 74*:1-4.

Live DH, Davis DG, Agosta WC , Cowburn D. 1984. Long-range hydrogen bond mediated effects in peptides: ^{15}N NMR study of gramacidin S in water and organic solvents. *J Am Chem Soc 106*:1934-1941.

Lodi PJ, Garrett DS, Kuszewski J, Tsang ML, Weatherbee JA, Leonard WJ, Gronenborn AM , Clore GM. 1994. High-resolution solution structure of the β chemokine hMIP-1 β by multidimensional NMR. *Science 263*:1762-1767.

Lyu PC , Wemmer DE. 1993. Capping Interactions in Isolated α Helices: Position-Dependent Substitution Effects and Structure of a Serine-Capped Peptide Helix. *Biochemistry 32*:421-425.

Macura S, Huang Y, Suter D , Ernst RR. 1981. Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J Magn Reson 43*:259-281.

Marion D, Ikura M , Bax A. 1989a. Improved solvent suppression in one-dimensional and 2-dimensional NMR spectra by convolution of time-domain data. *J Magn Reson 84*:425-430.

Marion D, Ikura M, Tschudin R , Bax A. 1989b. Rapid recording of 2D NMR spectra without phase cycling - application to the study of hydrogen exchange in proteins. *J Magn Reson 85*:393-399.

Marion D, Kay LE, Sparks SW, Torchia DA , Bax A. 1989c. Three dimensional heteronuclear NMR of ^{15}N labeled proteins.

Marion D , Wüthrich K. 1983. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem Biophys Res Commun 113*:967-974.

McIntosh L , Dahlquist FW. 1990. Biosynthetic Incorporation of ^{15}N and ^{13}C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins. *Quart Rev Biophys 23*:1-38.

Messerle BA, Wider G, Otting G, Weber C , Wüthrich K. 1989. Solvent suppression using a spin lock in 2D NMR and 3D NMR spectroscopy with H_2O solutions. *J Magn Reson 85*:608-613.

Metzler WJ, Constantine KL, Friedrichs MS, Bell AJ, Ernst EG, Lavoie TB , Mueller L. 1993. Characterization of the three-dimensional solution structure of human profilin: ^{1}H, ^{13}C, and ^{15}N NMR assignments and global folding pattern. *Biochemistry 32*:13818-13829.

Morimoto RI. 1993. Cells in stress: transcriptional activation of heat shock genes. *Science 259*:1409-1410.

Morimoto RI, Sarge KD , Abravaya K. 1992. Transcriptional regulation of heat shock genes. A paradigm for inducible genomic responses. *J Biol Chem 267*:21987-21990.

Morris AL, MacArthur MW, Hutchinson EG , Thornton JM. 1992. Stereochemical quality of protein structure coordinates. *Proteins 12*:345-364.

Morris GA , Freeman R. 1979. *J Am Chem Soc 101*:760-.

Mueller L. 1979. Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence. *J Am Chem Soc 101*:4481-4484.

Neri D, Billeter M, Wider G , Wüthrich K. 1992. NMR Determination of Residual Structure in a Urea-Denatured Protein, the 434-Repressor. *Nature 257*:1559-1563.

Neuhaus D , Williamson M. 1989. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. New York: VCH Publishers.

Nieto-Sotelo J, Wiederrecht G, Okuda A , Parker CS. 1990. The yeast heat shock transcription factor contains a transcriptional activation domain whose activity is repressed under nonshock conditions. *Cell 62*:807-817.

Nilges M, Clore GM , Gronenborn AM. 1988. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett 229*:317-324.

Pardi A, Billeter M , Wüthrich K. 1984. Calibration of the angular dependence of the amide proton-C alpha proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein. Use of $^{3}J_{HN\alpha}$ for identification of helical secondary structure. *J Mol Biol 180*:741-751.

Pelton JG, Torchia DA, Meadow ND, Wong CY , Roseman S. 1991. ^{1}H, ^{15}N, and ^{13}C NMR signal assignments of IIIGlc, a signal-transducing protein of *Escherichia coli*, using three-dimensional triple-resonance techniques. *Biochemistry 30*:10043-10057.

Peteranderl R , Nelson HCM. 1992. Trimerization of the heat shock transcription factor by a triple-stranded alpha-helical coiled-coil. *Biochemistry 31*:12272-12276.

Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM , Clore GM. 1993. The high-resolution, three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry 32*:6744-6762.

Rance M, Sorensen OW, Bodenhausen G, Wagner G, Ernst RR , Wüthrich K. 1983. Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem Biophys Res Commun 117*:479-485.

Seale JW, Srinivasan R , Rose GD. 1994. Sequence determinants of the capping box, a stabilizing motif at the N-termini of α-helices. *Protein Sci 3*:1741-1745.

Shaka AJ, Barker P , Freeman R. 1985. Computer-optimized decoupling scheme for wideband applications and low level operation. *J Magn Reson 64*:547-552.

Shaka AJ, Keeler J , Freeman R. 1983. Evaluation of a new broadband decoupling sequence: Waltz16. *J Magn Reson 53*:313-340.

Shaka AJ, Lee CJ , Pines A. 1988. Iterative schemes for bilinear operators: Application to spin decoupling. *J Magn Reson 77*:274-293.

Shellman C. 1980. In: R. Jaenicke, ed. *Protein Folding : proceedings of the 28th conference of the german biochemical society*. New York: Elsevier/North-Holland. pp 53-61.

Siegele DA, Hu JC, Walter WA , Gross CA. 1989. Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. *J Mol Biol 206*:591-603.

Sippl MJ , Sheraga HA. 1986. *Proc Natl Acad Sci USA 83*:2283-2287.

Sorger PK. 1990. Yeast heat shock factor contains separable transient and sustained response transcriptional activators. *Cell 62*:793-805.

Sorger PK , Nelson HCM. 1989. Trimerization of a yeast transcriptional activator via a coiled-coil motif. *Cell 59*:807-813.

Spera S , Bax A. 1991. Empirical correlation between protein backbone conformation and C_α and C_β ^{13}C nuclear magnetic resonance chemical shifts. *J Am Chem Soc 113*:5490-5492.

Steitz TA. 1990. Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Quart Rev Biophys 23*:205-280.

Studier FW, Rosenberg AH, Dunn JJ , Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60-89.

Vuister GW , Bax A. 1992. Resolution enhancement and spectral editing of uniformly ^{13}C-enriched proteins by homonuclear broadband ^{13}C decoupling. *J Magn Reson 98*:428-435.

Vuister GW , Bax A. 1993. Measurement of Two- and Three-Bond Proton to Methyl Carbon J Couplings in Proteins Uniformly Enriched with ^{13}C. *J Magn Reson 102B*:228-231.

Vuister GW, Kim S-J, Orosz A, Marquardt J, Wu C , Bax A. 1994a. Solution structure of the DNA-binding domain of Drosophila heat shock transcription factor. *Struct Biol 1*:605-614.

Vuister GW, Kim SJ, Wu C , Bax A. 1994b. NMR evidence for similarities between the DNA-binding regions of *Drosophila melanogaster* heat shock factor and the helix-turn-helix and HNF-3/forkhead families of transcription factors. *Biochemistry 33*:10-16.

Vuister GW, Wang AC , Bax A. 1993. Measurement of 3-bond nitrogen carbon-J couplings in proteins uniformly enriched in ^{15}N and ^{13}C. *J Am Chem Soc 115*:5334-5335.

Wagner G, Braun W, Havel TF, Schaumann T, Go N , Wüthrich K. 1987. Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J Mol Biol 196*:611-639.

Wiederrecht G, Seto D , Parker CS. 1988. Isolation of the gene encoding the *S. cerevisiae* heat shock transcription factor. *Cell 54*:841-853.

Williamson MP, Havel TF , Wüthrich K. 1985. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J Mol Biol 182*:295-315.

Wishart DS, Sykes BD , Richards FM. 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol 222*:311-333.

Wüthrich K. 1986. *NMR of Proteins and Nucleic Acids*. John Wiley & Sons.

Wüthrich K, Billeter M , Braun W. 1983. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J Mol Biol 169*:949-961.

Xiao H , Lis JT. 1988. Germline transformation used to define key features of heat-shock response elements. *Science 239*:1139-1142.

Xiao H, Perisic O , Lis JT. 1991. Cooperative binding of *Drosophila* heat shock factor to arrays of a conserved 5 bp unit. *Cell 64*:585-593.

Zhang H, Zhao D, Revington M, Lee W, Jia X, Arrowsmith C , Jardetsky O. 1994. The Solution Structures of the *trp* Repressor-Operator DNA Complex. *J Mol Biol 238*:592-614.

Zuiderweg ER, Boelens R , Kaptein R. 1985. Stereospecific assignments of ^{1}H NMR methyl lines and conformation of valyl residues in the *lac* repressor headpiece. *Biopolymers 24*:601-611.

Zuiderweg ERP , Fesik SW. 1989. Heteronuclear Three-Dimensional NMR Spectroscopy of the Inflammatory Protein C5a. *Biochemistry 28*:2387-2391.

Appendix A. Assignments table . *Assignments at pH 3.4 and 25.0° C for HSF[a] DNA-binding domain*

[a] Chemical shifts for ^{13}C and ^{1}H (in parentheses) are given in ppm.

[b] Both signals have the same chemical shift.

[c] Stereoassignments of Hβ resonances were obtained using experiments described in Chapter 3.

[t] Tentative assignment.

Residue	Amide	C α	C β	Other
A1			19.1[t] (1.34)[t]	
R2	120.9 (8.68)	54.6 (4.66)	30.1 (1.65,1.59)	Hγ (1.68), H^{δ} (2.94), N^{ε} 83.2 (7.07)
P3		63.3 (4.50)	32.5 (1.63$^{\beta 2}$, 2.57$^{\beta 3}$)	Cγ 28.2 (2.17, 2.23), C^{δ} 50.9 (3.46, 4.05)
A4	126.6 (8.79)	55.6 (4.16)	18.0 (1.51)	
F 5	116.6 (9.09)	62.6 (4.01)	39.9 (2.26$^{\beta 3}$, 3.36$^{\beta 2}$)	H^{δ} (6.92) 132.0
V6	115.4 (6.75)	65.4 (3.25)	31.6 (1.89)	$C^{\gamma 1}$ 21.6 (0.90), $C^{\gamma 2}$ 23.0 (0.80)
N7	117.5 (7.73)	56.2 (4.47)	37.8 (2.74$^{\beta 2}$, 2.80$^{\beta 3}$)	
K8	120.8 (7.97)	59.6 (3.94)	32.3 (1.34, 1.78)	Cγ 25.5 (1.25 1.50), C^{δ} 28.9 (1.46), C^{ε} 43.0 (3.21)
L9	120.5 (8.27)	58.0 (3.61)	41.7 (1.77$^{\beta 3}$, 0.75$^{\beta 2}$)	Cγ 26.5 (0.89), $C^{\delta 1}$ 22.9 (0.35), $C^{\delta 2}$ 25.1 (.09)
W10	117.7 (8.50)	61.4 (4.03)	28.8 (3.30, 3.33)	$C^{\delta 1}$ 126.2 (7.36), $C^{\varepsilon 3}$ 120.7 (7.68), $C^{\zeta 3}$ 121.9 (7.11)
				$C^{\eta 2}$ 124.5 (7.23), $C^{\zeta 2}$ 114.5 (7.53), $N^{\varepsilon 1}$ 127.7 (10.23)
S11	110.6 (7.83)	61.0 (4.10)	63.1 (4.00)	
M12	119.3 (7.92)	58.8 (3.38)	33.2 (1.01,1.83)	Cγ 30.0 (1.55, 1.66) C^{ε} 1.78 (16.5)
V13	115.7 (8.03)	65.0 (3.27)	31.3 (1.08)	$C^{\gamma 1}$ 21.1 (-0.41), $C^{\gamma 2}$ 22.2 (0.13)
N14	114.2 (7.84)	53.5 (4.70)	38.5 (2.79, 2.85)	

D15	118.6 (6.97)	53.1 (4.70)	38.6 (2.98, 3.14)	
K16	129.1 (9.16)	59.3 (4.10)	31.9 (1.94 deg[b])	C^{γ} 24.8 (1.60, 1.63), C^{δ} 28.9 (1.78), C^{ε} 42.0 (3.09)
S17	114.1 (8.34)	61.2 (4.33)	62.7 (3.83, 3.95)	
N18	117.0 (8.30)	53.1 (5.15)	39.2 (3.15, 3.02)	
E19	120.6 (7.42)	59.7 (4.57)	28.3 (2.17, 2.25)	C^{γ} 33.0 (2.60, 2.64)
K20	116.5 (8.48)	57.9 (3.91)	31.5 (1.00 deg)	C^{γ} 24.8 (1.01), C^{δ} 29.0 (1.46), C^{ε} 41.7 (2.86)
F21	114.4 (8.16)	56.4 (4.97)	42.3 ($3.09^{\beta 2}$, $3.23^{\beta 3}$)	C^{δ} 131.3 (7.11)
I22	117.8 (8.16)	60.0 (4.98)	37.7 (1.99)	$C^{\gamma 1}$ 14.1 ($0.47^{\gamma 12}$ $1.19^{\gamma 11}$), $C^{\gamma 2}$ 14.1 (0.88), $C^{\delta 2}$ 14.0 (0.30)
H23	116.6 (8.33)	53.8 (4.87)	30.4 (3.47, 3.65)	
W24	120.8 (8.93)	56.5 (5.32)	31.3 ($3.16^{\beta 2}$, $3.34^{\beta 3}$)	$C^{\delta 1}$ 128.7 (7.39), $C^{\varepsilon 3}$ 120.9 (7.70), $C^{\zeta 3}$ 121.7 (6.83)
				$C^{\eta 2}$ 123.2 (6.18), $C^{\zeta 2}$ 113.9 (6.42), $N^{\varepsilon 1}$ 128.6 (9.91)
S25	119.4 (9.05)	58.2 (4.87)	63.3 (3.94, 4.10)	
T26	116.4 (8.55)	64.5 (4.19)	69.1(4.25)	C^{γ} 22.1 (1.27)
S27	117.1 (8.58)	(3.85)		
G28	101.9 (6.70)	45.6 (3.43, 4.07)		
E29	116.7 (8.06)	56.4 (4.50)	29.1 (1.91, 2.27)	C^{γ} 32.6 (2.36, 2.40)
S30	112.0 (8.24)	57.4 (5.83)	65.3 ($3.94^{\beta 2}$, $4.01^{\beta 3}$)	
I31	122.3 (9.19)	60.9 (4.32)	41.8 (1.29)	$C^{\gamma 1}$ 29.2 ($0.20^{\gamma 12}$, $1.45^{\gamma 13}$), $C^{\gamma 2}$ 17.5 (0.34), $C^{\delta 2}$ 12.6 (-0.38)
V32	125.3 (8.92)	60.9 (4.52)	34.2 (1.10)	$C^{\gamma 1}$ 20.5 (0.60), $C^{\gamma 2}$ 20.8 (0.74)
V33	127.0 (9.08)	58.7 (4.57)	33.5 (2.27)	$C^{\gamma 1}$ 21.1 (0.76), $C^{\gamma 2}$ 20.2 (0.49)
P34		65.3 (4.13)	32.2(2.05, 2.50)	C^{γ} 27.9 (1.80, 2.18), C^{δ} 50.8 (3.73, 3.49)
N35	111.35 (8.24)	52.4 (4.80)	40.3 (2.74, 2.76)	
R36	131.3 (9.56)	60.7 (3.51)	29.6 (1.41, 1.70)	C^{γ} 24.9 (0.82, -0.01), C^{δ} 43.2 (2.32, 2.44), N^{ε} 83.4 (6.97)
E37	118.4 (8.50)	59.8 (3.96)	27.2 (2.09, 2.21)	C^{γ} 32.7 (2.41).

R38	116.8 (8.32)	58.8 (4.08)	30.0 (1.83, 1.92)	C^{δ} 42.8 (3.21, 3.26), N^{ε} 83.4 (7.32)
F39	117.8 (7.82)	61.6 (4.00)	39.8 (3.20, 3.23)	C^{δ} 131.2 (6.96), C^{ε} 130.3 (6.11), C^{ζ} 128.0 (6.10)
V40	121.8 (8.48)	66.1 (3.20)	31.3 (2.13)	$C\gamma^{1}$ 21.6 (0.96), $C\gamma^{2}$ 24.7 (1.11)
Q41	114.7 (7.83)	58.1 (3.99)	29.3 (1.95, 2.03)	$C\gamma$ 33.5 (2.31, 2.53)
E42	110.5 (7.79)	56.5 (4.29)	29.6 ($1.12^{\beta 2}$, $1.59^{\beta 3}$)	$C\gamma$ 32.6 (2.11, 2.16)
V43	116.2 (7.40)	64.1 (4.05)	33.2 (1.65)	$C\gamma^{1}$ 23.0 (0.87), $C\gamma^{2}$ 23.5 (0.84)
L44	118.9 (6.79)	60.1 (3.73)	39.6 ($1.65^{\beta 3}$, $0.56^{\beta 2}$)	$C\gamma$ 26.3 (1.35), $C^{\delta 2}$ 25.4 (0.56), $C^{\delta 1}$ 24.6 (0.48)
P45		64.9 (4.45)	31.5($1.62^{\beta 2}$,$0.56^{\beta 3}$)	C^{δ} 50.9 (2.82, 3.83)
K46	113.9 (7.36)	57.6 (3.80)	31.9 (1.73, 1.31)	$C\gamma$ 25.5 (1.08, 1.32), C^{δ} 28.9 (1.63), C^{ε} 41.7 (2.92)
Y47	114.3 (7.26)	59.1 (4.02)	42.4 ($2.22^{\beta 2}$, $2.15^{\beta 3}$)	C^{δ} 133.8 (6.96), C^{ε} 117.4 (6.68)
F48	117.6 (8.06)	55.8 (4.79)	40.3 ($2.61^{\beta 2}$, $3.12^{\beta 3}$)	C^{δ} 132.8 (7.35), C^{ε} 130.6 (7.20), C^{ζ} 129.3 (6.85)
K49	122.0 (8.45)	57.2 (4.07)	32.6 (1.48, 1.68)	$C\gamma$ 24.8 (1.19, 1.26), C^{δ} 28.9 (1.58), C^{ε} 41.7 (2.87)
H50	116.6 (8.89)	55.0 (4.62)	27.5 ($3.22^{\beta 2}$, $3.39^{\beta 3}$)	
S51	114.0 (8.02)	57.3 (4.65)	65.0 (3.65 deg)	
N52	116.6 (8.31)	51.8 (4.69)	39.2 (3.01, 3.16)	
F53	121.5 (9.41)	62.9 (4.03)	39.0 (2.85, 3.05)	C^{δ} 132.6 (6.92), C^{ε} 130.2 (6.76), C^{ζ} 128.8 (6.42)
A54	119.2 (8.49)	55.1 (3.95)	20.3 (1.43)	
S55	114.3 (8.33)	61.5 (4.08)	62.7 (3.90, 4.00)	
F56	124.2 (7.81)	60.7 (4.05)	38.7 (2.90, 3.02)	C^{δ} 131.5 (6.62), C^{ε} 130.6 (6.60), C^{ζ} 128.3 (6.47)
V57	118.5 (8.23)	66.7 (2.85)	31.1 (1.84)	$C\gamma^{1}$ 21.6 (0.96), $C\gamma^{2}$ 23.9 (0.27)
R58	120.0 (7.85)	59.7 (3.94)	29.8 (1.84 deg)	$H\gamma$ (1.48, 1.73), C^{δ} 43.2 (3.21 deg), N^{ε} 83.8 (7.25)
Q59	119.1 (7.67)	59.2 (3.66)	28.0 ($0.60^{\beta 3}$, $1.34^{\beta 2}$)	$C\gamma$ 34.5 (1.83, 2.07)
L60	118.9 (7.93)	58.8 (3.68)	39.9 ($1.06^{\beta 3}$, $1.53^{\beta 2}$)	$C\gamma$ 25.8 (0.50), $C^{\delta 2}$ 23.2 (-0.12), $C^{\delta 1}$ 24.8 (-0.92)
N61	116.2 (8.06)	56.2 (4.46)	39.0 (2.87, 2.95)	

M62	120.1 (7.79)	58.3 (4.00)	31.6 (1.77, 1.94)	C^{γ} 31.3 (2.17, 2.24)
Y63	115.4 (7.43)	58.2 (4.67)	38.0 ($2.69^{\beta 2}$, $3.76^{\beta 3}$)	C^{δ} 133.9 (7.25), C^{ε} 117.6 (6.71)
G64	105.5 (7.59)	45.8 (3.62, 4.17)		
W65	121.4 (8.39)	57.0 (4.84)	30.4 ($2.66^{\beta 3}$, $2.98^{\beta 2}$)	$C^{\delta 1}$ 126.1 (7.02), $C^{\varepsilon 3}$ 119.5 (7.37), $C^{\zeta 3}$ 121.9 (7.15)
				$C^{\eta 2}$ 124.5 (7.31), $C^{\zeta 2}$ 114.5 (7.70), $N^{\varepsilon 1}$ 128.8 (10.52)
H66	118.3 (8.91)	53.5 (5.12)	31.5 ($3.26^{\beta 3}$, $3.36^{\beta 2}$)	
K67	125.6 (9.36)	56.9 (4.49)	33.2 (1.72, 1.88)	C^{ε} 41.7 (2.12, 2.40)
V68	126.3 (8.48)	62.0 (4.15)	33.1 (1.82)	C^{γ} 21.1 (0.87), H^{γ} (0.87)
Q69	125.4 (8.60)	55.5 (4.28)	29.4 (1.91, 2.00)	C^{γ} 33.8 (1.95, 2.24)
D70	122.0 (8.45)	52.8 (4.65)	38.6 (2.70, 2.80)	
V71	120.1 (8.14)	62.4 (4.08)	32.4 (2.01)	C^{γ} 21.2 (0.85), C^{γ} 20.2 (0.85)
K72	123.9 (8.34)	56.6 (4.29)	32.6 (1.71, 1.80)	C^{γ} 25.5 (1.40, 1.45), C^{δ} 28.9 (1.68), C^{ε} 41.7 (2.96)
S73	115.5 (8.19)	58.3 (4.43)	63.8 (3.84, 3.88)	
G74	110.6 (8.45)	45.6 (3.92, 4.01)		
S75	115.2 (8.20)	58.2 (4.43)	63.8 (3.85, 3.93)	
M76	121.6 (8.43)	55.8 (4.48)	31.9 (2.05, 2.11)	C^{γ} 31.9 (2.52, 2.61)
L77	121.7 (8.18)	55.0 (4.32)	42.0 (1.58, 1.63)	C^{γ} 27.1 (1.59), C^{δ} 23.2 (0.85), C^{δ} 25.1 (0.89)
S78	115.0 (8.14)	58.3 (4.41)	63.7 (3.86, 3.93)	
N79	120.2 (8.42)	53.2 (4.70)	38.5 (2.84)	
N80	118.6 (8.42)			
D81	118.9 (8.37)			
S82	116.4 (8.20)	59.0 (4.55)	64.4 (4.00 deg)	
R83	122.5 (7.93)	55.2 (4.85)	31.0 (1.72, 1.80)	C^{γ} 27.5 (1.35, 1.50), C^{δ} 43.6 (3.12 deg), N^{ε} 84.4 (7.33)

W84	127.3 (9.00)	56.6 (4.52)	31.5 (2.93, 3.00)	$C^{\delta 1}$ 125.6 (7.20), $C^{\varepsilon 3}$ 120.3 (7.50), $C^{\zeta 3}$ 122.0 (7.33) $C^{\eta 2}$ 124.6 (7.47), $C^{\zeta 2}$ 114.8 (7.67), $N^{\varepsilon 1}$ 129.6 (10.08)
E85	121.9 (8.51)	53.5 (5.33)	31.0 (1.80, 1.89)	C^{γ} 33.4 (2.19, 2.36)
F86	120.7 (9.29)	55.7 (5.52)	43.1 (2.80$^{\beta 2}$, 3.01$^{\beta 3}$)	C^{δ} 132.0 (6.92) C^{ε} 130.5 (7.24)
E87	121.0 (9.38)	54.4 (5.33)	32.2 (1.99, 2.02)	C^{γ} 32.6 (2.43, 2.46)
N88	123.9 (8.56)	52.1 (4.37)	37.6 (0.96, 1.90)	
E89	123.0 (8.50)	55.9 (4.35)	28.5 (1.91, 2.11)	C^{γ} 32.6 (2.40, 2.42)
R90	120.4 (8.19)	56.0 (4.22)	30.1 (1.71, 1.77)	C^{δ} 43.4 (3.18), N^{ε} 84.2 (7.18)
H91	119.6 (8.29)	54.8 (4.66)	28.8 (3.22, 3.25)	
A92	128.3 (8.33)	52.5 (4.24)	19.1 (1.34)	

Appendix B Stereoassignment Table : Stereoassignment Table for HSF at 25° C, pH 3.4

[a]The first three columns list the single amino acid code for each residue in HSF, the stereoassignment of the H_β, and the chemical shifts of the two H_β nuclei respectively. The next eight columns list the intensity of the relevant crosspeak at the H_β chemical shift indicated for the given row in the following experiments: 3D ^{15}N separated NOESY-HMQC (Ch. 2) H_N-H_β, 3D ^{13}C-separated NOESY-HMQC (Ch. 2) H_α-H_β, 4D $^{13}C/^{13}C$-edited NOESY-HMQC (Ch. 2) Hα-Hβ, 3D ^{13}C-separated ROESY-HMQC H_α-H_β, short mixing time 3D ^{15}N-separated TOCSY-HMQC (Ch. 3) H_N-H_β, 2D ^1H-^1H DQF-COSY (Ch. 2) H_α-H_β, and 3D ^{15}N-separated HNHB (Ch. 2) H_N-H_β. The last column lists the dihedral angle assigned to the sidechain corresponding to the row.

[b]S, M, W and VW stand for strong, medium, weak, and very weak crosspeak observed in the relevent experiment at the indicated chemical shift. OL indicates that resonance overlap prevented quantification of the relevant crosspeak. deg indicates that degenerate H_β chemical shifts prevented stereoassignment. PO indicates that there was partial resonance overlap which may alter observed intensities for this crosspeak. NS indicates that the peak was not observed in the relevant spectrum.

[c]Uppercase t stands for tentative chemical shift assignment.

[d]These residues were stereoassigned using spin-echo difference ct-HSQC spectra described in Chapter 2.

[e]Proline H_β resonances were assigned by comparing intensities of NOEs from both H_β protons to the intraresidue H_α proton in 4D $^{13}C/^{13}C$-edited NOESY-HMQC (Ch. 2) and 3D ^{13}C-separated ROESY-HMQC spectra (Ch. 3) assigning $H_{\beta 3}$ to the H_β resonance with the stronger crosspeak.

[f]FR indicates that the χ_1 angle shows evidence of free rotation. NA stands for Not Applicable (for example methyl groups in Alanines). OL and a residue name indicate that overlap with resonances of the named residue prevented assignment. ? indicates a tentative stereoassignment not included in the dihedral and NOE restraints.

[g]Val-13 showed evidence of rotamer-averaging with a preference for $\chi_1 = 180°$.

Table B Stereoassignments data and analysis

Stereoassignment Table for HSF at 25° C, pH 3.4

[a]RES #	stereo pair	Hβ PPM[c]	15N NOE	13C NOE	4D NOE	13C ROE	15N TOC	DQF-COSY	15N HNHB	χ_1[f]
A1	NA									NA
R2	Hβ2	1.65	S[b]				S		S	
R2	Hβ3	1.59	S				S[PO]		S	
P3[e]	Hβ2	1.65			M	M				
P3	Hβ3	2.60			S	S				NA
A4	NA									NA
F5	Hβ2	3.36	S	M	S	VW	NS	W[PO]	NS	
F5	Hβ3	2.26	S	M	W	NS	NS	S[PO]	VW	180°
V6[d]	Hβ	1.89	S						M	180°
N7	Hβ2	2.74	S	OL		OL	S	OL	NS	
N7	Hβ3	2.80	M	OL		OL	NS	OL	S	-60°
K8		1.34	M				NS		S	
K8		1.78	M				NS		M	FR

[a]RES #	stereo pair	Hβ PPM	15N NOE	13C NOE	4D NOE	13C ROE	15N TOC	DQF-COSY	15N HNHB	χ1
L9	Hβ2	0.75	S	S		S	NS	W	VW	
L9	Hβ3	1.77	S[a]	NS		NS	NS	S	NS	180°
W10	deg	3.30								
W10	deg	3.33								deg
S11	deg									
S11	deg									deg
M12		1.01	S			NS	NS		M	
M12		1.83	S			NS	NS		M	
V13[d]		1.08	M			VW	NS	S	S	180°[g]
N14	deg	2.79								
N14	deg	2.85								deg
D15		2.98	S				streak		NS	
D15		3.14	S				streak		NS	?
K16	deg	1.94								
K16	deg	1.94								
S17		3.83								OL
S17		3.95								S55
N18		3.02	W			NS		NS	S	OL
N18		3.15	W			S		S	W	N52
E19		2.17	M				NS	S	S	
E19		2.25	M				NS	S	S	FR
K20	deg	1.00								
K20	deg	1.00								
F21	Hβ2	3.09	M	M		NS	S	S	NS	
F21	Hβ3	3.23	W	S		M	NS	NS	S	-60°
I22[d]		1.99	VW			S	NS	W	NS	180°
H23		3.47	W	M		H2O	NS		S	
H23		3.65	W	M		H2O	NS		NS	?
W24	Hβ2	3.16	S	M		?	W	S	W	
W24	Hβ3	3.34	M	S		NS	W	NS	S	-60°
S25		3.94	M				W		W	
S25		4.10	W				W		W	FR
T26	deg	4.19[t]								
T26	deg	4.25[t]								
S27		NA								
S27		NA								
G28	Hα	3.43								
G28	Hα	4.07								
E29		1.91	S	S		M	S	S	VW	
E29		2.27	W	S		M	W	W	S	?
S30	Hβ2	3.94	W	W		M	NS	S	NS	
S30	Hβ3	4.01	NS	M		S	NS	W	M	+60°?
I31[d]		1.29	M				M		NS	-60°
V32[d]		1.10	M			NS	M	S	NS	180°
V33[d]		2.27	M			NS	M	S	W?	180°
P34[e]	Hβ2	2.05			M					
P34	Hβ3	2.50			S					NA
N35	deg	2.74								
N35	deg	2.76								
R36		1.41	S				NS		NS	
R36		1.70	S				NS		NS	?

[a]RES #	stereo pair	Hβ PPM	15N NOE	13C NOE	4D NOE	13C ROE	15N TOC	DQF-COSY	15N HNHB	χ1
E37		2.09	S			M	NS		S	
E37		2.21	S			M	NS		M	FR
R38		1.83	M	OL		W	NS	S	W	
R38		1.92	S	OL		S	NS	W	S	?
F39	deg	3.20								
F39	deg	3.23								
V40[d]		2.13	S			W	NS		M/W	180°
Q41	deg	1.95	S			M	W(deg)		S(deg)	
Q41	deg	2.03	S			M	W(deg)		S(deg)	
E42		1.12	S	OL		NS	S	S	W	
E42		1.59	M	OL		M	NS	NS	S	-60°
V43		1.65	?			NS	M		NS	180°
L44		0.56	OL		S	M	NS	W	W	
L44		1.65	S		M	-	NS	S	W	180°
P45[e]	Hβ2	1.63			M					
P45	Hβ3	2.27			S					NA
K46		1.31	S	OL		NS	VW		S	
K46		1.73	S	OL		NS	W		W	?
Y47	Hβ2	2.22	M	OL		NS	S	S	NS	
Y47	Hβ3	2.15	W	OL		M	NS	NS	S	-60°
F48	Hβ2	2.61	M	M		NS	S	S	W	
F48	Hβ3	3.12	W	S		M	NS	VW	S	-60°
K49	Hβ2	1.68	OL	M		W	W	S	W	
K49	Hβ3	1.48	S	M		S	NS	W	S	-60°
H50	Hβ2	3.22	S			NS	S	S	M	
H50	Hβ3	3.39	W			S	W	W	S	-60°
S51	deg	3.65								
S51	deg	3.65								
N52		3.01	NS			W	NS	NS	NS	
N52		3.16	NS			S	NS	S	S	FR?
F53	Hβ2	2.85	S		W	NS	NS	?	NS	
F53	Hβ3	3.05	M		W	W	NS	?	W	-60°?
A54	NA									
S55		3.90								OL
S55		4.00								S17
F56		2.90	S		W	NS	NS	W	NS	
F56		3.02	M		W	NS	NS	S	W	?
V57[d]		1.84	S			NS	W		M	180°
R58	deg	1.84								
R58	deg	1.84								
Q59	Hβ2	1.34	M		VW	NS	W	S	W	
Q59	Hβ3	0.60	M		W	M	NS	NS	S	-60°
L60	Hβ2	1.53	M/S			S		S	W	
L60	Hβ3	1.06	M			W		NS	S	-60°
N61		2.87	S			M	NS		S	
N61		2.95	S			M	NS		W	FR?
M62		1.77	S				W		S	
M62		1.94	S				W		S	FR?
Y63	Hβ2	2.69	M	S		NS	S	S	W	
Y63	Hβ3	3.76	NS	M		W	NS	NS	S	-60°
G64	Hα	3.62								
G64	Hα	4.17								

[a]RES #	stereo pair	Hβ PPM	15N NOE	13C NOE	4D NOE	13C ROE	15N TOC	DQF-COSY	15N HNHB	χ1
W65	Hβ2	2.98	S		VW	M	W	S	W	
W65	Hβ3	2.66	M		W	M/W	W	W	S	-60°
H66	Hβ2	3.26	W			NS	NS	S	NS	
H66	Hβ3	3.36	M			S	W	W	S	-60°
K67		1.72	S			NS	W		W	
K67		1.88	S			NS	W		M	FR?
V68		1.82	S			M	M/W		M	FR?
Q69		1.91	OL				S		M	
Q69		2.00	M				S		S	FR
D70		2.70	NA				VW		S	
D70		2.80	NA				VW		M	FR?
V71		2.01	M				M		S	FR
K72		1.71	M				S		?	
K72		1.80	W				S		S	FR
S73	deg	3.84								
S73	deg	3.88								
G74	Hα	3.92								
G74	Hα	4.01								
S75		3.85	S						S(OL)	
S75		3.93	S						S(OL)	FR
M76		2.05	M						S(OL)	
M76		2.11	M						S(OL)	FR
L77		1.58	S			M(OL)		M(OL)	S(OL)	
L77		1.63	S			M(OL)		M(OL)	S(OL)	FR
S78		3.86	S						S	
S78		3.93	S						S	FR
N79	deg	2.84[t]								
N79	deg	2.84[t]								
N80		NA								
N80		NA								
D81		NA								
D81		NA								
S82		NA								
S82		NA								
R83		1.72	S				S		deg	
R83		1.80	S				M		deg	FR?
W84		2.93	S			M	M	W	S	
W84		3.00	S			M	NS	NS	NS	FR
E85		1.89	M			deg		S	NS	
E85		1.80	M			deg		W	NS	?
F86	Hβ2	2.80	S			NS	M	S	NS	
F86	Hβ3	3.01	W			VW	NS	NS	S	-60°
E87	deg	1.99								
E87	deg	2.02								
N88		0.96	W			NS	NS		NS	
N88		1.90	W			NS	M	OL	NS	?
E89		1.91	S			W	S		M	
E89		2.11	W			W	M		S	FR
R90		1.71	S						S	
R90		1.77	S						S	FR
H91	deg	3.22								
H91	deg	3.25								
A92	NA									

Appendix C AMX-600 pulse programs

P.S. #1 2D DQF-COSY (cosydfprtp.ok) 131

P.S. #2 2D TOCSY (mlevcprtp.ok) 132

P.S. #3 2D NOESY (noesyprtp.m) 133

P.S. #4 HMQC (inv4jr1d.bv) 133

P.S. #5 11-jr HMQC 1D (inv4jr1d.bv) 134

P.S. #6 11-jr HMQC (hmqcjrY.bv) 134

P.S. #7 fast HSQC (fasthsqc.al) 135

P.S. #8 3D ^{15}N-TOCSY-HMQC (invtoc3d.gab) 136

P.S. #9 3D ^{15}N-NOESY-HMQC (noeinv3d.ffd) 137

P.S. #10 CBCANH (cbcanh.sg) 138

P.S. #11 CBCA(CO)NH (cbcaconh.sg) 141

P.S. #12 ct-HSQC [filters for aromatics] (ct-hsqced31.jgp) 145

P.S. #13 ct-HSQC (cthsqcRNA.jgp) 146

P.S. #14 3D ^{13}C-NOESY-HMQC (noe3d13cX.jgp) 147

P.S. #15 3D HCCH TOCSY (hcchtoc.al) 148

P.S. #16 4D ^{13}C/^{13}C NOESY-HMQC (hcchnoe4d.al) 150

P.S. #17 4D ^{13}C/^{15}N NOESY-HMQC (chhn4d.jgp) 151

P.S. #18 2D ^{13}C-{^{13}CO} spin echo difference ct-HSQC (cthsqcjcc.jgp) 153

P.S. #19 2D ^{13}C-{^{15}N} spin echo difference ct-HSQC (cthsqcjnc.jgp) 155

P.S. #20 ct-HCCH-TOCSY (hcchdipct.jgp) 157

P.S. #21 2D ct-HCCH-TOCSY [aromatics] (hcchdpctar.jgp) 159

P.S. #22 2D ^{15}N{^{1}H} NOE HSQC (het15nonoe.jgp) 161

P.S. #23 2D ^{15}N{^{1}H} NOE HSQC (het15nnoe.jgp) 162

P.S. #24 HSQC (hsqcYdecC.jgp) 163

P.S. #25 2D LRCC (lrcc.jgp) 164

P.S. #26 2D LRCH (lrch.jgp) 166

P.S. #27 2D LRCH (lrchref.jgp) 167

P.S. #28 3D HNHB (hnhb.jgp) 168

P.S. #29 3D ^{15}N-HOHAHA-HMQC (hohaha15n.jgp) 169

P.S. #30 3D ^{13}C-ROESY-HMQC (roesyhmqc3d.jgp) 171

P.S. #1 2D DQF-COSY

(cosydfprtp.ok)

```
;cosydfprtp.ok  renamed from cosydfprtp.r
(3/26/93)
;2D homonuclear shift correlation
;phase sensitive using TPPI
;with double quantum filter
;and presaturation during relaxation delay

1 ze
2 d1
3 d12 hl2
  p18 ph29
  d13
  d12 hl1
  p1 ph1
  d0
  p1 ph2
  d13
  p1 ph3
  go=2 ph31
  d1 wr #0 if #0 id0 ip1 zd
  lo to 3 times td1
exit

;The following are the phase cycles released
by Bruker:
;ph1=0 2 1 3
;ph2=0 0 1 1 2 2 3 3
;ph3=0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1
;   2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3
;ph31=0 2 2 0 2 0 0 2 3 1 1 3 1 3 3 1
;    2 0 0 2 0 2 2 0 1 3 3 1 3 1 1 3


;The following are the improved phase
cycles that cancel rapid-
;pulsing double-frequency artifacts.  From
Derome and Williamson,
;JMR, 88, 177-185, (1990).
ph1=1 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2
ph2=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph3=1 2 3 0 2 3 0 1 3 0 1 2 0 1 2 3
ph31=0 3 2 1 1 0 3 2 0 3 2 1 1 0 3 2
ph29=0
;DP=hl1: ecoupler high power level
;p1 :  90 degree transmitter high power
pulse
;p18: presaturation during relaxation delay
;d0 : incremented delay (2D)
[3 usec]
;d1 : relaxation delay; 1-5 * T1
minimum: [30 msec]
;d12: delay for power switching
[20 usec]
;d13: short delay (e.g. to compensate delay
line)   [3 usec]
;in0: 1/(2 * SW) = DW
;nd0: 2
;NS: 16 * n
;DS: 4 (if old phase cycles are used, DS: 2
or 4)
;td1: number of experiments
;MC2: TPPI
;AQ_mod: qseq
```

P.S. #2 2D TOCSY
(mlevcprtp.ok)

```
;homonuclear Hartman-Hahn transfer using
;compensated MLEV17 sequence for mix
;one power level for excitation and SL
;phase sensitive using TPPI
;d17 delays in mixing period suppress
;effects of cross-relaxation
;C. Griesinger, et al. JACS, 110:7870
;A. Bax * D. G. Davis, J. Magn. Reson.
;65, 355-360 (1985)
;with MLEV17y instead of MLEV17x

1 ze
2 d11
3 d12 hl2
  d6
  p18 ph29
  d13
  d12 hl3
  p1 ph1
  d0
  (p17 ph26)
4 (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph22 d17 p7 ph23 d17 p6 ph22)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p6 ph24 d17 p7 ph25 d17 p6 ph24)
  (p5 ph23)
  lo to 4 times l1
  (p17 ph26)
  go=2 ph31
  d11 wr #0 if #0 id0 ip1 zd
  lo to 3 times td1
exit


ph1 =0 2 2 0 1 3 3 1
ph22=0 2 0 2 1 3 1 3
ph23=1 3 1 3 2 0 2 0
ph24=2 0 2 0 3 1 3 1
ph25=3 1 3 1 0 2 0 2
ph26=1 3 1 3 2 0 2 0
ph31=0 2 2 0 1 3 3 1
ph29=0

;hl2: ecoupler low power level for presat
;hl3: ecoupler power level for 1st pulse
;        and MLEV spinlock (7 dB)
;
;p1 : 90 degree ecoupler pulse
;p5 : 60 degree ecoupler pulse
;p6 : 90 degree ecoupler pulse (10-15usec)
;p7 :180 degree ecoupler pulse
;p17: trim pulse (2.5msec)
;p18: presaturation during recycle delay
;
;d0 : incremented delay (2D) (3usec)
;d17: compensation time during mixing
(10-40usec)
;d11: delay for disk I/O (30 msec)
;d12: delay for power switching (20 usec)
;d13: short delay (e.g. to compensate delay
line) (3 usec)
;
;L1 : loop for MLEV cyle:
;τm=((p6*64)+p5+(32*d17))*L1+(p17*2)
;
;in0: 1/(2*SW) = DW for AQ_mode=qseq
;nd0: 2
;NS : 8 * n
;DS : 2 or 4
;td1: number of experiments
;MC2: TPPI
```

P.S. #3 2D NOESY (noesyprtp.m)

```
;noesy
;with presat during RD and mixing time
;quad in t1 using TPPI

1 ze
2 d11
3 d12 hl2
  d6
  p18 ph29
  d13
  d12 hl1
  p1 ph1
  d0
  p1 ph2
  d9 hl2
  p19 ph29
  d13
  d12 hl1
  p1 ph3
  go=2 ph31
  d11 wr #0 if #0 id0 ip1 zd
  d11 id9
  lo to 3 times td1
exit

ph1= 0 2
ph2= 0 0 0 0 0 0 0 0
     2 2 2 2 2 2 2 2
ph3= 0 0 2 2 1 1 3 3
ph31=0 2 2 0 1 3 3 1
     2 0 0 2 3 1 1 3
ph29=0

;DP=hl1: ecoupler high power level
;hl2: ecoupler power level for presaturation
;p1 :  90 degree transmitter high power
;p18: presaturation during relaxation delay
;p19: presaturation during mixing time-
NOTE tm=p19+d9
;d0 : incremented delay  (2D)        [3 usec]
;d6 : additional relaxation delay
;d9 : set to 5ms
;d11: delay for disk I/O             [30 msec]
;d12: delay for power switching  [20 usec]
;d13: short delay                    [3 usec]
;NS : 8 * n
;DS : 2 or 4
;MC2: TPPI
;in0: 1/(2*swh)=dw
```

P.S. #4 HMQC (inv4jr1d.bv)

```
; hmqc15n.jgp most recently edited for X
channel=15N -ffd
;  states with dante presat
;
d8=aq
d11=50m
d12=11m
p22=d2

1 ze
2 d11
3 d12
  d12 dblo
4 d1 hl2
;5 (p18 ph15 p18 ph16 p18 ph17 p18
ph18)   ; dante
;  lo to 5 times 1800
  p18 ph0 ; on resonance presat
  10u hl1
6 p1 ph1
  d2 dbhi
  (p3 ph3):f3
  d0
  p1*2 ph2
  d0
  (p3 ph4):f3
  d7 dblo
  2u ph0
  p22:e    ; p22=d2
  2u adc
  d8 cpdb
  dbo
  rcyc=2 ph31
    d11 wr #0 if #0 ip3 zd
    lo to 3 times 2
    d12 ip31
    d12 ip31 id0
    lo to 4 times l3
exit
  ph0=0
  ph1=0 1 2 3
  ph2=0
  ph3=0 0 0 0 2 2 2 2
  ph4=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
  ph5=(360)100
  ph31=0 3 2 1 2 1 0 3 2 1 0 3 0 3 2 1
  ph15=0
  ph16=1
  ph17=2
  ph18=3
```

P.S. #5 11-jr HMQC 1D (inv4jr1d.bv)

```
;H-1/X correlation via heteronuclear zero
;and double quantum coherence 1-D
;phase sensitive using TPPI
;with 1-1JR excitation and
;refocusing pulses
;with decoupling during acquisition
;Sklenar, V. and Bax, A., J. Mag. Res. 74,
469-479 (1987)


1 ze
2 d11
3 d1 dbhi
 d13        ; 3us
 d12 hl1    ; 20us
 p1 ph1
 d5
 p0 ph5
 d2 dbhi
 (p3 ph2):f3
 d13
 (p3 ph4):f3
 d13
 p1 ph3
 d6
 p0 ph6
 d2 dblo
 go=2 ph31 cpdb   ; garp
 d11 wr #0        ; I/O 30 ms
exit

ph1=0
ph2=0
ph3=0 0 1 1 2 2 3 3
ph4=0 2
ph5=2
ph6=2 2 3 3 0 0 1 1
ph31=0 2 2 0 0 2 2 0
```

P.S. #6 11-jr HMQC (hmqcjrY.bv)

```
;2D H-X correlation via heteronuclear zero
and double quantum
;coherence
;phase sensitive using States-TPPI
;with 11-JR water suppression
;and decoupling during acquisition
;X-nucleus on Y channel
;BFV 2/16/94

1 ze
2 d11
3 d1 dbo
4 d13
 d12 hl1
 p1 ph1
 d5
 p0 ph5
 d2 dbhi
 p3:f3 ph2
 d0
 p1 ph3
 d6
 p0 ph6
 d0
 p3:f3 ph4
 d2 dblo
 go=2 ph31 cpdb
 d11 wr #0 if #0 ip2 zd
 d11 dbo
 lo to 3 times 2
 d11 ip31 id0
 d11 ip31
 lo to 2 times l0
exit

ph1=0
ph2=0
ph3=0 0 1 1 2 2 3 3
ph4=0 2 0 2 0 2 0 2 2 0 2 0 2 0 2 0
ph5=2
ph6=2 2 3 3 0 0 1 1
ph31=0 2 2 0 0 2 2 0 2 0 0 2 2 0 0 2
```

P.S. #7 fast HSQC (fasthsqc.al)

```
;HSQC with no phase cycling,
;TPPI-STATES in t1
;D. Marion et al., JMR 85, pp. 393-399
;(1989)
;X nucleus on Y channel
;Solvent suppr. with scrabling
;pulses and presat.


d6=3u
d8=aq
d11=50m
d12=5m
d25=p3-p1
p5=p1*(8/3)
p8=p1*111 ;10,000 deg non-selec. 1H pul
p9=p1*55 ;5,000 deg non-selec. 1H pulse
p24=d4 ;

;HL2=presat power
;p18=presat pulse
;d4=1/4J


1 ze
2 p8 ph1
  p9 ph2
3 d12
  d12
  d12
4 10u hl2 dbo
  p18 ph0
  10u hl1
  p1 ph1
  d4 dbhi
  (d25 p1*2 ph1) (p3*2 ph1):f3
  d4
  p22 ph1          ;Messerle pulse = 1.8 msec
  (d25 p1 ph2) (p3 ph3):f3  ;1H 90, X 90
  d0
  (p1 ph1 d6 p5 ph2 d6 p1 ph1) ;180 1H
  d0
  (d25 p1 ph1) (p3 ph1):f3
  d4
  (d25 p1*2 ph1) (p3*2 ph1):f3
  p24:e dblo
  d7                              ;DE
  d6 adc
  d8 cpdb
  rcyc=2 ph31              ;set ns=1
d11 dbo wr #0 if #0 zd    ;write to disk
d12 ip3
  p8 ph1
  p9 ph2
lo to 3 times 2
d12 ip31
d12 ip31
d12 id0
lo to 4 times l3
exit

  ph0=1
  ph1=0
  ph2=1
  ph3=0 2        ;cycle if ns > 1.  otherwise
                 ;set ph3=0
  ph31=0 2
```

P.S. #8 3D 15N-TOCSY-HMQC (invtoc3d.gab)

```
;invtoc3d.gab
;3D sequence with H-X correlation via
heteronuclear multiple quantum
;    coherence - homonuclear correlation
via "clean" MLEV17
;phase sensitive using TPPI
;no heteronuclear decoupling during
acquisition
;two power levels for excitation and spin-
locking
;Griesinger et al., JACS 110, 7870 (1988)
;Wijmenga et al., JMR 84, 634 (1989)
;Edison et al., JMR 92, 434 (1991)

1 d11 ze
2 d11 do
3 d11*2
4 d12 hl2
  p18 ph29
  d12 hl1
  p1 ph1
  d2 dhi
  p3:d ph2
  d0
  p2 ph3
  d0
  p3:d ph4
  d2 hl3
  d5
  (p17 ph26)
5 (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph22 d6 p7 ph23 d6 p6 ph22)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p6 ph24 d6 p7 ph25 d6 p6 ph24)
  (p5 ph23)
  lo to 5 times l1
  (p17 ph26)
  go=2 ph31
  d11 do wr #0 if #0 id0 ip2 zd
  lo to 3 times l0
  d11 rd0 ip1
  d11 id5 ip3
  lo to 4 times l5
exit

ph1=0 0 0 0
ph2=0 2
ph3=0
ph4=0 0 2 2
ph22=0
ph23=1
ph24=2
ph25=3
ph26=1
ph29=0
ph31=0 2 2 0

;hl1: ecoupler high power level
;hl2: ecoupler power level for solvent
presaturation
;hl3: ecoupler low power level for spin-lock
;p1 :  90 degree transmitter high power
pulse
;p2 : 180 degree transmitter high power
pulse
;p3 :  90 degree decoupler high power pulse
;p5 :  60 degree transmitter low power
pulse
;p6 :  90 degree transmitter low power
pulse
;p7 : 180 degree transmitter low power
pulse
;p17: trim pulse                    [1-2 ms]
;p18: relaxation delay with presaturation
;d0 : incremented delay (F1 in 3D)
[3 usec]
;d2 : 1/((2J)XH)                    [3 msec]
;d6 : clean TOCSY delay, d6=p6
;d5: incremented delay (F2 in 3D)
[3 usec]
;d11: delay for disk I/O
[30 msec]
;d12: delay for power switching
[200 usec]
;in0: 1/(4 * SW(X)) = (1/2) DW(X) for t1
;nd0: 4
```

```
;iN5: 1/(4 * SW(H)) = (1/2) DW(H) for t2
no, make -> iN5 = DW(H)
;NS: 8 * n
;DS: 16
;l0: number of experiments in F1
;l1: loop for MLEV cycle - ((p6 *
96)+p5)*l1)+p17*2 = mixing time
;l5: number of experiments in F2
;MC2: TPPI in F1
;MC2: TPPI in F2
```

P.S. #9 3D ^{15}N-NOESY-HMQC (noeinv3d.ffd)

```
;noeinv3d.ffd
;3D sequence with
;  homonuclear correlation via dipolar
coupling
;  dipolar coupling may be due to noe or
chemical exchange.
;  H-1/X correlation via heteronuclear zero
and double quantum
;     coherence
;phase sensitive using TPPI
;with decoupling during acquisition
;with presaturation during relaxation delay
and mixing time

1 d11 ze
2 d11 do
3 d11*2
4 d12 hl2 dhi
  p18 ph29
  d13
  d12 hl1
  p1 ph1
  d0
  p4:d ph7
  d0
  p1 ph2
  d13
  d12 hl2
  p19 ph29
  d13
  d12 hl1
  p1 ph3
  d2
  p3:d ph5
  d5
  p2 ph4
  d5
  p3:d ph6
  d13
  d2 dlo
  go=2 ph31 cpd
  d11 do wr #0 if #0 id0 ip1 zd
  lo to 3 times l0
  d11 rd0 ip5
  d11 id5
  lo to 4 times l5
exit

ph1=0 0 0 0 2 2 2 2
ph2=0
ph3=0
ph4=0
ph5=0 2
ph6=0 0 2 2
ph7=0
ph29=0
ph31=0 2 2 0 2 0 0 2

;hl1: ecoupler high power level
;hl2: ecoupler power level for presaturation
;dl0: power level for dlo mode
;p1 :  90 degree transmitter high power
pulse
;p2 : 180 degree transmitter high power
pulse
;p3 :  90 degree decoupler high power pulse
;p4 : 180 degree decoupler high power
pulse
;p18: presaturation during relaxation delay
;p19: presaturation during mixing time
;p31: 90 degree pulse for slave timer (cpd-
sequence)
;d0 : incremented delay (F1 in 3D)
[3 usec]
;d2 : 1/((2J)XH)                    [3 msec]
;d5 : Incremented delay (F2 in 3D)
;nd0: 4
;in5: 1/(4 * SW(X)) = (1/2) DW(X)
;nd5: 4
;NS: 4 * n
;DS: 32
;l0: number of experiments in F1
;l10: number of experiments in F2
;MC2: TPPI in F1
```

P.S. #10 CBCANH (cbcanh.sg)

```
;cbcanh
;ref JMR 99 (1992) p 201-207
;7/9/92 changed p29:c8 to d29 again!

#define      IN0    59.2u
#define LOOP1       36
#define LOOP2       16

#undef COMMENT
#ifdef COMMENT
;IN0 = in0
;LOOP1 = (d18-in0)%in0 + 1 = d18%in0
;LOOP1+LOOP2 = total complex 13C incs
= d17%in0
#endif

#undef CARBON
#define CARBON
#undef NITROGEN
#define NITROGEN


;t=f1 = hydrogen frequency on water
;p1=proton 90 at hl1
;p30=proton 90 at hl3
;hl1 = 1dB
;hl3 = dipsi decoupling power 15db

;f2 = carbon frequency switched from
46ppm to 177ppm
;p4=low power carbon 90 49us at dl0
;p6=low power carbon 90 22us at dl2
;p20=low power carbon 90 53us at dl4
;p21=low power carbon 90 23.5u at dl5
p11=202.4u;shaped offset sinc

p22=1.8m     ;water trim pulse
p23=0.1m     ;water trim pulse
;p24    ;opening the receiver

;f3 = nitrogen frequency
;p7 = high power nitrogen 90 45us
;p31 = low power pulse 160 us at dbl0

;p18=presat pulse = ca. 1s

p2=2.25m
d0=4u
;d7 ca 100u = dw
d5=5m        ;universal dd, id, ip delay
d8=aq
d11=50m
d13=219.2u
d14=d13-p21*2.0
d17=3.1m; 7.1m; 10.5m; 3.1m   ;cb to ca
transfer time/2
d18=2.1m+IN0
d19=d17-d18+IN0
d26=d0+p11+4u+d18-IN0
d27=4u
d28=d0+IN0*LOOP1+d17-d26-d27

;nitrogen constant time:
d15=11m; half constant time
d16=d15/2-5.4m
d10=d15/2
d20=d15/2
;in16=in20=in10

d21=p4-p1
d22=p4-p7
d23=p7-p1
d24=p7-p6
d25=p6-p1
d30=d17+d0
d29=d30-IN0*LOOP1
;in0=in30=in29=in18=in19=in27
;in10=in20=in15=in16

        ze
2       d5
3       d5 dd18
4       2u
5       2u
;*******     presaturation    ******
6       50m hl2
        50u dbhi
        1m dlo dl2
        50u o1
        d1
        p18 ph0
        10u hl1
;****    start 90-degree on ha/hb   ****
        (p1 ph2)
        1.5m
        2u
        (d25 p1*2 ph0) (p6*2 ph7):f2
        2u
        1.5m dl0
        (p22 ph2)     ;water trim pulse
        2u
;*******     inept to ca/cb   *******
        (p1 ph7) (p4 ph1):f2
```

```
        4u
        d0 hl3
        (p11 ph10):dp1          ;shaped
pulse on carbonyl
        4u
        d18 dl2
        ;d18+d19=d17, d0+d18 = 2.1m
        d19 cpdts
        (p6*2 ph17):f2
        d30
        (p11 ph10):dp1          ;shaped
pulse on carbonyl Bloch-Siegert
        4u
        4u dl0
;*****  inept to ca  *******
        (p4 ph19):f2
        10u
        11m dl2
        (d24 p6*2 ph9):f2 (p7*2 ph0):f3
        11m
        10u dl0
;*****  inept to nitrogen       *******
        (p4 ph0):f2
        4u to
        4u o1   ;shifting frequency to mid of
amid protons
        0.1m cpdts
        (p7 ph9):f3     ;inept to nitrogen
        d10
        (p11 ph10):dp1
        d10 dl2
        (d24 p6*2 ph10):f2 (p7*2 ph4):f3
        d20
        (p11 ph10):dp1
        d16
        2.4m to
        3m hl1 o1       ;shifting frequency
back to water
;*******  inept to protons     *******
        (d23 p1 ph0) (p7 ph0):f3
        2u
        (p23 ph2)       ;water trim pulse
        2.25m
        (d23 p1*2 ph0) (p7*2 ph10):f3
        2.25m
        (p23*2 ph0)     ;water trim pulse
        (2u ph0)
        d7 dblo
        p24:e
        3u adc
        d8 cpdb
dbo
rcyc=6 ph31
d11 wr #0 if #0 zd

#ifdef CARBON
d5 ip1
lo to 4 times 2
d5 id0
d5 dd30
d5 id19
d5 ip31
d5 ip31
lo to 3 times LOOP1

;*****  here starts LOOP2     *******
12      d5
        d5
13      20m
14      10m
15      30m
;;*****  presaturation   *******
16      50m hl2
        50u dbhi
        1m dlo dl2
        50u o1
        d1
        p18 ph0
        10u hl1
;****  start 90-degree on ha/hb  *****
        (p1 ph2)
        1.5m
        2u
        (d25 p1*2 ph0) (p6*2 ph7):f2
        2u
        1.5m dl0
        (p22 ph2)       ;water trim pulse
        2u
;*******       inept to ca/cb   *******
        (p1 ph7) (p4 ph1):f2
        4u
        d26 hl3
        d27 cpdts
        (p11 ph10):dp1          ;shaped
pulse on carbonyl
        4u
        d28 dl2
        (p6*2 ph17):f2
        d29
        (p11 ph10):dp1          ;shaped
pulse on carbonyl Bloch-Siegert
    4u
        4u dl0
;*******  inept to ca  *******
```

```
        (p4 ph19):f2
        10u
        11m dl2
        (d24 p6*2 ph9):f2 (p7*2 ph0):f3
        11m
        10u dl0
;*******  inept to nitrogen    *******
        (p4 ph0):f2
        4u to
        4u o1   ;shifting frequency to mid of
amid protons
        0.1m cpdts
        (p7 ph9):f3     ;inept to nitrogen
        d10
        (p11 ph10):dp1
        d10 dl2
        (d24 p6*2 ph10):f2 (p7*2 ph4):f3
        d20
        (p11 ph10):dp1
        d16
        2.4m to
        3m hl1 o1       ;shifting frequency
back to water
;*******  inept to protons     *******
        (d23 p1 ph0) (p7 ph0):f3
        2u
;       (p23 ph2)       ;water trim pulse
        2.25m
        (d23 p1*2 ph0) (p7*2 ph10):f3
        2.25m
;       (p23*2 ph0)     ;water trim pulse
        (2u ph0)
        d7 dblo
        p24:e
        3u adc
        d8 cpdb
dbo
rcyc=16 ph31
d11 wr #0 if #0 zd
d5 ip1
lo to 13 times 2
d5 id27
d5 dd29
d5 ip31
d5 ip31
lo to 14 times LOOP2
#endif

#ifdef NITROGEN
d5 ip9
d5 rd0
d5 rd30
d5 rd29
d5 rd27
d5 rd18
d5 rd19
lo to 2 times 2
d5 dd10
d5 id16
d5 id20
d5 ip31
d5 ip31
lo to 2 times l3
#endif
do
exit

ph0=0
ph1=0
ph2=1
ph4=0 0 0 0 2 2 2 2
ph7=0 2
ph9=0 0 2 2
ph17=0 0 0 0 0 0 0 0
   1 1 1 1 1 1 1 1
   2 2 2 2 2 2 2 2
   3 3 3 3 3 3 3 3
ph19=0
ph31=0 2 2 0 0 2 2 0
   2 0 0 2 2 0 0 2
ph10=0
ph20=0
```

P.S. #11 CBCA(CO)NH (cbcaconh.sg)

```
#undef COMMENT
#ifdef COMMENT
;sg 1/28/92

;IN0 = IN0
;LOOP1 = (d18-IN0)%in0 + 1 = d18%IN0
;LOOP1+LOOP2 = total complex 13C incs
= d17%IN0

#endif

#define      IN0     59.2u
#define LOOP1        36
#define LOOP2        14

#undef CARBON
#define CARBON
#undef NITROGEN
#define NITROGEN

;t=f1 = hydrogen frequency on water
;p1=proton 90 at hl1
;p30=proton 90 at hl3
;hl1 = 1dB
;hl3 = dipsi decoupling power 15db
;f2 = carbon frequency switched from
        46ppm to 177ppm
;p4=low power carbon 90 58.8us at dl0 at
        500 MHz
;p6=low power carbon 90 26.3us at dl2
;p20=low power carbon 90 63.9us at dl4
;p21=low power carbon 90 28.6u at dl5
p11=202.4u  ;shaped offset sinc
p22=1.8m     ;water trim pulse
p23=0.1m     ;water trim pulse
;p24    ;opening the receiver
;f3 = nitrogen frequency
;p7 = high power nitrogen 90 42us
;p31 = low power pulse 160 us at dbl0
;p18=presat pulse = ca. 1s

p2=2.25m
d0=4u
;d7 ca 100u = dw
d5=5m          ;universal dd, id, ip delay
d8=aq
d11=50m
d13=217.6u  ; freq switching 4/delta freq
d14=d13-p21*2.0
d15=5.4m
d16=4u
d17=3.1m; 7.1m; 10.5m; 3.1m  ;cb to ca
                transfer time/2
d18=2.1m+IN0
d19=d17-d18+IN0
d26=d0+p11+4u+d18-IN0
d27=4u
d28=d0+IN0*LOOP1+d17-d26-d27
d10=11m+d16+p21*2.0+10u
d20=11m-d15
d21=p4-p1
d22=p4-p7
d23=p7-p1
d24=p20-p7
d25=p6-p1
d30=d17+d0
d29=d30-IN0*LOOP1
;in0=in30=in29=in18=in19=in27
;in10=in20=in15=in16

        ze
2       d5
3       d5 dd18
4       2u
5       2u
;*******presaturation*******
6       50m hl2
        50u dbhi
        1m dlo dl2
        50u o1
        50u o2
        d1
        p18 ph0
        10u hl1
;*******start 90-degree on ha/hb*******
        (p1 ph2)
        1.5m
        2u
        (d25 p1*2 ph0) (p6*2 ph7):f2
        2u
        1.5m dl0
        (p22 ph2)       ;water trim pulse
        2u
;*******    inept to ca/cb     *******
        (p1 ph7) (p4 ph1):f2
        4u
        d0 hl3
        (p11 ph10):dp1 ;shaped carbonyl
        4u
        d18 dl2
        ;d18+d19=d17, d0+d18 = 2.1m
        d19 cpdts
```

```
        (p6*2 ph17):f2
        d30
        (p11 ph10):dp1  ; Bloch-Siegert
        4u
        4u dl0
;******* inept to ca  *******
        (p4 ph19):f2
        4u
        (p11 ph10):dp1  ;shaped carbonyl
        10u
        3.5m dl2        ;
        (p6*2 ph9):f2
        4u
        (p11 ph10):dp1;shaped carbonyl
        3.5m dl0
        10u
;*******inept to carbonyl      *******
        (p4 ph0):f2
        50u o2; shifting frequency to C'
        10u dl4
        (p20 ph8):f2
        1m
        3.5m dl5
        d13 o2;         shifting frequency to
                ; Cα    -18248 Hz
        (p21*2 ph10):f2
        d14
        5.5m o2   ;shifting frequency  to C'
        1m dl4
        (p20*2 ph10):f2(d24 p7*2 ph10):f3
        11m
        d13*2
;*******  inept to nitrogen    *******
        (p20 ph18):f2
        4u to
        4u o1   ;shifting frequency to mid of
                amid protons
        0.1m cpdts
        (p7 ph9):f3      ;inept to nitrogen
        d10
        (p20*2 ph10):f2 (d24 p7*2 ph4):f3
        d20 dl5
        d15 to
        10u o2          ;shifting frequency
                to cα -18248
        (p21*2 ph10):f2
        d16 hl1 o1      ;shifting frequency
                back to water
;*******inept to protons       *******
        (d23 p1 ph0) (p7 ph0):f3
        2u
;       (p23 ph2)          ;water trim pulse
        2.25m
        (d23 p1*2 ph0) (p7*2 ph10):f3
        2.25m
;       (p23*2 ph0)      ;water trim pulse
        (2u ph0)
        d7 dblo
        p24:e
        3u adc
        d8 cpdb
        dbo
        rcyc=6 ph31
        d11 wr #0 if #0 zd

#ifdef CARBON
d5 ip1
lo to 4 times 2
d5 id0
d5 dd30
d5 id19
d5 ip31
d5 ip31
lo to 3 times LOOP1

;*******here starts LOOP2*******
12      d5
        d5
13      20m
14      10m
15      30m
;*******presaturation*******
16      50m hl2
        50u dbhi
        1m dlo dl2
        50u o1
        50u o2
        d1
        p18 ph0
        10u hl1
;*******start 90-degree on ha/hb*******
        (p1 ph2)
        1.5m
        2u
        (d25 p1*2 ph0) (p6*2 ph7):f2
        2u
        1.5m dl0
        (p22 ph2)        ;water trim pulse
        2u
;*******      inept to ca/cb   *******
        (p1 ph7) (p4 ph1):f2
        4u
        d26 hl3
```

```
        d27 cpdts
        (p11 ph10):dp1 ;shaped pulse on C'
        4u
        d28 dl2
        (p6*2 ph17):f2
        d29
        (p11 ph10):dp1;shaped pulse on
                ;carbonyl Bloch-Siegert
        4u
        4u dl0
;*******inept to ca     *******
        (p4 ph19):f2
        4u
        (p11 ph10):dp1 ;shaped pulse on C'
        10u
        3.5m dl2        ;be careful when
                        ;using shaped pulses
        (p6*2 ph9):f2
        4u
        (p11 ph10):dp1;shaped pulse on C'
        3.5m dl0
        10u
;*******   inept to carbonyl   *******
        (p4 ph0):f2
        50u o2; shifting frequency to C'
        10u dl4
        (p20 ph8):f2
        1m
        3.5m dl5
        d13 o2          ;shifting frequency
                        ;to Ca -18248 Hz
        (p21*2 ph10):f2
        d14
        5.5m o2         ;shifting frequency
                        ;back to carbonyl
        1m dl4
        (p20*2 ph10):f2(d24 p7*2 ph10):f3
        11m
        d13*2
;*******   inept to nitrogen    *******
        (p20 ph18):f2
        10u to
        10u o1          ;shifting frequency
                ;to mid of amid protons
        0.1m cpdts
        (p7 ph9):f3     ;inept to nitrogen
        d10
        (p20*2 ph10):f2 (d24 p7*2 ph4):f3
        d20 dl5
        d15 to
        10u o2  ;shifting freq to ca -18248
        (p21*2 ph10):f2
        d16 hl1 o1      ;shifting frequency
                        ;back to water
;*******   inept to protons     *******
        (d23 p1 ph0) (p7 ph0):f3
        2u
;       (p23 ph2)       ;water trim pulse
        2.25m
        (d23 p1*2 ph0) (p7*2 ph10):f3
        2.25m
;       (p23*2 ph0)     ;water trim pulse
        (2u ph0)
        d7 dblo
        p24:e
        3u adc
        d8 cpdb
dbo
rcyc=16 ph31
d11 wr #0 if #0 zd
d5 ip1
lo to 13 times 2
d5 id27
d5 dd29
d5 ip31
d5 ip31
lo to 14 times LOOP2
#endif

#ifdef NITROGEN
d5 ip9
d5 rd0
d5 rd30
d5 rd29
d5 rd27
d5 rd18
d5 rd19
lo to 2 times 2
d5 dd10
d5 dd15
d5 id16
d5 id20
d5 ip31
d5 ip31
lo to 2 times 32
#endif
do
exit

ph0=0
ph1=0
ph2=1
ph4=0 0 0 0 0 0 0 0
  2 2 2 2 2 2 2 2
```

```
ph7=0 2
ph8=0 0 2 2
ph9=0 0 0 0 2 2 2 2
ph17=0 0 0 0 0 0 0 0
   1 1 1 1 1 1 1 1
   2 2 2 2 2 2 2 2
   3 3 3 3 3 3 3 3
ph18=(360)51            ;adjusted for BS
ph19=0
ph31=0 2 2 0 2 0 0 2
   2 0 0 2 0 2 2 0
ph10=0
ph20=0
```

P.S. #12 ct-HSQC [filters for aromatics] (ct-hsqced31.jgp)

```
 for filtering arom. or C'
;original expt
;2D 1H-13C hsqc with BB dec of 13C
;Grzesiek and Bax (1993) J. Biol. NMR 3,
;pp.185-204.
;constant time 2T tuned to 1/Jcc
;need extra 13C ext. amp !!must use dlo !!
;p3 ca 18-20us 13C90 at dl0
;p13 26.3us 90 13C at dl1 null on carbonyl
;          when carrier set to 43 ppm
;p31 ca 60-65 us at dl2 for cpd decoupling
;p6 dp0 sinc1.0 180 13C' ( 240 us ) with
;dp0 40db approx 20 kHz offset
;set d7 to get 0 linear phase corr. 1.6*DW
;states with messerle water trim pulse
d26=p3-p1
d2=1.7m
d8=aq
p11=1m
d12=7m
d20=10u
d21=13.3m
d14=13.3m
p22=1.7m
d0=10u
d15=3m
d16=10.3m


1 ze
2 d12 dlo
3 d12*4
4 d12*3
5 d1 hl2
 p18 ph0      ;on resonance presat at Hl2
 10u hl1        ;set to HP 1H
 p1 ph0         ;1H 90
 d2 dl0         ;1/4JXH 1.7ms
 2u
 (d26 p1*2 ph0) (p3*2 ph1):d  ;1H,X 180
 2u
 d2
 (p11 ph0)      ;Messerele trim pulse < 1ms
 2u
 (d26 p1 ph1) (p3 ph2):d     ;1H,X 90
 d0                  ;t1/2
 (p1*2 ph0)              ;180 1H,
 (p6 ph0):dp0          ;180 C' shaped pulse
 10u
 d14 dl1                ;time T
 (p13*2 ph6):d       ;apply selective X 180
 d21                 ;T-t1/2
 10u dl0               ;set low power X
 (p6 ph0):dp0     ;apply shaped at pwr dp0
 (p1*2):c8          ;1H 180 compensation
 10u dl0
 (d26 p1 ph0) (p3 ph3):d     ;90 1H, 90 X
 d2
 (d26 p1*2 ph4) (p3*2 ph5):d ;180 1H&X
 d7          ;adjust for flat baseline 1.6*DW
 2u ph0                ;set rec phase
 p22:e dl2         ;open rec, wait period d2
 10u adc cpds            ;start adc and cpd
 d8                  ;set to acq time
 do
 rcyc=2 ph31
 d11 wr #0 if #0 zd  ;write + increment file
 d12 ip2           ;increment ph2 for States
lo to 3 times 2
 d12 ip31    ;increment rec for States TPPI
 d12 ip31
 d12 id0               ;increment d0 by in0
 d12 dd21          ;decrement d21 by in21
lo to 4 times l3
exit
ph0=0
ph1=1 3
ph2=0
ph3=0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
ph4=0
ph5=0
ph6=(360) 83 83 263 263 173 173 353 353
;adjust for low power
ph31=0 2 0 2 2 0 2 0
   2 0 2 0 0 2 0 2
;hl1 ecoupler high power
; p1 90 1H high power at hl1
; d1 relax delay
; d2 1/4JXH
; in0 dw/2 X dimension
; in21 = in0
; ns = 16*n
; ds = 16
```

P.S. #13 ct-HSQC (cthsqcRNA.jgp)

```
;2D 1H-13C hsqc with broad-band
;decoupling of 13C
;Vuister & Bax (1992) JMR 98, pp 428
;Santoro + King (1992) JMR 97, 202-207
;constant time 2T tuned to 1/Jcc
;states quad in t1 with messerle water trim
;pulse
;15N decoupling removed 11/4/93 jgp
d26=p3-p1
d8=aq
p22=2m
d12=11m
d11=50m
d0=3u
p9=10u
d5=(p1*2)+(d0*2)

1 ze
2 d12
3 d12*4
4 d12*3
5 d1 hl2          ;recycle delay - approx
1.5s
; p18 ph0          ;on resonance presat at
Hl2
 10u hl1          ;set to Hi Power 1H
 p1 ph0           ;1H 90
 d2 dhi           ;1/4JXH 1.4ms  - 13C hi
pwr
 2u
 (d26 p1*2 ph0) (p3*2 ph1):d      ;1H,X
180
 2u
 d2
 (p22 ph0)                ;Messerele trim
pulse  2ms
 2u
 (d26 p1 ph1) (p3 ph2):d      ;1H,X 90
 d0                   ;t1/2
 (p1*2 ph0)                ;180 1H
 d0                   ;t1/2
 d21                  ;T-t1/2
 (p3 ph6 2u p3*2 ph7 2u p3 ph6):d  ;13C
180 , composite
 d21                  ;T-t1/2
 d5                   ; compensate for phase
 (d26 p1 ph0) (p3 ph3):d      ;90 1H, 90 X
 d2
 (d26 p1*2 ph4) (p3*2 ph5):d  ;180 1H,
180 C
 d7                   ;adjust for flat baseline
1.6*DW
 p9:e ph0                 ;set rec phase
 d2 dlo                   ;open rec, wait
period d2
 10u adc cpds                ;start adc and cpd
 d8                   ;set to acq time
 do
 rcyc=2 ph31
 d11 do wr #0 if #0 zd        ;write +
increment file
 d12 ip2                  ;increment ph2 for
States
lo to 3 times 2
 d12 ip31                 ;increment rec for
States TPPI
 d12 ip31
 d12 id0                  ;increment d0 by in0
 d12 dd21                 ;decrement d21 by
in21
lo to 4 times l3
exit
ph0=0
ph1=1 3
ph2=0
ph3=0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
ph4=0
ph5=0
ph6=1 1 3 3 2 2 0 0
ph7=2 2 0 0 3 3 1 1
ph31=0 2 0 2 2 0 2 0
   2 0 2 0 0 2 0 2
;hl1 ecoupler high power
; p1 90 1H high power at hl1
; p3 13C 90 high power - approx 16us
; p31 13C low power for Garp decoupling
approx 60us
; d1 relax delay
; d2 1/4JXH approx 1.4ms
; in0 dw/2 X dimension
; in21 = in0
; d12 =7ms switch
; p18 - for presat if needed at hl2 -
; p22 spinlock trim pulse for h2o
suppression <2ms
; d0 - inital t1 period 3us
; d21 1/Jcc or 2/Jcc - either 9ms or 18ms
; d7 1.6*dw
; p9 2us
; d8 set to acq time
;
```

P.S. #14 3D ^{13}C-NOESY-HMQC (noe3d13cX.jgp)

```
;noe3d13cX.jgp
;noesy 13C 3d
;abx 5/17/91
d4=1.65m
p6=d4
d8=aq
d11=50m
1 ze
2 d11
  5m
3 15m
4 10m
5 15m
6 d1 hl2 dhi ; o1          ;add in for presat
f1list hcch
;7 p18 ph0                 ;if presat is needed
  10u ; o1                 ;add in for presat at
4.77
  10u hl1
  p1 ph1        ;1H 90
  d0            ;t1/2
  (p3 ph11):d
  2u
  (p3*2 ph10):d   ;13C 180 decouple
composite
  2u
  (p3 ph11):d
  d0            ;t1/2
  p1 ph2        ;90 1H
  d19           ;mixing time not presat
  p1 ph3        ;90 1H
  p22 ph8       ;messerele spin lock 0.5ms
  d4
  (p1*2 ph0) (p3*2 ph0):d
  d4
  p23 ph0       ;messerele spin lock
1.8ms
  (p3 ph5):d    ;13C 90
  d10           ;t2/2
  p1*2 ph4      ;1H 180
  d10           ;t2/2
  (p3 ph9):d    ;13C 90
  d4
  (p1*2 ph0) (p3*2 ph0):d
  d7 ph7
  p6:e dlo      ;wait d4, turn on rec, set
13C lo
  3u adc
  d8 cpd        ;start 13C cpd - garp
do
rcyc=2 ph31
d11 wr #0 if #0 zd
5m ip1
lo to 3 times 2
5m ip31
5m id0
5m ip31
lo to 4 times 126   ;states - TPPI #
complex 1H in t1
5m ip5
5m rd0
lo to 5 times 2
5m id10
5m ip31
5m ip31
lo to 6 times 31   ;states - TPPI # complex
in t2 15N
do
exit
ph0=0
ph1=0 2
ph2=0
ph3=0
ph4=0 0 2 2
ph5=0
ph6=0
ph7=(360)174
ph16=(6)0
ph17=(6)1
ph18=(6)2
ph19=(6)3
ph26=(6)4
ph27=(6)5
ph30=0
ph31=0 2 0 2 2 0 2 0
ph8=1
ph9=0 0 0 0 2 2 2 2
ph10=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph11=1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3
10us,2d10+4/pi*13C90+1H 180=dw/2
;p3 HP 13C 18us,d19 mix time 100ms,d7
0.6*dw,d4=1/4jCH
;p22 messerele 0.5ms, p23 messerele
1.8ms
;t1 4807 H in0 104us carrier 4.0ppm
;t2 3592.8H 139.2us=in10 carrier 43 ppm
;t3 approx 4807 H  dw=104us*2
```

P.S. #15 3D HCCH TOCSY (hcchtoc.al)

```
;3D 13C HCCH-TOCSY
;sg 10-15-91
;f2 = carbon frequency at 46 ppm
;p3 = high power carbon 90 ca. 10us
;p4 = shaped pulse on carbonyl 180 degree
p5 = 2m          ;low power trim pulse at dl0
;p6 low power carbon 90 30-35 us for
dipsi/flopsy at dl0
p10=p6*.55555
;p31 low power carbon ca. 50us 90 for
waltz decoupling dl1
d21=p3-p1
d6=p3*4+d0*2+4u
;d7=de
d8=aq
d12=p4 + d10*2 +p3*0.6366 +p1*2
p7=1.5m
p8=p3+1.5u
p9=p3-1.5u

1       ze
2       50m
3       10m
4       10m
5       15m
6       10u hl2
        d1 o1
        p18 ph0
        10u o1
        10u hl1
;***************************
        p1 ph1          ;starting 90
        1.5m dhi
        d0
        (p3 ph2):f2  ;composite 180 13C
     2u
        (p3*2 ph13):f2
        2u
        (p3 ph2):f2
        d0
        p1*2 ph3
        1.5m
        d6      ;compensation delay
;*******************************
        (d21 p1 ph4) (p3 ph5):f2       ;first
inept
        2u
        d10 dlo
        (p4 ph10):dp1           ;Carbonyl
180
        4u
        1m
        (p1*2 ph6)
        2u
        d10 dhi
        (p3*2 ph7):f2
        2u
        0.5m dlo
        0.5m dl0
        6u
        d12
;*********************************
        (p5 ph12):f2   ;SL-pulse
;*******       start DIPSIy*********
8       (p10*4.9 ph15):f2
        (p10*7.9 ph16):f2
        (p10*5.0 ph15):f2
        (p10*5.5 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.6 ph16):f2
        (p10*7.2 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.4 ph15):f2
        (p10*6.8 ph16):f2
        (p10*7.0 ph15):f2
        (p10*5.2 ph16):f2
        (p10*5.4 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.5 ph15):f2
        (p10*7.3 ph16):f2
        (p10*5.1 ph15):f2
        (p10*7.9 ph16):f2
        (p10*4.9 ph15):f2
        (p10*7.9 ph16):f2
        (p10*5.0 ph15):f2
        (p10*5.5 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.6 ph16):f2
        (p10*7.2 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.4 ph15):f2
        (p10*6.8 ph16):f2
        (p10*7.0 ph15):f2
        (p10*5.2 ph16):f2
        (p10*5.4 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.5 ph15):f2
        (p10*7.3 ph16):f2
        (p10*5.1 ph15):f2
        (p10*7.9 ph16):f2
;
        (p10*4.9 ph16):f2
```

```
(p10*7.9 ph15):f2
(p10*5.0 ph16):f2
(p10*5.5 ph15):f2
(p10*0.6 ph16):f2
(p10*4.6 ph15):f2
(p10*7.2 ph16):f2
(p10*4.9 ph15):f2
(p10*7.4 ph16):f2
(p10*6.8 ph15):f2
(p10*7.0 ph16):f2
(p10*5.2 ph15):f2
(p10*5.4 ph16):f2
(p10*0.6 ph15):f2
(p10*4.5 ph16):f2
(p10*7.3 ph15):f2
(p10*5.1 ph16):f2
(p10*7.9 ph15):f2
(p10*4.9 ph16):f2
(p10*7.9 ph15):f2
(p10*5.0 ph16):f2
(p10*5.5 ph15):f2
(p10*0.6 ph16):f2
(p10*4.6 ph15):f2
(p10*7.2 ph16):f2
(p10*4.9 ph15):f2
(p10*7.4 ph16):f2
(p10*6.8 ph15):f2
(p10*7.0 ph16):f2
(p10*5.2 ph15):f2
(p10*5.4 ph16):f2
(p10*0.6 ph15):f2
(p10*4.5 ph16):f2
(p10*7.3 ph15):f2
(p10*5.1 ph16):f2
(p10*7.9 ph15):f2
lo to 8 times 3
(p5 ph12):f2      ;sl pulse
;total mixtime0.55*p6*391.2*loop
;*****  end DIPSI  ******
2u
1m dhi
(d21 p1*2 ph0) (p3*2 ph10):f2
2u
1m
(d21 p1 ph9) (p3 ph10):f2      ;inept
1.5m
(d21 p1*2 ph0) (p3*2 ph10):f2
d7
p7:e
3u adc
(p8 ph10):f2
(p9 ph11):f2
2u
2u dlo
2u dl1
d8 cpds      ;garp 60us
do
rcyc=2 ph31
50m wr #0 if #0 zd
5m ip1
lo to 3 times 2
5m id0
5m ip31
5m ip31
lo to 4 times 128
5m rd0
5m ip5
lo to 5 times 2
5m id10
5m ip31
5m ip31
lo to 6 times 32
do
exit

ph0=0
ph1=0
ph2=0 0 0 0 0 0 0 0
  2 2 2 2 2 2 2 2
ph3=0 0 0 0 1 1 1 1
  2 2 2 2 3 3 3 3
ph4=1 3
ph5=0
ph6=0
ph7=0 0 1 1 2 2 3 3
ph9=0 0 0 0 0 0 0 0
  2 2 2 2 2 2 2 2
ph10=0
ph11=0 2
ph12=(360)5 5 5 5 185 185 185
185;should be adjusted
ph13=1 1 1 1 1 1 1 1
  3 3 3 3 3 3 3 3
ph15=(360)95          ;should be adjusted
ph16=(360)275         ;should be adjusted
ph21=(6)5
ph22=(6)4
ph23=(6)3
ph24=(6)2
ph25=(6)1
ph26=(6)0
ph31=0 2 2 0 2 0 0 2
     2 0 0 2 0 2 2 0
```

P.S. #16 4D $^{13}C/^{13}C$ NOESY-HMQC (hcchnoe4d.al)

```
;4d noesy (1H-13C-13C-1H)
;11/28/93 A. Lee,hcchnoe4d.abx as templat
;ref: Clore, Kay, Bax, Gronenborn
;Biochemistry, 1991, 30, pp. 12-18
;set d10 and d20 properly for 1/2 dw
p13=1m
p24=3.4m
d11=40m
d12=6m
d6=p3*4
d16=d6+10u
d8=aq
1 ze
 thi
2 d11
 d12*2
3 d12*6
4 d12*2
5 d12*3
6 d12*2
7 d12*3
8 d12 dlo hl2 o1
 p18 ph29 dl0 o2
 (p13 ph0 2u):f2
 d12 dhi hl1 o1
 p1 ph1
 3.4m
 d16
 (p3 ph11):f2
 d21
 p1*2 ph2
 d21
 (p3 ph0):f2                ;end of first hmqc
 3.4m o2
 d0
 (p3 ph22 2u p3*2 ph12 2u p3 ph22):f2
 d0
 (p2 ph3)                   ;trim
 2u
 p1 ph4
 40m o2
 (p3 ph0):f2
 60m
 10u dlo
 (p13 ph0):f2
 (p1 ph5)
 3.4m dhi
 10u
 (p3 ph13):f2               ;begin final hmqc
 d20
 (p1*2 ph6)
 d20
 (p3 ph0):f2
 p24:e o2
 d7                         ;pre-scan delay
 3u adc
 (p3 ph0):f2
 2u
 (p3 ph14):f2
 2u
 5u dlo
 5u dl1
 d8 cpds
 10u do
 rcyc=2 ph31
 d11 wr #0 if #0 zd
 d12 dp3
 d12 dp4        ;states part of states-tppi (t2)
 lo to 3 times 2
 d12 id0
 d12 rp3
 d12 rp4
 d12 ip1
 d12 ip1
 d12 ip31
 d12 ip31       ;tppi part of states-tppi (t2)
 lo to 4 times 64       ;64 complex in f1
 d12 rd0                ;reset d0
 d12 ip11               ;quad in t1 (13C)
 lo to 5 times 2
 d12 ip31
 d12 id21
 d12 ip31
 lo to 6 times 8
 d12 rd21
 d12 ip13               ;quad in t3 (13C)
 lo to 7 times 2
 d12 ip31
 d12 id20
 d12 ip31
 lo to 8 times 8
 exit
 ph0=0
 ph1=0
 ph2=0 0 0 0 1 1 1 1
 ph3=3
 ph4=0
 ph5=0
 ph6=0 0 0 0 1 1 1 1
 ph11=0 2
 ph12=1 1 3 3
 ph22=0 0 2 2
```

```
ph13=0 0 2 2
ph14=0 2
ph29=0
ph31=0 2 2 0
ph7=2
```

P.S. #17 4D $^{13}C/^{15}N$ NOESY-HMQC (chhn4d.jgp)

```
; chhn4d.jgp
;13c, 15n edited 4d noesy
;ref: Kay et al., 1990 Science, 249, 411.
;Phase cycle can be 4 or 8

;p1     proton 90 at hl1=2
;p18    proton presat at hl3
;p3     high power c13 90 on f2 (56ppm)
;p7     high power n15 90 on f3
;p31    low power n15 90 (160ms) on f3 at
dbl0
;p15    noesy mix with presat
;p24    refocusing delay = d4

d0=in0*0.5-p1-p3*0.637        ;carbon
incremental delay
d2=3m                         ;hmqc h to
c13
d3=p3*4+d10*2.0+4u            ;refocusing
delay for protons
d4=4.5m
;d7=de
d8=aq
;d0              13C incremental delay
;d10                   proton incremental
delay
d20=in20*0.5-p1-p7*0.637      ;nitrogen
incremental delay
d11=50m
d12=11m

        ze
2       d11
        d12
3       d12*3.0
4       d12*2.0
5       d12*3.0
6       d12*3.0
7       d12*3.0
;*****          presaturation    *****
8       10u hl2
        50u dbhi dhi
        p18 ph0
        10u hl1
;*****          start 90-degree on h-c13
        *****
        p1 ph1
        d2
        d3
        (p3 ph2):f2
;*****          carbon evolution *****
        d0
        (p1*2 ph0)
        d0
        (p3 ph0):f2
        d2
;*****          proton evolution *****
        d10
        (p3 ph0 2u p3*2 ph5 2u p3 ph0):f2
        d10
        (p1 ph0)
;*****          noesy mixing   *****
        10u hl2
        p15 ph0
        2u
        5u hl1
        p1 ph0
        d4
;*****          hmqc to nitrogen
        *****
        (p7 ph3):f3
;*****          nitrogen evolution *****
        d20
        (p1*2 ph12)
        d20
;*****          hmqc to proton *****
        (p7 ph0):f3
        (2u ph0)
        d7 dblo
        p24:e
        3u adc
        d8 cpdb
dbo
rcyc=2 ph31
d11 wr #0 if #0 zd
d12 ip1                       ;protons
lo to 3 times 2
d12 id10
d12 ip31
d12 ip31
lo to 4 times 64
d12 ip2                       ;carbons
d12 rd10
lo to 5 times 2
d12 id0
d12 ip31
```

```
d12 ip31
lo to 6 times 8
d12 ip3                    ;nitrogens
d12 rd0
lo to 7 times 2
d12 id20
d12 ip31
d12 ip31
lo to 8 times 8
do
dbo
exit


ph0=0
ph1=0
ph2=0 2 0 2
ph3=0 0 2 2
ph4=1
ph5=1
ph12=0 0 0 0 1 1 1 1
ph31=0 2 2 0 2 0 0 2
```

P.S. #18 2D ^{13}C-{^{13}CO} spin echo difference ct-HSQC (cthsqcjcc.jgp)

```
;2D 1H-13C hsqc with broad-band
decoupling of 13c for
;JCC coupling constants
;Gresiek, Vuister, Bax (1992) JBiomol.
NMR 3 (1993) 487-493.
;constant time 2T tuned to 2/Jcc = 58ms
;need extra 13C external amp !!must use
dlo !! ENI setup
;p3 ca 18-20us 13C90 at dl0 10dB
;p13 26.3us 90 13C at dl1 17.8 db null on
carbonyl
;           when carrier set to 46 ppm
;p31 ca 60-65 us at dl2 28.1 db for cpd
decoupling
;p6 dp0 sinc1.0 180 13C' (294 us) with
dp0 40db
;     approx 20 kHz offset
;p4 15N 90 High Power
;set d7 to assure 0 linear phase correct
1.6*DW
;states with messerle water trim pulse
d26=p3-p1
d27=(p6/2)-p1
d28=(p6/2)-p4
d29=p4-p13
d30=p4-p1
d2=1.7m
d8=aq
p11=1m
d12=7m
d20=10u
d21=28.7m
p22=1.7m
d11=50m
d0=10u
;********** zero buffers
*******************
d12 st0
 ze
d12 st
 ze
d12 st0
1 d11
 d12
2 d12*4
3 d12 dlo dbhi
; st0            ;start at buffer 0
 d1 hl2          ;recycle delay
 ; p18 ph0        ;on resonance presat at
Hl2
 10u hl1         ;set to HP 1H
 p1 ph0          ;1H 90
 d2 dl0          ;1/4JXH 1.7ms
 2u
 (d26 p1*2 ph0) (p3*2 ph1):d      ;1H,X
180
 2u
 d2
 (p11 ph0)       ;Messerele trim pulse <
1ms
 2u
 (d26 p1 ph1) (p3 ph2):d      ;1H,X 90
 d0                    ;t1/2
 (p4):c8
 (d27 p1*2 ph0) (p6 ph0):dp0 (d28 p4*2
ph7):f3
;     180 1H, 180 shaped C'O, 15N 180
 (p4):c8
 d0                    ;t1/2
 d21                   ;T-t1/2 set sel X 180
power
 10u dl1
 (p13*2 ph6):d            ;apply selective X
180
 d21                   ;T-t1/2
 20u dl0               ;set X low power
 (p4):c8
 (p6 ph0):dp0             ;apply shaped at
pwr dp0
 (p4):c8
 10u dl0
 (d26 p1 ph0) (p3 ph3):d      ;90 1H, 90 X
 d2
 (d26 p1*2 ph4) (p3*2 ph5):d  ;180 1H,
180 X
 d7                    ;adjust for flat baseline
1.6*DW
 2u ph0
 p9:e ph0              ;set rec phase
 d2 dl2                ;open rec, wait
period d2
 10u adc cpds             ;start adc and cpd
 d8                    ;set to acq time
 do
 rcyc=1 ph31
;**************second
experiment************************
*
d12 st
10 d11
```

```
  d12
  d12*4
  d12 dlo dbhi
  d1 hl2          ;recycle delay
 ; p18 ph0          ;on resonance presat at
Hl2
  10u hl1          ;set to HP 1H
  p1 ph0           ;1H 90
  d2 dl0           ;1/4JXH 1.7ms
  2u
  (d26 p1*2 ph0) (p3*2 ph1):d      ;1H,X
180
  2u
  d2
  (p11 ph0)        ;Messerele trim pulse <
1ms
  2u
  (d26 p1 ph1) (p3 ph2):d      ;1H,X 90
  d0                    ;t1/2
  (d30 p1*2 ph0) (p4*2 ph7):f3 ;180 1H,
180 15N
  d0                    ;t1/2
  d21                   ;T-t1/2
  (p6 ph0):dp0          ;carbonyl pulse
  10u dl1
  (p13*2 ph6):d         ;apply selectiveX180
  d21                   ;T-t1/2
  20u dl0              ;set X power
  (p4):c8
  (p6 ph0):dp0         ;apply shaped at pwr
dp0
  (p4):c8
  10u dl0
  (d26 p1 ph0) (p3 ph3):d      ;90 1H, 90 X
  d2
  (d26 p1*2 ph4) (p3*2 ph5):d  ;180 1H,
180 X
  d7                    ;adjust for flat
baseline 1.6*DW
  2u ph0
  p9:e ph0                 ;set rec phase
  d2 dl2                ;open rec, wait period
d2
  10u adc cpds                ;start adc and cpd
  d8                    ;set to acq time
  do
  rcyc=10 ph31
  3u
;  lo to 1 times l2
;******* write the data
********************
```

```
  d11 do wr #0 if #0 st0 zd   ;write +
increment file
  d12 ip2                ;increment ph2 for
States
lo to 2 times 2
  d12 ip31               ;increment rec for
States TPPI
  d12 ip31
  d12 id0                ;increment d0 by in0
  d12 dd21               ;decrement d21 by
in21
lo to 3 times l3
exit
ph0=0
ph1=1 3
ph2=0
ph3=0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
ph4=0
ph5=0
ph6=(360) 353 353 173 173 83 83 263 263
;adjust for low power
ph7=0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
ph31=0 2 0 2 2 0 2 0
   2 0 2 0 0 2 0 2
;hl1 ecoupler high power
; p1 90 1H high power at hl1
; d1 relax delay
; d2 1/4JXH
; in0 dw/2 X dimension
; in21 = in0
; p18 presat if needed
; nbl = 2
;number of scans = 32*l2
;ns=32
;complex t1 points t1 dimension = l3
;td = 2* normal td
;td1 = 4*l3
; total time approx 17 hrs.
;13C carrier 43 ppm;sw 33ppm; in0
100.4us;284 expt
;1H HDO;8ppm;
```

P.S. #19 2D ^{13}C-{^{15}N} spin echo difference ct-HSQC (cthsqcjnc.jgp)

```
;2D 1H-13C hsqc with broad-band
decoupling of 13c for JNC coupling
;constants
;Vuister, Wang & Bax (1992) JACS 1993,
115, 5334-5335
;constant time 2T tuned to 2/Jcc = 58ms
;need extra 13C external amp !!must use
dlo !! ENI setup
;p3 ca 18-20us 13C90 at dl0 10dB
;p13 26.3us 90 13C at dl1 17.8 db null on
carbonyl
;           when carrier set to 43 ppm
;p31 ca 60-65 us at dl2 28.1 db for cpd
decoupling
;p6 dp0 sinc1.0 180 13C' ( 240 us ) with
dp0 40db
;     approx 20 kHz offset
;p4 15N 90 High Power
;set d7 to assure 0 linear phase correct
1.6*DW
;states with messerle water trim pulse
d26=p3-p1
d27=(p6/2)-p1
d28=(p6/2)-p4
d29=p4-p13
d2=1.7m
d8=aq
p11=1m
d12=7m
d20=10u
d21=28.6m
p22=1.7m
d11=50m
;*********** zero buffers
********************
d12 st0
 ze
d12 st
 ze
d12 st0
1 d11
 d12
2 d12*4
3 d12 dlo dbhi
; st0            ;start at buffer 0
 d1 hl2          ;recycle delay
 ; p18 ph0         ;on resonance presat at
Hl2
 10u hl1          ;set to HP 1H
 p1 ph0          ;1H 90
 d2 dl0          ;1/4JXH 1.7ms
 2u
 (d26 p1*2 ph0) (p3*2 ph1):d      ;1H,X
180
 2u
 d2
 (p11 ph0)        ;Messerele trim pulse <
1ms
 2u
 (d26 p1 ph1) (p3 ph2):d      ;1H,X 90
 d0                    ;t1/2
 (d27 p1*2 ph0) (p6 ph0):dp0 (d28 p4*2
ph7):f3
;     180 1H, 180 shaped C'O, 15N 180
 d0                    ;t1/2
 d21 dl1                ;T-t1/2 set sel X 180
power
 d29                   ;comp delay for 15N
180 expt 2
 (p13*2 ph6):d            ;apply selective X
180
 d29                   ;comp delay for 15N
180 expt 2
 d21                   ;T-t1/2
 10u dl0                ;set low power X
 (p6 ph0):dp0             ;apply shaped at
pwr dp0
 10u dl0
 (d26 p1 ph0) (p3 ph3):d      ;90 1H, 90 X
 d2
 (d26 p1*2 ph4) (p3*2 ph5):d  ;180 1H,
180 X
 d7                    ;adjust for flat baseline
1.6*DW
 2u ph0
 p9:e ph0                ;set rec phase
 d2 dl2                 ;open rec, wait
period d2
 10u adc cpds             ;start adc and cpd
 d8                    ;set to acq time
 do
 rcyc=1 ph31
;**************second
experiment************************
*
d12 st
10 d11
  d12
  d12*4
  d12 dlo dbhi
  d1 hl2          ;recycle delay
```

```
 ; p18 ph0          ;on resonance presat at
Hl2
  10u hl1          ;set to HP 1H
  p1 ph0           ;1H 90
  d2 dl0           ;1/4JXH 1.7ms
  2u
  (d26 p1*2 ph0) (p3*2 ph1):d      ;1H,X
180
  2u
  d2
  (p11 ph0)        ;Messerele trim pulse <
1ms
  2u
  (d26 p1 ph1) (p3 ph2):d      ;1H,X 90
  d0                    ;t1/2
  (d27 p1*2 ph0) (p6 ph0):dp0      ;180
1H, 180 shaped C'O
  d0                    ;t1/2
  d21 dl1                 ;T-t1/2 set sel X
180 power
  (d29 p13*2 ph6):d (p4*2 ph7):f3  ;apply
selective X 180 + 15N 180
  d21                   ;T-t1/2
  10u dl0                 ;set low power X
  (p6 ph0):dp0              ;apply shaped at
pwr dp0
  10u
  (d26 p1 ph0) (p3 ph3):d      ;90 1H, 90 X
  d2
  (d26 p1*2 ph4) (p3*2 ph5):d  ;180 1H,
180 X
  d7                    ;adjust for flat
baseline 1.6*DW
  2u ph0
  p9:e ph0                  ;set rec phase
  d2 dl2                ;open rec, wait period
d2
  10u adc cpds              ;start adc and cpd
  d8                    ;set to acq time
  do
  rcyc=10 ph31
  3u
  ;lo to 1 times l2
;******* write the data
*********************
 d11 do wr #0 if #0 st0 zd       ;write +
increment file
 d12 ip2                ;increment ph2 for
States
lo to 2 times 2
 d12 ip31                ;increment rec for
States TPPI
 d12 ip31
 d12 id0                        ;increment d0 by in0
 d12 dd21                        ;decrement d21 by
in21
lo to 3 times l3
exit
ph0=0
ph1=1 3
ph2=0
ph3=0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
ph4=0
ph5=0
ph6=(360) 353 353 173 173 83 83 263 263
;adjust for low power
ph7=0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
ph31=0 2 0 2 2 0 2 0
    2 0 2 0 0 2 0 2
;hl1 ecoupler high power
; p1 90 1H high power at hl1
; d1 relax delay
; d2 1/4JXH
; in0 dw/2 X dimension
; in21 = in0
; p18 presat if needed
; nbl = 2
;number of scans = 32*l2
;ns=32
;complex t1 points t1 dimension = l3
;td = 2* normal td
;td1 = 4*l3
; total time approx 17 hrs.
;13C carrier 43 ppm;sw 33ppm; in0
100.4us;284 expt
;1H HDO;8ppm;
```

P.S. #20 ct-HCCH-TOCSY (hcchdipct.jgp)

```
;constant time hcch-dipsi with t1-trick
;p1= proton 90 at hl1
;p18 = proton presaturation at hl2
;p11 messerle pulse h2o suppression
;f2 = carbon frequency at 43 ppm
;p3 = high power carbon 90 ca. 12us
;p13=180 with null at CO (21.4 90 degr)
;p4 = shaped pulse on carbonyl 180 degree
;dl0=high
;dl1=null at CO
;dl2=dipsi
;dl3=13c decouple
p5 = 1m          ;low power trim pulse at dl0
;p6 low power carbon 90 30-35 us for
dipsi/flopsy at dl0
p10=p6*.55555
;p31 low power carbon ca. 50us 90 for
waltz decoupling dl1
d6= p3*2 + d0*2
;d7 = de
d8 = aq
d10=4u
d20=1.6m
d21=1.6m
d11 =p1*2 + 13.6m + d10
p7 = 1.5m
p8 = p3+1.5u
p9 = p3-1.5u
p11=1m

1       ze
2       50m
;      11m
;3     33m
;4      22m
5       33m
6       10u hl2
7     d1 dlo
;       p18 ph0
        10u hl1 dl0
;*********************************
        p1 ph1          ;starting 90
     d20
        d0
        (p3*2 ph4):f2
        d0
        p1*2 ph3
        d21
        d6      ;compensation delay
        p11 ph1
;*********************************
        (p1 ph4) (p3 ph5):f2   ;first inept
        2u
        d10 dlo
        (p4 ph10):dp0          ;Carbonyl
180
        1.1m
        (p1*2 ph0)
        12.5m dl1   ;for 2/JCC use 25.2m
        (p13*2 ph6):f2
        2u
        d11 dl2  ;for 1/Jcc use 13.6m for
2/JCC 26.3m
        (p4 ph10):dp0          ;Carbonyl
180
        4u
        4u dl2
;*********************************
        (p5 ph12):f2    ;SL-pulse
;***  start DIPSIy  ***********
8       (p10*4.9 ph15):f2
        (p10*7.9 ph16):f2
        (p10*5.0 ph15):f2
        (p10*5.5 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.6 ph16):f2
        (p10*7.2 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.4 ph15):f2
        (p10*6.8 ph16):f2
        (p10*7.0 ph15):f2
        (p10*5.2 ph16):f2
        (p10*5.4 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.5 ph15):f2
        (p10*7.3 ph16):f2
        (p10*5.1 ph15):f2
        (p10*7.9 ph16):f2
        (p10*4.9 ph15):f2
        (p10*7.9 ph16):f2
        (p10*5.0 ph15):f2
        (p10*5.5 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.6 ph16):f2
        (p10*7.2 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.4 ph15):f2
        (p10*6.8 ph16):f2
        (p10*7.0 ph15):f2
        (p10*5.2 ph16):f2
        (p10*5.4 ph15):f2
        (p10*0.6 ph16):f2
```

```
        (p10*4.5 ph15):f2
        (p10*7.3 ph16):f2
        (p10*5.1 ph15):f2
        (p10*7.9 ph16):f2
;
        (p10*4.9 ph16):f2
        (p10*7.9 ph15):f2
        (p10*5.0 ph16):f2
        (p10*5.5 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.6 ph15):f2
        (p10*7.2 ph16):f2
        (p10*4.9 ph15):f2
        (p10*7.4 ph16):f2
        (p10*6.8 ph15):f2
        (p10*7.0 ph16):f2
        (p10*5.2 ph15):f2
        (p10*5.4 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.5 ph16):f2
        (p10*7.3 ph15):f2
        (p10*5.1 ph16):f2
        (p10*7.9 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.9 ph15):f2
        (p10*5.0 ph16):f2
        (p10*5.5 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.6 ph15):f2
        (p10*7.2 ph16):f2
        (p10*4.9 ph15):f2
        (p10*7.4 ph16):f2
        (p10*6.8 ph15):f2
        (p10*7.0 ph16):f2
        (p10*5.2 ph15):f2
        (p10*5.4 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.5 ph16):f2
        (p10*7.3 ph15):f2
        (p10*5.1 ph16):f2
        (p10*7.9 ph15):f2
        lo to 8 times l2
        (p5 ph12):f2    ;SL-pulse
;*******        end DIPSIy   *********
        2u
        1.1m dl0
        (p1*2 ph0) (p3*2 ph0):f2
        2u
        1.1m
        (p1 ph0) (p3 ph9):f2    ;inept
        1.5m
        (p1*2 ph0) (p3*2 ph30):f2
        d7
        p7:e
        3u adc
        (p8 ph0):f2
        (p9 ph11):f2
        2u
        2u dl3
        d8 cpds
        do
rcyc=2 ph31
50m wr #0 if #0 zd
;11m ip1
;lo to 3 times 2
;11m id0
;11m id20
;11m dd21
;11m ip31
;11m ip31
;lo to 4 times 64
11m rd0
11m ip5
11m ip6
lo to 5 times 2
11m id10
11m dd11
11m ip31
11m ip31
lo to 6 times l3
do
exit

ph0=0
ph1=0
ph3=0 0 0 0 2 2 2 2 1 1 1 1 3 3 3 3
ph30=0
ph4=1 3
ph5=0
ph6=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph9=0 0 2 2
ph10=0
ph11=0 0 2 2
ph12=0          ;should be adjusted
ph15=(360)90          ;should be adjusted
ph16=(360)270          ;should be adjusted
ph21=(6)0
ph22=(6)1
ph23=(6)2
ph24=(6)3
ph25=(6)4
ph26=(6)5
ph31=0 2 2 0 2 0 0 2
   2 0 0 2 0 2 2 0
```

P.S. #21 2D ct-HCCH-TOCSY [aromatics] (hcchdpctar.jgp)

```
;hcchdpctar.jgp
;derived from hcchdipct.abx
;constant time hcch-dipsi with t1-trick
;optimized for aromatic H-C pairs 11/4/93
;-jgp & ffd
;p1= proton 90 at hl1
;p18 = proton presaturation at hl2
;f2 = carbon frequency at 125 ppm
;p3 = high power carbon 90 ca. 12us
;dl0=dipsi
;dl1=13c decouple
p5 = 1m          ;low power trim pulse at dl0
;p6 low power carbon 90 30-35 us for
dipsi/flopsy at dl0
p10=p6*.55555
;p31 low power carbon ca. 50us 90 for
waltz decoupling dl1
d6= p3*2 + d0*2
;d7 = de
d8 = aq
d10=4u
;d20=1.3m
;d21=1.3m
d11 =p1*2 + 7.14m + d10
p7 = 1.3m
p8 = p3+1.5u
p9 = p3-1.5u

1       ze
2       55m
3       15m
4       10m
5       15m
6       10u hl2
        d1
7     1s
;       p18 ph0
        10u hl1 dhi      ;switch to hp 1H &
13C
;*********************************
        p1 ph1           ;starting 90
      1.3m
        d0
        (p3*2 ph4):f2
        d0
        p1*2 ph3
        1.3m
        d6               ;compensation delay
;*********************************
        (p1 ph4) (p3 ph5):f2   ;first inept
        2u
        d10
        1.3m
        (p1*2 ph0)
        5.84m
        (p3*2 ph6):f2
        2u
        d11 dlo dl0  ;switch to Dipsi 13C
power
        4u
        4u
;*********************************
        (p5 ph12):f2   ;SL-pulse
;*******       start DIPSIy   *********
8       (p10*4.9 ph15):f2
        (p10*7.9 ph16):f2
        (p10*5.0 ph15):f2
        (p10*5.5 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.6 ph16):f2
        (p10*7.2 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.4 ph15):f2
        (p10*6.8 ph16):f2
        (p10*7.0 ph15):f2
        (p10*5.2 ph16):f2
        (p10*5.4 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.5 ph15):f2
        (p10*7.3 ph16):f2
        (p10*5.1 ph15):f2
        (p10*7.9 ph16):f2
        (p10*4.9 ph15):f2
        (p10*7.9 ph16):f2
        (p10*5.0 ph15):f2
        (p10*5.5 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.6 ph16):f2
        (p10*7.2 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.4 ph15):f2
        (p10*6.8 ph16):f2
        (p10*7.0 ph15):f2
        (p10*5.2 ph16):f2
        (p10*5.4 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.5 ph15):f2
```

```
        (p10*7.3 ph16):f2
        (p10*5.1 ph15):f2
        (p10*7.9 ph16):f2
;
        (p10*4.9 ph16):f2
        (p10*7.9 ph15):f2
        (p10*5.0 ph16):f2
        (p10*5.5 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.6 ph15):f2
        (p10*7.2 ph16):f2
        (p10*4.9 ph15):f2
        (p10*7.4 ph16):f2
        (p10*6.8 ph15):f2
        (p10*7.0 ph16):f2
        (p10*5.2 ph15):f2
        (p10*5.4 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.5 ph16):f2
        (p10*7.3 ph15):f2
        (p10*5.1 ph16):f2
        (p10*7.9 ph15):f2
        (p10*4.9 ph16):f2
        (p10*7.9 ph15):f2
        (p10*5.0 ph16):f2
        (p10*5.5 ph15):f2
        (p10*0.6 ph16):f2
        (p10*4.6 ph15):f2
        (p10*7.2 ph16):f2
        (p10*4.9 ph15):f2
        (p10*7.4 ph16):f2
        (p10*6.8 ph15):f2
        (p10*7.0 ph16):f2
        (p10*5.2 ph15):f2
        (p10*5.4 ph16):f2
        (p10*0.6 ph15):f2
        (p10*4.5 ph16):f2
        (p10*7.3 ph15):f2
        (p10*5.1 ph16):f2
        (p10*7.9 ph15):f2
        lo to 8 times 11
        (p5 ph12):f2    ;SL-pulse
;*******        end DIPSIy    *********
        2u
        1.3m dhi                ;switch back
to hp 13C
        (p1*2 ph0) (p3*2 ph0):f2
        2u
        1.3m
        (p1 ph0) (p3 ph9):f2    ;inept
        1.3m
        (p1*2 ph0) (p3*2 ph30):f2
        d7
        p7:e
        3u adc
        (p8 ph0):f2
        (p9 ph11):f2
        2u
        2u dl1
        d8 cpds
        do
rcyc=2 ph31
50m wr #0 if #0 zd
;5m ip1
;lo to 3 times 2
;5m id0
;5m id20
;5m dd21
;5m ip31
;5m ip31
;lo to 4 times 64
;5m rd0
5m ip5
5m ip6
lo to 5 times 2
5m id10
5m dd11
5m ip31
5m ip31
lo to 6 times 56
do
exit

ph0=0
ph1=0
ph3=0 0 0 0 2 2 2 2 1 1 1 1 3 3 3 3
ph30=0
ph4=1 3
ph5=0
ph6=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph9=0 0 2 2
ph10=0
ph11=0 0 2 2
ph12=0          ;should be adjusted
ph15=(360)90            ;should be adjusted
ph16=(360)270           ;should be adjusted
ph21=(6)0
ph22=(6)1
ph23=(6)2
ph24=(6)3
ph25=(6)4
ph26=(6)5
ph31=0 2 2 0 2 0 0 2
   2 0 0 2 0 2 2 0
```

P.S. #22 2D 15N{1H} NOE HSQC (het15nonoe.jgp)

```
;reference pulse sequence
;For measurement of heteronuclear NOE
;REf. Nino Barbato, et al. Biochemistry
1992, 31 5269.
;Y channel for 15N
;States-TPPI
;p1 1H, p3 15N hi pwr, p31 15N low pwr
;set relax delay >3T1 15N and > 3T1 1H
;can get errors for fast exchange amides
because of exchange
d8=aq
d11=50m
d12=11m
p4=2.25m
p8=d0*2 + p1*2
p19=20m
d26=p3-p1
d9=p1*1.33
p22=10m
1 ze
2 d11
  d12
3 d12*3
4 d1 dbhi
;***********Presat ****************
10u hl2
p18
10u hl1
;*************Saturate 1H 120x
pulses*no pulse in ref******
10      d9
        20m
lo to 10 times l2
;******Purge to remove H2O *********
p22 ph0
1m
p22 ph0
10u
;****** Start on 15N ******
(p3 ph1):f3
2.75m
d0
(p1*2 ph0)
d0
(p3*2 ph2):f3
(p8):c8
2.75m
(d26 p1 ph3) (p3 ph4):f3
2.25m
(d26 p1*2 ph0) (p3*2 ph0):f3
3u
(3u ph7)
d7 dblo
p4:e
3u adc
d8 cpdb
dbo
rcyc=2 ph31
d11 wr #0 if #0 zd
d12 ip1
lo to 3 times 2
d12 ip31
d12 id0
d12 ip31
lo to 4 times l3
exit
ph0=0
ph1=0 2
ph2=0 0 0 0 0 0 0 0
   1 1 1 1 1 1 1 1
   2 2 2 2 2 2 2 2
   3 3 3 3 3 3 3 3
ph3=0 0 2 2
ph4=1 1 1 1
   3 3 3 3
ph7=0
ph31=0 2 2 0 2 0 0 2
   2 0 0 2 0 2 2 0
```

P.S. #23 2D 15N{1H} NOE HSQC (het15nnoe.jgp)

```
;For measurement of heteronuclear NOE
;REf. Nino Barbato, et al. Biochemistry
1992, 31 5269.
;Y channel for 15N
;States-TPPI
;p1 1H, p3 15N hi pwr, p31 15N low pwr
;set relax delay >3T1 15N and > 3T1 1H
;can get errors for fast exchange amides
because of exchange
d8=aq
d11=50m
d12=11m
p4=2.25m
p8=d0*2 + p1*2
p19=20m
d26=p3-p1
p22=10m
1 ze
2 d11
  d12
3 d12*3
4 d1 dbhi
;*************Presat
****************
10u hl2
p18 ph0
10u hl1
;*************Saturate 1H 120x
pulses****************
10 (p1*1.33 ph0)
20m
  lo to 10 times l2
;*******Purge pulses to remove H2O
*****
p22 ph0
1m
p22 ph0
10u
; *******Start on 15N ******
(p3 ph1):f3
2.75m
d0
(p1*2 ph0)
d0
(p3*2 ph2):f3
(p8):c8
2.75m
(d26 p1 ph3) (p3 ph4):f3
2.25m
(d26 p1*2 ph0) (p3*2 ph0):f3
3u
(3u ph7)
d7 dblo
p4:e
3u adc
d8 cpdb
dbo
rcyc=2 ph31
d11 wr #0 if #0 zd
d12 ip1
lo to 3 times 2
d12 ip31
d12 id0
d12 ip31
lo to 4 times l3
exit
ph0=0
ph1=0 2
ph2=0 0 0 0 0 0 0 0
   1 1 1 1 1 1 1 1
   2 2 2 2 2 2 2 2
   3 3 3 3 3 3 3 3
ph3=0 0 2 2
ph4=1 1 1 1
   3 3 3 3
ph7=0
ph31=0 2 2 0 2 0 0 2
   2 0 0 2 0 2 2 0
```

P.S. #24 HSQC (hsqcYdecC.jgp)

```
;bodenruben hetcor - single quantum in t1
;ref
;states-tppi in t1
;X nucleus on X channel
; 180 13C pulses to decouple 13C during t1
d8=aq
d11=50m
d12=5m
d6=3u
d25=p3-p1
p13=68u
p24=d4

1 ze
2 d12 o2
  d12
  d12
3 d1
  10u hl2
  ;4 (p7 ph15 d5 p7 ph16 d5 p7 ph17 d5 p7
ph18 d5)  ;Dante presat
  ;  lo to 3 times 1800
  p18 ph0
  10u hl1
  p1 ph8                    ;1H 90
  d4 dbhi dlo                    ;1/4JXH X
dec on hi pwr
  (d25 p1*2 ph8) (p3*2 ph1):f3     ;1H, X
180 hi pwr
  d4
  p22 ph8                   ;messerele trim
H20 pulse
  (d25 p1 ph9) (p3 ph11):f3       ;1H, X 90
Hi pwr
  d0                        ;t1/2
  p13:d ph1                      ;C'
decoupling
  ;(p1 ph0 d6 p1*2 ph1 d6 p1 ph0)
;composite 180 1H
  p1*2 ph0 o2                    ;1H decoupling
and 13C freq shift
  p13:d ph1                 ;C(alpha)
decoupling
  d0                        ;t1/2
  (d25 p1 ph10) (p3 ph12):f3      ;1H, X 90
H pwr
  d4
  (d25 p1*2 ph8) (p3*2 ph1):f3
  p24:e dblo                ;open rec, wait
d4, set dec pow lev
  d7                        ;0.6*DW or
1.6*DW to flatten baseline
  d6 adc
  d8 cpdb
  rcyc=2 ph31
d11 dbo wr #0 if #0 zd
d12 ip11
lo to 2 times 2
d12 ip31
d12 ip31
d12 id0
lo to 3 times l3
exit
ph0=1
ph1=0
ph2=2
ph3=1
ph4=1
ph5=0
ph10=0 0 0 0 2 2 2 2
ph15=0
ph16=1
ph17=2
ph18=3
ph8=0
ph9=1 3
ph11=0 0 2 2
ph12=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph31=0 2 2 0 2 0 0 2 2 0 0 2 0 2 2 0
;p1 1H 90  p3 X 90  p31 cpd 90 at dbl0
(dblo)
;p13 39us semi-soft pulse to invert C' with
null at Ca
;     or Ca but not C'
```

P.S. #25 2D LRCC

(lrcc.jgp)

```
;lrcc.jgp
;Measure long-range carbon-carbon
couplings
;ref: Bax, Max,Zax 1992, JACS 114,6923-
6925
;ENI setup
;need extra 13C ENI external amp must use
dlo for 13C
;p3 ca 18-20us 13C 90 at dl0
;p13 26.3us 90 13C at dl1 null on carbonyl
;          when carrier set to 43 ppm
;p31 ca 60-65 us at dl2 for cpd decoupling
;p6 dp0 sinc1.0 180 13C' ( 240 us ) with
dp0 40db
;     approx 20 kHz offset
;set d7 to assure 0 linear phase correct
1.6*DW
;states-tppi with messerle water trim pulse
d26=p3-p1
d2=1.7m
d8=aq
p11=1m
d12=11m
d11=50m
;d21=100u ; comment in for test of 1st
block
d21=14.7m
p9=2u

1 ze
 d12 dlo
2 d11
d12*4
3 d12*2
 d12*4
3 d12*2
4 d1 hl2          ;recycle delay
; p18 ph0          ;on resonance presat at
Hl2
 10u hl1          ;set to HP 1H
 p1 ph0           ;1H 90
 d2 dl0           ;1/4JXH 1.7ms
 (d26 p1*2 ph0) (p3*2 ph1):d      ;1H,X
180
 d2
 (p11 ph0)        ;Messerele trim pulse <
1ms
 (d26 p1 ph1) (p3 ph7):d    ;1H,X 90
 4u
 (p6 ph0):dp0
 ;** long range to antiphase **
 d21 dl1                ;constant time T
14.7ms
 (p13*2 ph3):f2           ;med power 13C
180
 4u
 (p6 ph0):dp0
 d21 dl0                ;constant time T
 ;** COSY to other carbons **
 (p3 ph4):d
 d0                ;t1/2
 p1*2 ph0
 d0                ;t1/2
 (p3 ph5):d
 ;** get back **
 d21 dl1                ;T
 (p6 ph0):dp0             ;apply shaped C'
at pwr dp0
 4u dl1
 (p13*2 ph6):d            ;med power 13C
180
 d21 dl0                ;T
 (p6 ph0):dp0             ;180C' for
compensation
 4u dl0
 (d26 p1 ph0) (p3 ph2):d     ;90 1H, 90 X
 d2
 (d26 p1*2 ph0) (p3*2 ph8):d  ;180 1H,
180 X
 d7                ;adjust for flat baseline
1.6*DW
 2u ph0
 p9:e ph0                ;set rec phase
 d2 dl2            ;open rec, wait period
d2
 10u adc cpds             ;start adc and cpd
 d8                ;set to acq time
 do
 rcyc=2 ph31
 d11 do wr #0 if #0 zd       ;write +
increment file
 d12 ip2
 d12 ip5
 d12 ip6
 d12 ip8
lo to 3 times 2
 d12 ip31                ;increment rec for
States TPPI
 d12 ip31 id0
```

```
lo to 4 times l3
exit
ph0=0
ph1=1 3
ph2=1 1 3 3
ph3=0 0 2 2 1 1 3 3
ph4=0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
ph5=0 0 0 0 0 0 0 0
   0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
   2 2 2 2 2 2 2 2
ph6=0                ;adjust for low power
ph7=1
ph8=0
ph31=0 2 2 0 2 0 0 2
;hl1 ecoupler high power
; p1 90 1H high power at hl1
; d1 relax delay
; d2 1/4JXH
; in0 dw/2 X dimension
;in0
; ns = 16*n
; ds = 16
```

P.S. #26 2D LRCH (lrch.jgp)

```
; 3D lrch - fully labeled samples
;Coupling constants between methyl
carbons and 2+3 bond protons
;ref Vuister + Bax JMR Series B 102,228-
231 (1993).
;standard setup
;HP + LP 13C
;HP 1H
;in18=in22=in23=in24=t2/4=36.1us
;in10=t2/2=72.2us
;in0=t1/2
p21=p1*2-(p3*0.636)+(p1*0.636)
d0=3u
d10=(in10-p1*1.27-p3*2)*0.5
d18=3m
d21=23.6m
d22=3m
d23=25.6m
d24=4m
d26=p3-p1
p22=1.7m
d8=aq
d11=50m
d12=11m
1 ze
2 d11
; d12
;3 d12*4
;4 d12
 d12
5 d12*7
6 d1 dhi
  10u hl1
;**** Begin saturation of 1H *****
8  p1*1.5 ph0
  10m
  lo to 8 times 75 ;(for 1sec sat of 1H)
  10u hl1
;**** Start sequence  ****
  (p3 ph1):f2
  p21:c8
  d0
  (p3*2 ph2):f2
  d18                ;zeta-t2/4
  p1*2 ph0
  d21                ;23.6m-t1/2
  d22                ;3m -t2/4
  p1 ph3
  d10                ;t2/2
  (p3*2 ph4):f2
  d10
  p1 ph4
  d23                ;25.6m-t2/4
  p1*2 ph0
  d24                ;e-t2/4 = 4m - t2/4
  (d26 p1 ph0) (p3 ph5):f2
  1.7m
  (d26 p1*2 ph0) (p3*2 ph0):f2
  d7
  (2u ph0)
  p22:e dlo dl0
  10u adc cpd
  d8
  do
  rcyc=2 ph31
  d11 wr #0 if #0 zd
;  d12 ip1
;  lo to 3 times 2
;  d12 id0
;  d12 dd21
;  d12 ip31
;  d12 ip31
;  lo to 4 times 32
;  d12 rd0
  d12 ip3
  lo to 5 times 2
  d12 dd18
  d12 dd22
  d12 id10
  d12 dd23
  d12 dd24
  d12 ip31
  d12 ip31
  lo to 6 times 83
  exit
  ph0=0
  ph1=0
  ph2=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
  ph3=0 2
  ph4=0 0 2 2
  ph5=1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
     3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
  ph31=0 2 2 0 2 0 0 2 0 2 2 0 2 0 0 2
        2 0 0 2 0 2 2 0 2 0 0 2 0 2 2 0
```

P.S. #27 2D LRCH (lrchref.jgp)

```
; 3D lrch - fully labeled samples
;Coupling constants between methyl
carbons and 2+3 bond protons
;ref Vuister + Bax JMR Series B 102,228-
231 (1993).
;standard setup
;HP + LP 13C
;HP 1H
;in18=in22=in23=in24=t2/4=36.1us
;in10=t2/2=72.2us
;in0=t1/2
p21=p1*2-(p3*0.636)+(p1*0.636)
d0=3u
d10=(in10-p1*1.27-p3*2)*0.5
d18=3m
d21=23.6m
d22=3m
d23=25.6m
d24=4m
d26=p3-p1
p22=1.7m
d8=aq
d11=50m
d12=11m
1 ze
2 d11
 d12
3 d12*4
4 d1 dhi
  10u hl1
;**** Begin saturation of 1H *****
8  p1*1.5 ph0
  10m
  lo to 8 times 75 ;(for 1sec sat of 1H)
  10u hl1
;**** Start sequence  ****
  (p3 ph1):f2
  p21:c8
  d0
  (p3*2 ph2):f2
  d18                ;zeta-t2/4
  p1*2 ph0
  d21                ;23.6m-t1/2
  d22                ;3m -t2/4
  p1 ph3
  d10                ;t2/2
  (p3*2 ph4):f2
  d10
  p1 ph4
  d23                ;25.6m-t2/4
  p1*2 ph0
  d24                ;e-t2/4 = 4m - t2/4
  (d26 p1 ph0) (p3 ph5):f2
  1.7m
  (d26 p1*2 ph0) (p3*2 ph0):f2
  d7
  (2u ph0)
  p22:e dlo dl0
  10u adc cpd
  d8
  do
  rcyc=2 ph31
  d11 wr #0 if #0 zd
  d12 ip1
  lo to 3 times 2
  d12 id0
  d12 dd21
  d12 ip31
  d12 ip31
  lo to 4 times 64
;  d12 rd0
;  d12 ip3
;  lo to 5 times 2
;  d12 dd18
;  d12 dd22
;  d12 id10
;  d12 dd23
;  d12 dd24
;  d12 ip31
;  d12 ip31
;  lo to 6 times 50
  exit
  ph0=0
  ph1=0
  ph2=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
  ph3=0 2
  ph4=0 0 2 2
  ph5=1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
     3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
  ph31=0 0 0 0 2 2 2 2 0 0 0 0 2 2 2 2
        2 2 2 2 0 0 0 0 2 2 2 2 0 0 0 0
```

P.S. #28 3D HNHB (hnhb.jgp)

```
;hnhb.jgp
;for correlation of 15N with hb
;Use with hn(co)hb to get hb
stereoassignments
;and chi-1 angles in proteins
;Ref Archer et al., (1991) JMR 95,636-641
;STATES-Tppi with presat
;
d26=p3-p1
d2=2.25m
p22=2m
d11=50m
d12=11m
d3=38m
1 ze
2 d11 dbo
3 d12*2
4 d12*3
5 d12*2
6 d1 hl2
 p18 ph0   ;presat
 d15 hl1
 ; *** Start Inept to 15N ***
 p1 ph0
 d2 dbhi
 (d26 p1*2 ph0) (p3*2 ph1):f3
 d2
 p22 ph0   ;messerele spin H2o
suppression
 2u
 (p1 ph1) (p3 ph2):f3
 ;***Evolve long range N-Hb via HMQC
+ Hb t1***
 d3
 (p1*2):c8  ;compensation pulse
 p1 ph6
 d0
 (p3*2 ph8):f3
 d0
 p1 ph7
 d3
 ;*** Evolve 15N chem shift ***
 d10
 p1*2 ph0
 d10
 ;*** Reverse inept to 1H ***
 (d26 p1 ph0) (p3 ph3):f3
 d2
 (d26 p1*2 ph4) (p3*2 ph5):f3
 2u
 d2 dblo dbl0
 go=2 ph31 cpdb
 d11 do wr #0 if #0 ip6 zd
lo to 3 times 2
 d12 ip31
 d12 ip31 id0
lo to 4 times 256 ;States-TPPI Hb
; d12 ip3 rd0
; d12 ip3
; d12 ip3
;lo to 5 times 2  ;increment 3* because after
evolution
; d12 ip31
; d12 ip31 id10
;lo to 6 times 30 ;States-TPPI 15N
exit
ph0=0
ph1=1 3
ph2=0
ph3=0
ph4=0
ph5=0
ph6=0 0 2 2
ph7=0
ph8=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph31=0 2 2 0 2 0 0 2
ph20=0

;p1 1H 90
;p3 15N 90 Hipower
;p31 15N low power
;d0 hb  t1
;d10 15N t2
;d2=1/4JNH
;d3=38ms
```

P.S. #29 3D ^{15}N-HOHAHA-HMQC (hohaha15n.jgp)

```
; hohaha15n.jgp
;p1=proton 90
;p3=nitrogen 90 at dbhi
p4=2.25m
;p5 1H 90 at hl2 approx 15 dB
;p5=p1 proton at hl3
p22=2.0m; Messerle pulse
p9=700u        ;SL-pulse
p10=p5*320.0/90.0
p11=p5*410.0/90.0
p12=p5*290.0/90.0
p13=p5*285.0/90.0
p14=p5*30.0/90.0
p15=p5*245.0/90.0
p16=p5*375.0/90.0
p17=p5*265.0/90.0
p18=p5*370.0/90.0
p19=p5*60.0/90.0
;p21=presat pulse = ca. 1s
;p31=nitrogen 90 at dbl0 ca. 150us

;d0=25u -p1/3.14159
d1=1.0m
;d3 = half mixing time
d4=2.25m
;d7 ca 100u = 1.6*dw
d8=aq
d11=50m
d12=11m
d26=p3-p1
;in0= proton inc = 1/sw
;in10=nitrogen inc = 1/2/sw
;hl1 = 3dB
;hl2 = 70dB
;hl3 = 15db
;t=f1 = hydrogen frequency on water
;f3 = nitrogen frequency
        ze
        d11
2       d11
      d12*2
3     d12*3
4     d12*2
5     d12*3
;*******       presaturation
        *******
6       d1 hl2 dbhi
        p21 ph0
        5u
        20u hl1
;*******       start 90-degree *******
        p1 ph2
;*******       proton evolution  *******
        d0      ;t1
;*******       SL pulse        *******
;       p9 ph1
        2u
        2u hl3
;*******       dipsi-2 mixing  *******
7 (p10 ph20 p11 ph19 p12 ph20 p13 ph19
p14 ph20 p15 ph19 p16 ph20)
 (p17 ph19 p18 ph20)
 (p10 ph19 p11 ph20 p12 ph19 p13 ph20
p14 ph19 p15 ph20 p16 ph19)
 (p17 ph20 p18 ph19)
 (p10 ph19 p11 ph20 p12 ph19 p13 ph20
p14 ph19 p15 ph20 p16 ph19)
 (p17 ph20 p18 ph19)
 (p10 ph20 p11 ph19 p12 ph20 p13 ph19
p14 ph20 p15 ph19 p16 ph20)
 (p17 ph19 p18 ph20)
;;      (p19 ph21)
        lo to 7 times l2
;;*******      total duration 115.11 * p5 *
loop l2
        2u
        2u hl1
;*******       SL pulse        *******
;       (p9 ph1)
        2u
;*******       read pulse      *******
        (p1 ph0)
;       10u hl2
;       p20 ph0
;       10u hl1
        p8:h
        d3      ;half mixing time
        (p1 ph0)
        d4
        (d26 p1*2 ph4) (p3*2 ph10):f3
        d4
;*******       water trim pulse  *******
        (p22 ph0)
        2u
        (p1 ph1) (p3 ph11):f3
;*******       N15 evolution  *******
        d10     ;t2/2
        (p1*2 ph0)
        d10     ;t2/2
        (p1 ph0) (p3 ph10):f3
```

```
        d4
        (d26 p1*2 ph3) (p3*2 ph10):f3
        2u
        (2u ph0)
        d7 dblo
        p4:e
        3u adc
        d8 cpdb
dbo
rcyc=2 ph31
d11 wr #0 if #0 zd
d12 ip2
d12 ip2
lo to 3 times 2
d12 id0
d12 ip31
d12 ip31
lo to 4 times 128
d12 ip11
d12 rd0
lo to 5 times 2
d12 id10
d12 ip31
d12 ip31
lo to 6 times 128
dbo
exit

ph0=0
ph1=1
ph2=(8) 1 5
ph3=0 0 1 1 2 2 3 3
ph4=0 0 2 2 1 1 3 3
ph10=0
ph11=0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
ph19=(3600)0
ph20=(3600)1800
ph21=(3600)900
ph31=0 2 2 0 2 0 0 2
   2 0 0 2 0 2 2 0
```

P.S. #30 3D ^{13}C-ROESY-HMQC (roesyhmqc3d.jgp)

```
;roesyhmqc3d.ffd
;3D-13C-separated ROESY-HMQC
;3D sequence with
;  homonuclear correlation via dipolar
coupling during a cw
;  spinlock
;  dipolar coupling may be due to noe or
chemical exchange
;  but these will have opposite sign (noe
negative wrt diagonal)
;  H-1/X correlation via heteronuclear zero
and double quantum
;     coherence
;phase sensitive using States-TPPI
;with decoupling during acquisition
;with presaturation during relaxation delay
;JBNMR , 1 (1991) 13-22
; Clore, Bax, and Gronenborn
;
d2=3m
d11=50m
d12=11m
p4=3m
d8=aq

1 d11 ze
2 d11
 d12
3 d12*3
4 d12*2
5 d12*3
6 d1 o1 hl2 dhi
 p18 ph29      ; presat
 10u o1
 10u hl1
 p1 ph1         ; 1H 90
 d0
        ; t1/2
 (p3 ph4):f2    ; 13C 90x
 (p3*2 ph5):f2          ; 13C 180y (13C cs
refocus)
 (p3 ph4):f2    ; 13C 90x
 d0
        ; t1/2
 10u hl3
 p21 ph2        ; cw
spinlock=22ms/8.33kHz for Il-4b
 d2 hl1         ; 1/2J=3 ms/switch to
4.77ppm
 (p3 ph6):f2    ; 13C 90
 d10
        ; t2/2
 p1*2 ph3       ; 1H 180 (1H cs refocus)
 d10
        ; t2/2
 (p3 ph8):f2    ; 13C 90
 2u
 (2u ph0)
 d7
 p4:e
 3u adc
 (p3 ph9):f2    ; 13C 90x
 (p3 ph7):f2    ; 13C 90 (to reduce mod
sidebands)
 2u
 2u dlo dl0                          ;
switch ti lp for GARP dec
 d8 cpds                             ; acq
+ cGARP=3.85kHz
 do
 rcyc=2 ph31
 d11 wr #0 if #0 zd
 d12 ip1
 lo to 3 times 2
 d12 id0
 d12 ip31
 d12 ip31
 lo to 4 times 124              ; t1
incr=190us  f1=7ppm
 d12 ip6
 d12 rd0
 lo to 5 times 2
 d12 id10
 d12 ip31
 d12 ip31
 lo to 6 times 28               ; t2
incr=320us  f2=43ppm
 do
exit

ph0=0
ph1=0 0 0 0 1 1 1 1
ph2=1 1 1 1 2 2 2 2 3 3 3 3 0 0 0 0
ph3=0 0 1 1 3 3 2 2
ph4=0 0 0 0 2 2 2 2
ph5=1 1 1 1 3 3 3 3
ph6=0 2
ph7=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph8=0
```

```
ph9=0
ph29=0
ph31=0 2 2 0 1 3 3 1

;hl1: ecoupler high power level
;hl2: ecoupler power level for presaturation
;hl3: ecoupler power level for spinlock
;dl0: power level for dlo mode
;p1 :  90 degree transmitter high power
pulse
;p3 :  90 degree decoupler high power pulse
;p4 : refocus delay =d4
;p18: presaturation during relaxation delay
;p21: spinlock pulse
;p31: 90 degree pulse for slave timer (cpd-
sequence)
;d0 : incremented delay (F1 in 3D)
[3 usec]
;d2 : 1/((2J)XH)                          [3
msec]
;d10 : Incremented delay (F2 in 3D)
[3 usec]
;d11: delay for disk I/O
[30 msec]
;d12: delay for power switching
[20 usec]
;d13: short delay (e.g. to compensate delay
line)   [3 usec]
;in0: 1/(4 * SW(H)) = (1/2) DW(H)
;nd0: 2
;in5: 1/(4 * SW(X)) = (1/2) DW(X)
;nd10: 2
;NS: 4 * n
;DS: 32
;l0: number of experiments in F1
;l10: number of experiments in F2
;MC2: TPPI in F1
;MC2: TPPI in F2
;cpd: cpd-decoupling according to sequence
defined by cpdprg=garp
```

Appendix D Software for analysis

Table of software code

C Software for assignments analysis 174

chhnassign.c 174

chhnsortpippn.c 179

filterpippn.c 181

filter.c 183

hbpairs.c 184

hcchassignpipp.c 187

hcchassign.c 192

C Software for Analysis of Structures and Contraint Data 199

dist.c 199

hbond.c 201

hbondrev2.c 204

hbondrev2.com 208

energyrmsd.c 216

standev.c 219

noeclass.c 220

noeclass_writeall.c 221

ptnoe.c 224

ptbkrmsd.c 226

ptnonhrmsd.c 227

phicalc_s.c 230

psicalc_s.c 236

chi1_s.c 241

chi2_s.c 247

X-PLOR macros for structure analysis 254

ave.com 254

ave.inp 254

avecycle.com 258

C Software for assignments analysis

chhnassign.c

```
/* chhnassign.c */
/* Program to analyze peak picks from 4D chhn expt */
/* to compile use: cc chhnassign.c -lm */
/*The -lm is for inclusion of the math.h file */
/* To be used for 13C - 15N 4D NOESY only */

/* For a description of how this program works */
/* See the header of the source program: */
/* hcchassign.c */


/* ________ FILES __________ */
/* Input file  : filterresultpippn.d (filtered peak pick list) */
/* Input file  : assign.d (assignments file 1H/13C and 1H/15N) */
/* Input file  : xyz.d (coordinates for structure from previous */
/*                 round of refinement) */
/* Output file : chhnresult.d (output of new NOE assignments) */

#include <math.h>
#include <stdio.h>
#include <string.h>

main()
{
FILE *fp, *fp1, *fx2, *fp3;
int i, resnum[3000], atomnumc, resnumc[3000], peakno[5000], posneg[5000];
int ptr2, ptr3, ptr4;
char atname[3000][10], resname[3000][5], atomc[10], atnamec[3000][10];
char resnamec[3000][5],c212c[10];
float xc[3000], yc[3000], zc[3000], c1, bfc, h[3000], c[3000], h1[5000];
float wid1, h2[5000], wid2, c3[5000], wid3, c4[5000], wid4, cc;
float x[3000], y[3000], z[3000];
int xyzc, peaklistc, assignc, j;
int k, l, m, n, m1resnum[100], m2resnum[100];
char m1resname[100][5], m2resname[100][5], m1atname[100][10],m2atname[100][10];
float m1h1[100], m1c4[100], m2h2[100], m2c3[100], m1x[100], m1y[100];
float m1z[100], m2x[100], m2y[100], m2z[100];
float diffh1, diffc3, diffh2, diffc4;
double a, b, cd, d, dd;
float h1cut, h2cut, ccut, cfold, cfold3, cfold4, h1dev, c4dev, h2dev, c3dev;
int t, assign[5000],zindex[3000];
float c4pl[5000],c3pl[5000],dcut;
float ncut,nfold1,intensity[3000];
int aindex[5000];
/************** added correction of 0.08 to h2 (13C) chem shifts ***/

/* Define 1H and 13C chemical shift ranges to be tested */
```

```
/*  when assigning peaks from the peak list */
/* hcut for 1H and ccut for 13C */

printf(" input files assign.d , xyz.d, filterresultpippn.d \n");
printf(" output file  chhnresult.d \n");
h1cut=0.12;  /* cutoff for d1 dimension 1H - with c4 */
h2cut=0.15;  /* cutoff for d2 dimension 1H - with c3 */
ncut=0.5;
ccut=0.6;
cfold=23.8;
cfold3=3*23.8; /* for 3 folded aromatics */
cfold4=4*23.8; /* for 4 folded aromatics */
nfold1=16.88;
h1dev=0.0;   /*deviation h1[j]-h1dev=h[j] (h1[j] is peak picked data)*/
h2dev=0.08;   /* h[j] is chem shift assignment data */
c3dev=0.0;
c4dev=0.0;
dcut=15.0;  /* distance cutoff for printing possible proton pairs */
/* Open file to write results */
   if ((fp3=fopen("chhnresult.d", "w")) ==NULL) {
     printf("Error opening chhnresult.d");
     exit(1); }


/* Open data bases and read to arrays */
/* Open assignments + read data as assign.d */
  if ((fp=fopen("assign.d", "r")) ==NULL) {
     printf("Error opening assign.d/n");
     exit(1); }
printf("Opened chemical shift list assign.d");
printf(" Reading in the file \n");
/* resnum = residue number; resname= residue name; assign =assignment;*/
/* h = 1H chemical shift; c = carbon chemical shift */
   i=-1;
   assignc=-1;
   while(!feof(fp)) {
   i=i+1;
   assignc=assignc+1;
   fscanf(fp, "%i%s%s%f%f", &resnum[i], &resname[i], &atname[i], &h[i], &c[i]);
/*   printf("%i %s %s %2.2f %3.2f\n", resnum[i], resname[i], atname[i],h[i],c[i]);*/
       }
   fclose(fp);

/* Read in peak pick list */

   if ((fp=fopen("filterresultpippn.d", "r")) == NULL) {
     printf("Error opening peaklist.PCK\n");
     exit(1); }
printf(" Opened peak pick list peaklist.PCK - reading in the file \n");
   i=-1;
   peaklistc=-1;
   while(!feof(fp)) {
```

```
  i=i+1;
  peaklistc=peaklistc+1;
fscanf(fp,"%d%d%d%d%f%f%f%f%e", &peakno[i], &assign[i], &zindex[i],
&aindex[i],&h1[i], &h2[i], &c3[i], &c4[i], &intensity[i]);
/*  Calculate plane number for each c3 and c4    */
c4pl[i]=17-((c4[i]-117.8)/0.5275);
c3pl[i]=17-((c3[i]-43.0)/0.74);

printf("%d %2.2f %2.2f %2.2f %2.2f", peakno[i], h1[i], h2[i], c3[i], c4[i]);
printf(" %3.2e c4 plane %2.1f  c3 plane %2.1f\n\n", intensity[i], c4pl[i], c3pl[i]);
}
 fclose(fp);

/*  Read in xyz coordinates for a structure */

  if ((fp1=fopen("xyz.d", "r")) == NULL) {
   printf("Error opening xyz.d\n");
   exit(1); }
printf("Opened xyz coordinate file xyz.d - reading coordinates \n");
  i=-1;
  xyzc=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc=xyzc+1;
   fscanf(fp, "%s%d", &atomc, &atomnumc);
   fscanf(fp, "%s%s%d%f%f%f", &atnamec[i],
&resnamec[i],&resnumc[i],&xc[i],&yc[i],&zc[i]);
   fscanf(fp, "%f%f%s", &c1, &bfc, &c212c);
   printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[i], resnamec[i], resnumc[i], xc[i],
yc[i], zc[i]);
}
   fclose(fp);

/*  **  Match chemical shifts with xyz coordinates  ** */

printf("Matching assignments with coordinates\n");
 for(i=0; i<assignc; i++) {
   for(j=0; j<xyzc; j++) {
    if(strcmp(atname[i],atnamec[j])==0 && strcmp(resname[i],resnamec[j])==0 &&
        resnum[i]==resnumc[j]) {
                              x[i]=xc[j];
                              y[i]=yc[j];
                              z[i]=zc[j];
      printf("%s %s %d assigned to", resnamec[j], atnamec[j], resnumc[j]);
      printf(" %s %s %d", resname[i], atname[i], resnum[i]);
      printf(" xyz coordinates %2.3f %2.3f %2.3f\n", x[i], y[i], z[i]);
      if(x[i]==0.0 || y[i]==0.0 || z[i]==0.0) printf("Error -  no match");
}
}
}
/*  Match peak list to assignment table  */
```

```
  for(i=0; i<peaklistc ; i++) {
k=-1;
l=-1;

/* Check to see if peak in peak list is an aromatic */
/* Use a cutoff : if 1H shift is above 6.0 ppm */

/*if(h1[i]>6.0) c4[i]=c4[i]+(4*23.8);*/
/*if(h2[i]>6.0) c3[i]=c3[i]+(4*23.8);*/

    for(j=0; j<assignc; j++) {

if(fabs(h1[i]-h[j]-h1dev)<h1cut && (strcmp(atname[j],"HN")==0 ||
strcmp(atname[j],"HE1")==0 || strcmp(atname[j],"HE")==0 ||
strcmp(atname[j],"HD21")==0 || strcmp(atname[j],"HE21")==0) && (fabs(c4[i]-c[j]-
c4dev)<ccut || fabs(c4[i]+nfold1-c4dev-c[j])<ccut || fabs(c4[i]+(nfold1*2)-c4dev-c[j])<ccut
|| fabs(c4[i]-nfold1-c4dev-c[j])<ccut)) {
          k=k+1;
          strcpy(m1atname[k],atname[j]);
          strcpy(m1resname[k],resname[j]);
          m1resnum[k]=resnum[j];
          m1h1[k]=h[j];
          m1c4[k]=c[j];
          m1x[k]=x[j];
          m1y[k]=y[j];
          m1z[k]=z[j];
          }
if(fabs(h2[i]-h[j]-h2dev)<h2cut && (fabs(c3[i]-c[j]-c3dev)<ccut || fabs(c3[i]+cfold-c[j]-
c3dev)<ccut || fabs(c3[i]-cfold-c[j]-c3dev)<ccut || fabs(c3[i]+cfold3-c[j]-c3dev)<ccut ||
fabs(c3[i]+cfold4-c[j]-c3dev)<ccut)) {
          l=l+1;
          strcpy(m2atname[l],atname[j]);
          strcpy(m2resname[l],resname[j]);
          m2resnum[l]=resnum[j];
          m2h2[l]=h[j];
          m2c3[l]=c[j];
          m2x[l]=x[j];
          m2y[l]=y[j];
          m2z[l]=z[j];
          }
/*Use info on parameter posneg to filter some choices - */
/*if(h1[i]<6.0 && h2[i]<6.0) {

if(posneg[i]==2 && (fabs(c4[i]-c[j])<ccut && fabs(c3[i]-c[j])<ccut)) continue;
if(posneg[i]==0 && (fabs(c4[i]-c[j])<ccut || fabs(c3[i]-c[j])<ccut)) continue;}*/


    }
  /* Write out matches and calculate distance for each */
h1[i]=h1[i]-h1dev;
h2[i]=h2[i]-h2dev;
c3[i]=c3[i]-c3dev;
```

```
c4[i]=c4[i]-c4dev;
fprintf(fp3,"\n1     %d    %2.2f %2.2f ",peakno[i],h1[i],c4[i]);
fprintf(fp3, " %3.2e   %2.2f %2.2f",intensity[i],h2[i],c3[i]);
fprintf(fp3, "  c4 plane  %2.1f   c3plane  %2.1f\n\n", c4pl[i], c3pl[i]);

if(k==-1 || l==-1) { fprintf(fp3,"3          **No pairs found**\n");
                }

if(k>-1 && l>-1) {
        for(m=0; m<=k; m++) {


     for(n=0; n<=l; n++) {
        a=(m1x[m]-m2x[n])*(m1x[m]-m2x[n]);
        b=(m1y[m]-m2y[n])*(m1y[m]-m2y[n]);
        cd=(m1z[m]-m2z[n])*(m1z[m]-m2z[n]);
        dd=a+b+cd;

        if(dd<=0.0) d=0.0;
        else d=sqrtf(a+b+cd);
        if(m1x[m]==0.0 || m2x[n]==0.0) d=0.0;
        /*if(d >= dcut) continue;*/

        diffh1=h1[i]-m1h1[m];
        diffh2=h2[i]-m2h2[n];
        diffc3=c3[i]-m2c3[n];
        diffc4=c4[i]-m1c4[m];

fprintf(fp3,"2 %s %d %s",m1resname[m],m1resnum[m],m1atname[m]);
fprintf(fp3," %2.2f %2.2f",m1h1[m],m1c4[m]);

fprintf(fp3,"  %s %d %s ",m2resname[n], m2resnum[n],m2atname[n]);
fprintf(fp3, " %2.2f %2.2f   %2.2f", m2h2[n],m2c3[n],d);

fprintf(fp3, "  %2.2f %2.2f %2.2f %2.2f\n",diffh1,diffc4,diffh2,diffc3);
        }
        }
     }
if(k>-1 && l==-1) {
        for(m=0; m<=k; m++) {
fprintf(fp3,"2 %s %d %s",m1resname[m],m1resnum[m],m1atname[m]);
fprintf(fp3," %2.2f %2.2f\n",m1h1[m],m1c4[m]);  } }

if(k==-1 && l>-1) {
   for(n=0; n<=l; n++) {
fprintf(fp3,"\t\t\t  %s %d %s ",m2resname[n], m2resnum[n],m2atname[n]);
fprintf(fp3, "\t\t\t %2.2f %2.2f   %2.2f\n", m2h2[n],m2c3[n],d); } }

   }
  }
```

chhnsortpippn.c

```
/* chhnsortpippn.c  */
/* Program to sort peak pick list h1,c4 high to low  */
/* Sorts peak pick data generated by peak picking */
/* a 4D 13C/15N NOESY using CAPP followed by PIPP */

/* _________ FILES __________ */
/* Input file  : peaklistpippn.PCK (peak pick list) */
/* Output file : hcchsortresultpippn.d (sorted output) */

#include <math.h>
#include <stdio.h>
#include <string.h>

main()
{

FILE *fp, *fp1;
int c, i, j, peaklistc,peakno[5000],peaknot[5000],posneg[5000],posnegt[5000];
float h1[5000], h1t[5000], h2[5000], h2t[5000];
float c3[5000], c3t[5000], c4[5000], c4t[5000];
float wid1,wid2,wid3,wid4,cc;
float intensity[5000], intensityt[5000];
int ptr2,ptr2t;
int ptr3,ptr4;
int zindex[3000], aindex[3000];

printf("starting the program\n");
for(c=0;c<958;c++) {c4[c]=99.99; c4t[c]=99.9;}
/* Open peak list file */
if((fp=fopen("peaklistpippn.PCK", "r"))==NULL) {
   printf("Error opening peaklistpippn.PCK\n");
   exit(1); }

if((fp1=fopen("hcchsortresultpippn.d","w"))==NULL) {
   printf("Error opening hcchsortresultpippn.d\n");
   exit(1); }
printf(" input file peaklistpippn.PCK\n");
printf(" output file hcchsortresultpippn.d\n");
i=-1;
peaklistc=-1;
while(!feof(fp)) {
i=i+1;
peaklistc=peaklistc+1;

fscanf(fp,"%d%d%d%f%f%f%f%e",
&peakno[i],&zindex[i],&aindex[i],&h1[i],&h2[i],&c3[i],&c4[i],&intensity[i]);
}
printf("peaklistc = %d\n", peaklistc);
fclose(fp);
i=0;
/* Do sort  */
```

```
printf("starting sort\n");
for(i=1;i<peaklistc-1;++i) {
   for(j=peaklistc-2;j>=i;--j) {
     if(c4[j-1]>c4[j] || (c4[j-1]==c4[j] && h1[j-1]>h1[j])) {
/*printf("exchange !\n");
printf("%d %d %2.2f %2.2f\n",i, j, c4[j], c4[j-1]);*/

peaknot[j]=peakno[j-1];
h1t[j]=h1[j-1];
intensityt[j]=intensity[j-1];
h2t[j]=h2[j-1];
c3t[j]=c3[j-1];
c4t[j]=c4[j-1];

peakno[j-1]=peakno[j];
h1[j-1]=h1[j];
intensity[j-1]=intensity[j];
h2[j-1]=h2[j];
c3[j-1]=c3[j];
c4[j-1]=c4[j];

peakno[j]=peaknot[j];
h1[j]=h1t[j];
intensity[j]=intensityt[j];
h2[j]=h2t[j];
c3[j]=c3t[j];
c4[j]=c4t[j];
/*printf("%f %f",c4[j], c4[j-1]);*/
} }
}
printf("print the sorted table\n");
for(i=0;i<peaklistc;i++) {
fprintf(fp1,"%d 0 %d %d %3.2f %3.2f %3.2f %3.2f %2.3e \n",
peakno[i],zindex[i],aindex[i],h1[i],h2[i],c3[i],c4[i],intensity[i]);


}
fclose(fp1);
}
```

filterpippn.c

```
/*  filterpippn.c  */
/*  For 4d chhn NOESY */
/*  Program to filter the peak list depending */
/* on whether the peak is unassigned (0), intraresidue (1) */
/* sequential (2), medium range (3), or long range (4) */
/*  These numbers follow the peak number in the output */
/*  from hcchsortpippn.c  - hcchsortpippn.c puts a zero in this place */
/*  the variable containing the number is assign[i] - integer */

/* _________ FILES _______ */
/* Input file  : hcchsortresultpippn.d (sorted peak picks) */
/* Output file : filterresultpippn.d (filtered peak picks) */

#include <math.h>
#include <stdio.h>
#include <string.h>
#include <ctype.h>

main()
{
FILE *fp,*fp1;
int filter, i, peakno, posneg, ptr2,ptr3,ptr4, assign;
float h1, h2, c3, c4,wid1,wid2,wid3,wid4,cc;
float intensity;
int zindex,aindex;



/* Prompt for the type of sort */
printf("Filter the peak list and print out the following\n");
printf("unassigned -   type 0\n");
printf("intraresidue - type 1\n");
printf("sequential -   type 2\n");
printf("medium range - type 3\n");
printf("long-range -   type 4\n");
printf("Type a number:\n");
scanf("%d", &filter);

/* Open peak list table - output from hcchsort.c */
if((fp=fopen("hcchsortresultpippn.d", "r"))==NULL) {
  printf("Error opening hcchsortresultn.d");
  exit(1); }

/*  Open output - filtered list */
if((fp1=fopen("filterresultpippn.d", "w"))==NULL) {
   printf("Error opening filterresultpippn.d");
   exit(1); }
printf(" input file hcchsortresultpippn.d\n");
printf(" output file filterresultpippn.d\n");
/* read in the peak list and write out only those specified */
```

```
while(!feof(fp)) {

fscanf(fp,"%d%d%d%d%f%f%f%f%e",&peakno,&assign,&zindex,&aindex,&h1,&h2,
&c3,&c4,&intensity);
/*printf(" %d  %d\n",peakno,assign);*/
if(assign==filter) {

fprintf(fp1," %d %d %d %d %2.2f %2.2f %2.2f %2.2f %2.2e
\n",peakno,assign,zindex,aindex,h1,h2,c3,c4,intensity);
    }

        }
fclose(fp);
fclose(fp1);
```

```
}
```

filter.c

```
/* filter.c */
/* Program to filter the peak list depending */
/* on whether the peak is unassigned (0), intraresidue (1) */
/* sequential (2), medium range (3), or long range (4) */
/* These numbers follow the peak number in the output */
/* from hcchsort.c - hcchsort.c puts a zero in this place */
/* the variable containing the number is assign[i] - integer */

/* _____ FILES _____ */
/* input file hcchsortresult.d (sorted peak picks) */
/* generated by hcchsort.c */
/* output file filterresult.d */
/* (peak picks with peak catagory data) */


#include <math.h>
#include <stdio.h>
#include <string.h>
#include <ctype.h>

main()
{
FILE *fp,*fp1;
int filter, i, peakno, posneg, ptr2,ptr3,ptr4, assign;
float h1, h2, c3, c4,wid1,wid2,wid3,wid4,cc;
char intensity[10];



/* Prompt for the type of sort */
printf("Filter the peak list and print out the following\n");
printf("unassigned -  type 0\n");
printf("intraresidue - type 1\n");
printf("sequential -  type 2\n");
printf("medium range - type 3\n");
printf("long-range -  type 4\n");
printf("Type a number:\n");
scanf("%d", &filter);

/* Open peak list table - output from hcchsort.c */
if((fp=fopen("hcchsortresult.d", "r"))==NULL) {
  printf("Error opening hcchsortresult.d");
  exit(1); }

/* Open output - filtered list */
if((fp1=fopen("filterresult.d", "w"))==NULL) {
   printf("Error opening filterresult.d");
   exit(1); }

/* read in the peak list and write out only those specified */
```

```
while(!feof(fp)) {

fscanf(fp,"%d%d%f%f%d%s",&peakno,&assign,&h1,&wid1,&posneg,&intensity);
fscanf(fp,"%f%f%d%f%f%d",&h2,&wid2,&ptr2,&c3,&wid3,&ptr3);
fscanf(fp,"%f%f%d%f", &c4,&wid4,&ptr4,&cc);
printf(" %d  %d\n",peakno,assign);
if(assign==filter) {

fprintf(fp1," %d %d  %2.2f %2.2f %d %s",peakno,assign,h1,wid1,posneg,intensity);
fprintf(fp1," %2.2f %2.2f %d %2.2f %f %d",h2,wid2,ptr2,c3,wid3,ptr3);
fprintf(fp1," %2.2f %2.2f %d %2.2f\n", c4,wid4,ptr4,cc);  }

        }
fclose(fp);
fclose(fp1);


}
```

hbpairs.c

```
/*  hbpairs.c  */
/*  Program to pick out pairs of peaks */
/*  ie. Pattern recognition to make assignments */
/*  in 4D HCCH NOESY spectra */
/*  Input - pickresult.d from hcchassign.c */

/* _________ FILES _________ */
/* Input file  : pickresult.d (NOE assignments file) */
/* Output file : hbpairsresult.d */
/*            ( matched HB for pair  and same CB on */
/*              two NOE assignments) */


#include <math.h>
#include <stdio.h>
#include <string.h>

main()
{
FILE *fp, *fp1;
int lineid, peakno[5000], posneg[5000], i, j, k, l, totpeaks, pairs[5000];
float h1[5000], c4[5000], h2[5000], c3[5000];
char intensity[5000][30];
char m1resname[5000][50][5], m1atname[5000][50][5], hb1[5], hb2[5];
char m2resname[5000][50][5], m2atname[5000][50][5], nopairs[5000][50][30];
int m1resnum[5000][50], m2resnum[5000][50];
float m1h1[5000][50], m1c4[5000][50], m2h2[5000][50], m2c3[5000][50];
float d[5000][50], diffh1[5000][50], diffh2[5000][50], diffc3[5000][50];
float diffc4[5000][50];
float c4pl[5000],c3pl[5000];
char pl[5],atnam[5][5][5];
```

```
int h;

/* Open pickresult.d file for read */

 if((fp=fopen("pickresult.d", "r"))== NULL) {
    printf("Error opening pickresult.d");
    exit(1); }

/* Open file to write results */

  if((fp1=fopen("hbpairsresults.d", "w"))==NULL) {
    printf("Error opening hbpairsresults.d");
    exit(1); }

/* Read pickresult.d file into 2d Arrays */

/* Read first number to determine format and what */
/* the line is */
i=-1;
totpeaks=0;
while(!feof(fp)) {

 fscanf(fp,"%d", &lineid);
 if(lineid==1) {
     i=i+1;
     pairs[i]=0;
     totpeaks=totpeaks + 1;
     fscanf(fp, "%d%f%f%s%f%d", &peakno[i], &h1[i], &c4[i], &pl, &c4pl[i],
&posneg[i]);
     fscanf(fp, "%s%f%f%s%f", &intensity[i], &h2[i], &c3[i],&pl,&c3pl[i]);}
 if(lineid==2) { j=j+1;
    pairs[i]=pairs[i]+1;
    fscanf(fp, "%s%d%s", &m1resname[i][j],&m1resnum[i][j],&m1atname[i][j]);
    fscanf(fp, "%f%f", &m1h1[i][j], &m1c4[i][j]);
    fscanf(fp, "%s%d%s", &m2resname[i][j], &m2resnum[i][j], &m2atname[i][j]);
    fscanf(fp, "%f%f%f", &m2h2[i][j], &m2c3[i][j], &d[i][j]);
    fscanf(fp, "%f%f%f%f", &diffh1[i][j],&diffc4[i][j],&diffh2[i][j],&diffc3[i][j]); }
 if(lineid==3)
    fscanf(fp, "%s", &nopairs[i][j]);
}

strcpy(atnam[0][0],"HB1");
strcpy(atnam[0][1],"HB2");
strcpy(atnam[1][0],"HG1");
strcpy(atnam[1][1],"HG2");
strcpy(atnam[2][0],"HD1");
strcpy(atnam[2][1],"HD2");

/* Check for pattern - one CH pair going to two hb1 or hb2 */
for(h=0; h<2; h++) {

 for(i=0; i<totpeaks; i++) {
```

```
    for(j=0; j<pairs[i]; j++) {

if(strcmp(m2atname[i][j],atnam[h][0])==0 || strcmp(m2atname[i][j],atnam[h][1])==0) {
  printf("made it through m2atname");
     for(k=0; k<totpeaks; k++) {
       for(l=0; l<pairs[k]; l++) {

if(strcmp(m1atname[i][j],m1atname[k][l])==0 &&
strcmp(m1resname[i][j],m1resname[k][l])==0 && m1resnum[i][j]==m1resnum[k][l] &&
(strcmp(m2atname[k][l],atnam[h][0])==0 || strcmp(m2atname[k][l],atnam[h][1])==0) &&
strcmp(m2resname[i][j],m2resname[k][l])==0 && m2resnum[i][j]==m2resnum[k][l] &&
peakno[i] != peakno[k]) {
  /* Print the results  */
fprintf(fp1,"\n   *** X - %s Pair  ***", atnam[h][0]);

fprintf(fp1,"\n %d %2.2f %2.2f pl %2.1f %d",peakno[i],h1[i],c4[i],c4pl[i], posneg[i]);
fprintf(fp1," %s  %2.2f %2.2f pl %2.1f\n", intensity[i], h2[i], c3[i], c3pl[i]);
fprintf(fp1,"%s %d %s", m1resname[i][j],m1resnum[i][j],m1atname[i][j]);
fprintf(fp1,"  %2.2f %2.2f", m1h1[i][j], m1c4[i][j]);
fprintf(fp1,"  %s %d %s", m2resname[i][j],m2resnum[i][j],m2atname[i][j]);
fprintf(fp1,"  %2.2f %2.2f  %2.2f\n", m2h2[i][j], m2c3[i][j],d[i][j]);

fprintf(fp1," %d %2.2f %2.2f %d",peakno[k],h1[k],c4[k],posneg[k]);
fprintf(fp1," %s  %2.2f %2.2f \n", intensity[k], h2[k], c3[k]);
fprintf(fp1,"%s %d %s", m1resname[k][l],m1resnum[k][l],m1atname[k][l]);
fprintf(fp1,"  %2.2f %2.2f", m1h1[k][l], m1c4[k][l]);
fprintf(fp1,"  %s %d %s", m2resname[k][l],m2resnum[k][l],m2atname[k][l]);
fprintf(fp1,"  %2.2f %2.2f  %2.2f\n", m2h2[k][l], m2c3[k][l],d[k][l]);

}  }}  }} }

/*  Check for HB1 - X pair */

for(i=0; i<totpeaks; i++) {
   for(j=0; j<pairs[i]; j++) {

if(strcmp(m1atname[i][j],atnam[h][0])==0 || strcmp(m1atname[i][j],atnam[h][1])==0) {

    for(k=0; k<totpeaks; k++) {
     for(l=0; l<pairs[k]; l++) {

if((strcmp(m1atname[k][l],atnam[h][0])==0 || strcmp(m1atname[i][j],atnam[h][1])==0)
&& strcmp(m1resname[k][l],m1resname[i][j])==0 && m1resnum[k][l]==m2resnum[i][j]
&& strcmp(m2atname[k][l],m2atname[i][j])==0 &&
strcmp(m2resname[k][l],m2resname[i][j])==0 && m2resnum[k][l]==m2resnum[i][j] &&
peakno[i]!= peakno[k]) {

fprintf(fp1,"\n    *** %s, %s - X Pair\n", atnam[h][0], atnam[h][1]);

fprintf(fp1,"\n %d %2.2f %2.2f pl %2.1f %d",peakno[i],h1[i],c4[i],c4pl[i],posneg[i]);
fprintf(fp1," %s %2.2f %2.2f pl %2.1f\n", intensity[i],h2[i],c3[i],c3pl[i]);
fprintf(fp1," %s %d %s",m1resname[i][j],m1resnum[i][j],m1atname[i][j]);
```

```
fprintf(fp1," %2.2f %2.2f",m1h1[i][j],m1c4[i][j]);
fprintf(fp1,"  %s %d %s",m2resname[i][j], m2resnum[i][j],m2atname[i][j]);
fprintf(fp1,"  %2.2f %2.2f  %2.2f\n",m2h2[i][j],m2c3[i][j],d[i][j]);

fprintf(fp1," %d %2.2f %2.2f %d",peakno[k],h1[k],c4[k],posneg[k]);
fprintf(fp1," %s %2.2f %2.2f \n", intensity[k],h2[k],c3[k]);
fprintf(fp1," %s %d %s", m1resname[k][l],m1resnum[k][l],m1atname[k][l]);
fprintf(fp1,"  %2.2f %2.2f", m1h1[k][l], m1c4[k][l]);
fprintf(fp1,"  %s %d %s",m2resname[k][l],m2resnum[k][l],m2atname[k][l]);
fprintf(fp1,"  %2.2f %2.2f  %2.2f\n",m2h2[k][l],m2c3[k][l],d[k][l]);
}}}}}}
  }
}
```

hcchassignpipp.c

```
/* hcchassignpipp.c */
/* derived from hcchassign.c */
/* Written by JGP */
/* Program to analyze peak picks from 4D expt */
/* generated by program PIPP */
/* to compile use: cc hcchassignpipp.c -lm */
/*The -lm is for inclusion of the math.h file */
/* To be used for 13C - 13C 4D NOESY only */

/* for a short description of the rationale */
/* behind this program - see the header of */
/* the program: hcchassign.c */

/* _________ FILES READ IN _________ */
/* reads in pickresult.d (raw peak pick file)  */
/* reads in assign.d (frequency assignments file) */
/* reads in filterresultpipp.d (assignment type file) */
/* not assigned=0, short=1, medium=2, long=3 */
/* reads in xyz.d (coordinate file) */


#include <math.h>
#include <stdio.h>
#include <string.h>

main()
{
FILE *fp, *fp1, *fx2, *fp3;
int i, resnum[3000], atomnumc, resnumc[3000], peakno[5000], posneg[5000];
int ptr2, ptr3, ptr4;
char atname[3000][10], resname[3000][5], atomc[10], atnamec[3000][10];
char resnamec[3000][5],c212c[10];
float xc[3000], yc[3000], zc[3000], c1, bfc, h[3000], c[3000], h1[5000];
float wid1, h2[5000], wid2, c3[5000], wid3, c4[5000], wid4, cc;
float x[3000], y[3000], z[3000];
int xyzc, peaklistc, assignc, j;
int k, l, m, n, m1resnum[100], m2resnum[100];
```

```
char m1resname[100][5], m2resname[100][5], m1atname[100][10],m2atname[100][10];
float m1h1[100], m1c4[100], m2h2[100], m2c3[100], m1x[100], m1y[100];
float m1z[100], m2x[100], m2y[100], m2z[100];
float diffh1, diffc3, diffh2, diffc4;
double a, b, cd, d, dd;
float h1cut, h2cut, ccut, cfold, cfold3, cfold4, h1dev, c4dev, h2dev, c3dev;
int t, assign[5000];
float c4pl[5000],c3pl[5000],dcut,intensity[3000];
int aindex[3000], zindex[3000];

printf("input file filterresultpipp.d, assign.d, xyz.d \n");
printf(" output peak picks/assignments pickresultpipp.d \n");
/*  Define 1H and 13C chemical shift ranges to be tested */
/*  when assigning peaks from the peak list */
/* hcut for 1H and ccut for 13C */

h1cut=0.12;  /* cutoff for d1 dimension 1H - with c4 */
h2cut=0.15;  /* cutoff for d2 dimension 1H - with c3 */
ccut=0.5;
cfold=23.8;
cfold3=3*23.8; /* for 3 folded aromatics */
cfold4=4*23.8; /* for 4 folded aromatics */
h1dev=0.0;   /*deviation h1[j]-h1dev=h[j] (h1[j] is peak picked data)*/
h2dev=0.0;   /* h[j] is chem shift assignment data */
c3dev=0.0;
c4dev=0.0;
dcut=15.0;  /* distance cutoff for printing possible proton pairs */
/* Open file to write results */
  if ((fp3=fopen("pickresultpipp.d", "w")) ==NULL) {
    printf("Error opening pickresultpipp.d");
    exit(1); }


/* Open data bases and read to arrays */
/* Open assignments + read data as assign.d */
  if ((fp=fopen("assign.d", "r")) ==NULL) {
    printf("Error opening assign.d/n");
    exit(1); }
printf("Opened chemical shift list assign.d");
printf(" Reading in the file \n");
/* resnum = residue number; resname= residue name; assign =assignment;*/
/* h = 1H chemical shift; c = carbon chemical shift */
  i=-1;
  assignc=-1;
  while(!feof(fp)) {
  i=i+1;
  assignc=assignc+1;
  fscanf(fp, "%i%s%s%f%f", &resnum[i], &resname[i], &atname[i], &h[i], &c[i]);
/*   printf("%i %s %s %2.2f %3.2f\n", resnum[i], resname[i], atname[i],h[i],c[i]); */
      }
  fclose(fp);
```

```
/* Read in peak pick list */

  if ((fp=fopen("filterresultpipp.d", "r")) == NULL) {
    printf("Error opening filterresultpipp.d\n");
    exit(1); }
printf(" Opened peak pick list peaklist.d - reading in the file \n");
  i=-1;
  peaklistc=-1;
  while(!feof(fp)) {
  i=i+1;
  peaklistc=peaklistc+1;
fscanf(fp, "%d%d%d%d%f%f%f%f%e", &peakno[i], &assign[i], &zindex[i],
&aindex[i], &h1[i], &h2[i], &c3[i], &c4[i], &intensity[i]);

/*  Calculate plane number for each c3 and c4   */
c4pl[i]=17-((c4[i]-43.0)/0.74);
c3pl[i]=17-((c3[i]-43.0)/0.74);

printf("%d %2.2f %2.2f %2.2f %2.2f", peakno[i], h1[i], h2[i], c3[i], c4[i]);
printf(" %2.3e c4 plane %2.1f  c3 plane %2.1f\n\n", intensity[i], c4pl[i], c3pl[i]);
}
 fclose(fp);

/*  Read in xyz coordinates for a structure */

  if ((fp1=fopen("xyz.d", "r")) == NULL) {
   printf("Error opening xyz.d\n");
   exit(1); }
printf("Opened xyz coordinate file xyz.d - reading coordinates \n");
  i=-1;
  xyzc=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc=xyzc+1;
   fscanf(fp, "%s%d", &atomc, &atomnumc);
   fscanf(fp, "%s%s%d%f%f%f", &atnamec[i],
&resnamec[i],&resnumc[i],&xc[i],&yc[i],&zc[i]);
   fscanf(fp, "%f%f%s", &c1, &bfc, &c212c);
/*   printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[i], resnamec[i], resnumc[i], xc[i],
yc[i], zc[i]); */
}
   fclose(fp);

/* ** Match chemical shifts with xyz coordinates ** */

printf("Matching assignments with coordinates\n");
 for(i=0; i<assignc; i++) {
   for(j=0; j<xyzc; j++) {
    if(strcmp(atname[i],atnamec[j])==0 && strcmp(resname[i],resnamec[j])==0 &&
        resnum[i]==resnumc[j]) {
                              x[i]=xc[j];
                              y[i]=yc[j];
```

```
                        z[i]=zc[j];
/*      printf("%s %s %d assigned to", resnamec[j], atnamec[j], resnumc[j]);
      printf(" %s %s %d", resname[i], atname[i], resnum[i]);
      printf(" xyz coordinates %2.3f %2.3f %2.3f\n", x[i], y[i], z[i]);
      if(x[i]==0.0 || y[i]==0.0 || z[i]==0.0) printf("Error -  no match");*/
}
}
}
/* Match peak list to assignment table  */

  for(i=0; i<peaklistc ; i++) {
k=-1;
l=-1;

/* Check to see if peak in peak list is an aromatic */
/* Use a cutoff : if 1H shift is above 6.0 ppm */

/*if(h1[i]>6.0) c4[i]=c4[i]+(4*23.8);*/
/*if(h2[i]>6.0) c3[i]=c3[i]+(4*23.8);*/

   for(j=0; j<assignc; j++) {

if(fabs(h1[i]-h[j]-h1dev)<h1cut && (fabs(c4[i]-c[j]-c4dev)<ccut || fabs(c4[i]+cfold-c4dev-
c[j])<ccut || fabs(c4[i]-cfold-c4dev-c[j])<ccut || fabs(c4[i]+cfold3-c4dev-c[j])<ccut ||
fabs(c4[i]+cfold4-c4dev-c[j])<ccut)) {
        k=k+1;
        strcpy(m1atname[k],atname[j]);
        strcpy(m1resname[k],resname[j]);
        m1resnum[k]=resnum[j];
        m1h1[k]=h[j];
        m1c4[k]=c[j];
        m1x[k]=x[j];
        m1y[k]=y[j];
        m1z[k]=z[j];
        }
if(fabs(h2[i]-h[j]-h2dev)<h2cut && (fabs(c3[i]-c[j]-c3dev)<ccut || fabs(c3[i]+cfold-c[j]-
c3dev)<ccut || fabs(c3[i]-cfold-c[j]-c3dev)<ccut || fabs(c3[i]+cfold3-c[j]-c3dev)<ccut ||
fabs(c3[i]+cfold4-c[j]-c3dev)<ccut)) {
        l=l+1;
        strcpy(m2atname[l],atname[j]);
        strcpy(m2resname[l],resname[j]);
        m2resnum[l]=resnum[j];
        m2h2[l]=h[j];
        m2c3[l]=c[j];
        m2x[l]=x[j];
        m2y[l]=y[j];
        m2z[l]=z[j];
        }
/*Use info on parameter posneg to filter some choices - */
/* if(h1[i]<6.0 && h2[i]<6.0) {

if(posneg[i]==2 && (fabs(c4[i]-c[j])<ccut && fabs(c3[i]-c[j])<ccut)) continue;
```

```
if(posneg[i]==0 && (fabs(c4[i]-c[j])<ccut || fabs(c3[i]-c[j])<ccut)) continue;}
*/
      }
   /* Write out matches and calculate distance for each */

fprintf(fp3,"\n1     %d %d   %2.2f %2.2f  ",peakno[i],assign[i],h1[i],c4[i]);
fprintf(fp3, " %2.3e   %2.2f %2.2f",intensity[i],h2[i],c3[i]);
fprintf(fp3, "  c4 plane  %2.1f   c3plane  %2.1f\n\n", c4pl[i], c3pl[i]);

if(k==-1 || l==-1) { fprintf(fp3,"3          **No pairs found**\n");
                continue; }

if(k>-1 && l>-1) {
        for(m=0; m<=k; m++) {


     for(n=0; n<=l; n++) {
        a=(m1x[m]-m2x[n])*(m1x[m]-m2x[n]);
        b=(m1y[m]-m2y[n])*(m1y[m]-m2y[n]);
        cd=(m1z[m]-m2z[n])*(m1z[m]-m2z[n]);
        dd=a+b+cd;

        if(dd<=0.0) d=0.0;
        else d=sqrtf(a+b+cd);
        if(m1x[m]==0.0 || m2x[n]==0.0) d=0.0;
        if(d >= dcut) continue;

        diffh1=h1[i]-m1h1[m];
        diffh2=h2[i]-m2h2[n];
        diffc3=c3[i]-m2c3[n];
        diffc4=c4[i]-m1c4[m];

fprintf(fp3,"2 %s %d %s",m1resname[m],m1resnum[m],m1atname[m]);
fprintf(fp3," %2.2f %2.2f",m1h1[m],m1c4[m]);

fprintf(fp3,"  %s %d %s ",m2resname[n], m2resnum[n],m2atname[n]);
fprintf(fp3, " %2.2f %2.2f   %2.2f", m2h2[n],m2c3[n],d);

fprintf(fp3, "  %2.2f %2.2f %2.2f %2.2f\n",diffh1,diffc4,diffh2,diffc3);
        }
        }
     }
   }
  }
```

hcchassign.c

```
/* hcchassign.c */
/* Program to analyze peak picks from 4D expt */
/* to compile use: cc hcchassign.c -lm */
/*The -lm is for inclusion of the math.h file */
/* To be used for 13C - 13C 4D NOESY only */

/* ________ SHORT DESCRIPTION ___________ */
/* This program written by Jeff Pelton */
/* is the workhorse of our assignment */
/* of crosspeaks peak picked in the 4D */
/* 13C/13C NOESY using the FELIX program */
/* and macros for manual peak picking */
/* developed by A.Lee and modified slightly */
/* by Fred Damberger */
/* In summary it generates an output file */
/* containing all NOE assignments for each */
/* crosspeak within a certain ppm cutoff of */
/* the assigned chemical shifts along with the */
/* distance between proton pairs in each proposed */
/* assignment as computed from a reference */
/* structure from a previous round of structure */
/* calculations */
/* Typically, the generated file is printed out */
/* and the user manually inspects the listing to */
/* find unambiguous NOE assignments - this is
/* loosely defined as a single assigned proton pair */
/* is within 6 Angstroms (A) and all others are much*/
/* further away for a given crosspeak. These guidelines */
/* for assigning new NOEs change as the structure */
/* refinement proceeds. A tightly defined family allows */
/* more precise determination of distances and allows */
/* NOEs to be assigned when one distance is within the
/* range to give rise to an NOE (say <5 A) and all others */
/* are outside that range (say >6 A). NOTE that at this */
/* stage we used the program dist.c to get distance ranges */
/* for the family of structures rather than just one as */
/* used in this program  -FFD */

/* _________ FILES _________ */
/* Input file  : assign.d (frequency assignments file) */
/* Input file  : filterresult.d (processed peak pick file) */
/* not assigned=0, short=1, medium=2, long=3 */
/* Input file  : reads in xyz.d (coordinate file) */
/* Output file : pickresult.d (NOE assignment output) */

/* _______________ LONG DESCRIPTION _____________ */

/* pickresult.d is generated by FELIX as an output */
/* list file from the peak pick data base generated */
```

```
/* by the felix macros: 4dpeakpick.mac */

/* This program reads in filterresult.d to get the */
/* location of 4D peaks in the 13C/13C NOESY */
/* filterresult.d is derived from peaklist.d */
/* It has been sorted by hcchsort.c and modified */
/* by filter.c to include information on whether peaks */
/* have been picked in a previous round of assignment */
/* This program then reads in assign.d to obtain */
/* the most recent chemical shift assignments */
/* It also reads in the coordinates from the structure */
/* file xyz.d (which should be a converged structure */
/* from the previous round of structure refinement) */
/* Then it takes each crosspeak in filterresult.d */
/* and trys to assign both 1H/13C chemical shift pairs */
/* to 1H/13C pairs in the assign.d file */
/* typical maximal ppm differences accepted as a potential */
/* assignment are : delta ppm 1H  = +/- 0.15  */
/*     parameters: h1cut (c4) and h2cut (c3) */
/*             : delta ppm 13C = +/- 0.5   */
/*     parameters: ccut                    */
/* each potential assignment is written to the file: pickresult.d */
/* along with the errors in ppm values from assign.d */
/* and the distance between two protons in the two assigned */
/* groups (as determined from the xyz.d structure). */
/* Note that this program does not used pseudo-atoms to */
/* calculate the distances - it arbitrarily picks one 1H from */
/* each group participating in the NOE, e.g. H12 in a valine */
/* methyl group. The program also uses these distances as a */
/* mild filter excluding assignments with distances corresponding */
/* to a distance greater than 15 A. */

/*_____ SUPPORTING PROGRAMS _____*/

/* FELIX */
/
/* FELIX 2.30 is used to generate the peakpick.d file. */
/* After processing the raw 4D data matrix using FELIX */
/* The macros developed by Andrew Lee are used to pick */
/* crosspeaks which occur in each plane manually. The  */
/* following macros were used :                  */
/* 2dfrom4dmnu.mac generates a 2D plane from a 4D matrix */
/* In the default mode, the user picks the plane numbers */
/* for the two 13C dimensions (D2 and D3) and the macro */
/* plots the corresponding 2D 1H/1H plane */
/* manpick4dmnu.mac then allows the user to interactively */
/* pick crosspeaks in the spectrum. The macro obtains the */
/* 1H/1H position of the cursor and plots the corresponding */
/* 13C/13C plane to locate the crosspeak position in the 13C */
/* dimensions. It then asks the user if he wishes to enter */
/* the peak and if so asks for a qualitative assessment of */
/* intensity (strong - medium or weak). If the peak was chosen */
```

```
/* for entry, the 1H/13C -1H/13C position and intensity are
/* entered into the open data base along with the peak */
/* intensity. These two macros are integrated into the menu */
/* interface of FELIX under 2D mode Tools option by modification */
/* of the file display2d.mnu and the creation of several
/* z_filename.mnu files placed in the directory pmnu */
/* The list option is used to generate an output file of the */
/* data base which should be named (or renamed) peaklist.d */

/* C programs */

/* peaklist.d is input file for the program hcchsort.c */
/* which sorts the peak pick list by 1H and 13C chemical shift */
/* output file is  hcchsortresult.d */
/* hcchsortresult.d is input file for the program filter.c */
/* which filters the peak pick list based on which peaks have */
/* already been assigned and includes a column in the output
/* file for type of assignment of each peak: not assigned, */
/* short range, medium range, and long range */
/* the output file is filterresult.d */
/* filterresult.d is input file for this program (hcchassign.c) */


/* _____ END OF HEADER _____ */

#include <math.h>
#include <stdio.h>
#include <string.h>

main()
{
FILE *fp, *fp1, *fx2, *fp3;
int i, resnum[3000], atomnumc, resnumc[3000], peakno[5000], posneg[5000];
int ptr2, ptr3, ptr4;
char atname[3000][10], resname[3000][5], atomc[10], atnamec[3000][10];
char resnamec[3000][5],c212c[10], intensity[5000][10];
float xc[3000], yc[3000], zc[3000], c1, bfc, h[3000], c[3000], h1[5000];
float wid1, h2[5000], wid2, c3[5000], wid3, c4[5000], wid4, cc;
float x[3000], y[3000], z[3000];
int xyzc, peaklistc, assignc, j;
int k, l, m, n, m1resnum[100], m2resnum[100];
char m1resname[100][5], m2resname[100][5], m1atname[100][10],m2atname[100][10];
float m1h1[100], m1c4[100], m2h2[100], m2c3[100], m1x[100], m1y[100];
float m1z[100], m2x[100], m2y[100], m2z[100];
float diffh1, diffc3, diffh2, diffc4;
double a, b, cd, d, dd;
float h1cut, h2cut, ccut, cfold, cfold3, cfold4, h1dev, c4dev, h2dev, c3dev;
int t, assign[5000];
float c4pl[5000],c3pl[5000],dcut;

/*  Define 1H and 13C chemical shift ranges to be tested */
/*  when assigning peaks from the peak list */
```

```
/* hcut for 1H and ccut for 13C */

h1cut=0.12;  /* cutoff for d1 dimension 1H - with c4 */
h2cut=0.15;  /* cutoff for d2 dimension 1H - with c3 */
ccut=0.5;
cfold=23.8;
cfold3=3*23.8; /* for 3 folded aromatics */
cfold4=4*23.8; /* for 4 folded aromatics */
h1dev=0.0;   /*deviation h1[j]-h1dev=h[j] (h1[j] is peak picked data)*/
h2dev=0.0;   /* h[j] is chem shift assignment data */
c3dev=0.0;
c4dev=0.0;
dcut=15.0;  /* distance cutoff for printing possible proton pairs */
/* Open file to write results */
  if ((fp3=fopen("pickresult.d", "w")) ==NULL) {
    printf("Error opening pickresult.d");
    exit(1); }


/* Open data bases and read to arrays */
/* Open assignments + read data as assign.d */
  if ((fp=fopen("assign.d", "r")) ==NULL) {
    printf("Error opening assign.d/n");
    exit(1); }
printf("Opened chemical shift list assign.d");
printf(" Reading in the file \n");
/* resnum = residue number; resname= residue name; assign =assignment;*/
/* h = 1H chemical shift; c = carbon chemical shift */
  i=-1;
  assignc=-1;
  while(!feof(fp)) {
  i=i+1;
  assignc=assignc+1;
  fscanf(fp, "%i%s%s%f%f", &resnum[i], &resname[i], &atname[i], &h[i], &c[i]);
  printf("%i %s %s %2.2f %3.2f\n", resnum[i], resname[i], atname[i],h[i],c[i]);
      }
  fclose(fp);

/* Read in peak pick list */

  if ((fp=fopen("filterresult.d", "r")) == NULL) {
    printf("Error opening filterresult.d\n");
    exit(1); }
printf(" Opened peak pick list peaklist.d - reading in the file \n");
  i=-1;
  peaklistc=-1;
  while(!feof(fp)) {
  i=i+1;
  peaklistc=peaklistc+1;
fscanf(fp, "%d%d%f%f%d%s", &peakno[i], &assign[i], &h1[i], &wid1, &posneg[i],
&intensity[i]);
fscanf(fp, "%f%f%d%f%f%d", &h2[i], &wid2, &ptr2, &c3[i], &wid3, &ptr3);
```

```
fscanf(fp, "%f%f%d%f", &c4[i], &wid4, &ptr4, &cc);

/*  Calculate plane number for each c3 and c4    */
c4pl[i]=17-((c4[i]-43.0)/0.74);
c3pl[i]=17-((c3[i]-43.0)/0.74);

printf("%d %2.2f %2.2f %2.2f %2.2f", peakno[i], h1[i], h2[i], c3[i], c4[i]);
printf(" %d %s c4 plane %2.1f  c3 plane %2.1f\n\n", posneg[i], intensity[i], c4pl[i],
c3pl[i]);
}
 fclose(fp);

/*  Read in xyz coordinates for a structure */

  if ((fp1=fopen("xyz.d", "r")) == NULL) {
   printf("Error opening xyz.d\n");
   exit(1); }
printf("Opened xyz coordinate file xyz.d - reading coordinates \n");
  i=-1;
  xyzc=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc=xyzc+1;
   fscanf(fp, "%s%d", &atomc, &atomnumc);
   fscanf(fp, "%s%s%d%f%f%f", &atnamec[i],
&resnamec[i],&resnumc[i],&xc[i],&yc[i],&zc[i]);
   fscanf(fp, "%f%f%s", &c1, &bfc, &c212c);
   printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[i], resnamec[i], resnumc[i], xc[i],
yc[i], zc[i]);
}
   fclose(fp);

/*  **  Match chemical shifts with xyz coordinates  ** */

printf("Matching assignments with coordinates\n");
 for(i=0; i<assignc; i++) {
   for(j=0; j<xyzc; j++) {
    if(strcmp(atname[i],atnamec[j])==0 && strcmp(resname[i],resnamec[j])==0 &&
        resnum[i]==resnumc[j]) {
                                x[i]=xc[j];
                                y[i]=yc[j];
                                z[i]=zc[j];
      printf("%s %s %d assigned to", resnamec[j], atnamec[j], resnumc[j]);
      printf(" %s %s %d", resname[i], atname[i], resnum[i]);
      printf(" xyz coordinates %2.3f %2.3f %2.3f\n", x[i], y[i], z[i]);
      if(x[i]==0.0 || y[i]==0.0 || z[i]==0.0) printf("Error -  no match");
}
}
}
/*  Match peak list to assignment table  */

  for(i=0; i<peaklistc ; i++) {
```

```
k=-1;
l=-1;

/* Check to see if peak in peak list is an aromatic */
/* Use a cutoff : if 1H shift is above 6.0 ppm */

/*if(h1[i]>6.0) c4[i]=c4[i]+(4*23.8);*/
/*if(h2[i]>6.0) c3[i]=c3[i]+(4*23.8);*/

   for(j=0; j<assignc; j++) {

if(fabs(h1[i]-h[j]-h1dev)<h1cut && (fabs(c4[i]-c[j]-c4dev)<ccut || fabs(c4[i]+cfold-c4dev-
c[j])<ccut || fabs(c4[i]-cfold-c4dev-c[j])<ccut || fabs(c4[i]+cfold3-c4dev-c[j])<ccut ||
fabs(c4[i]+cfold4-c4dev-c[j])<ccut)) {
        k=k+1;
        strcpy(m1atname[k],atname[j]);
        strcpy(m1resname[k],resname[j]);
        m1resnum[k]=resnum[j];
        m1h1[k]=h[j];
        m1c4[k]=c[j];
        m1x[k]=x[j];
        m1y[k]=y[j];
        m1z[k]=z[j];
        }
if(fabs(h2[i]-h[j]-h2dev)<h2cut && (fabs(c3[i]-c[j]-c3dev)<ccut || fabs(c3[i]+cfold-c[j]-
c3dev)<ccut || fabs(c3[i]-cfold-c[j]-c3dev)<ccut || fabs(c3[i]+cfold3-c[j]-c3dev)<ccut ||
fabs(c3[i]+cfold4-c[j]-c3dev)<ccut)) {
        l=l+1;
        strcpy(m2atname[l],atname[j]);
        strcpy(m2resname[l],resname[j]);
        m2resnum[l]=resnum[j];
        m2h2[l]=h[j];
        m2c3[l]=c[j];
        m2x[l]=x[j];
        m2y[l]=y[j];
        m2z[l]=z[j];
        }
/*Use info on parameter posneg to filter some choices - */
if(h1[i]<6.0 && h2[i]<6.0) {

if(posneg[i]==2 && (fabs(c4[i]-c[j])<ccut && fabs(c3[i]-c[j])<ccut)) continue;
if(posneg[i]==0 && (fabs(c4[i]-c[j])<ccut || fabs(c3[i]-c[j])<ccut)) continue;}


    }
  /* Write out matches and calculate distance for each */

fprintf(fp3,"\n1     %d %d   %2.2f %2.2f  %d",peakno[i],assign[i],h1[i],c4[i],posneg[i]);
fprintf(fp3, " %s   %2.2f %2.2f",intensity[i],h2[i],c3[i]);
fprintf(fp3, "  c4 plane  %2.1f   c3plane  %2.1f\n\n", c4pl[i], c3pl[i]);

if(k==-1 || l==-1) { fprintf(fp3,"3          **No pairs found**\n");
```

```
                }

if(k>-1 && l>-1) {
        for(m=0; m<=k; m++) {


     for(n=0; n<=l; n++) {
        a=(m1x[m]-m2x[n])*(m1x[m]-m2x[n]);
        b=(m1y[m]-m2y[n])*(m1y[m]-m2y[n]);
        cd=(m1z[m]-m2z[n])*(m1z[m]-m2z[n]);
        dd=a+b+cd;

        if(dd<=0.0) d=0.0;
        else d=sqrtf(a+b+cd);
        if(m1x[m]==0.0 || m2x[n]==0.0) d=0.0;
        /*if(d >= dcut) continue;*/

        diffh1=h1[i]-m1h1[m];
        diffh2=h2[i]-m2h2[n];
        diffc3=c3[i]-m2c3[n];
        diffc4=c4[i]-m1c4[m];

fprintf(fp3,"2 %s %d %s",m1resname[m],m1resnum[m],m1atname[m]);
fprintf(fp3," %2.2f %2.2f",m1h1[m],m1c4[m]);

fprintf(fp3,"  %s %d %s ",m2resname[n], m2resnum[n],m2atname[n]);
fprintf(fp3, " %2.2f %2.2f   %2.2f", m2h2[n],m2c3[n],d);

fprintf(fp3, "  %2.2f %2.2f %2.2f %2.2f\n",diffh1,diffc4,diffh2,diffc3);
        }
        }
     }
if(k>-1 && l==-1) {
        for(m=0; m<=k; m++) {
fprintf(fp3,"2 \t\t\t %s %d %s",m1resname[m],m1resnum[m],m1atname[m]);
fprintf(fp3," %2.2f %2.2f\n",m1h1[m],m1c4[m]);  }}

if(k==-1 && l>-1) {
     for(n=0; n<=l; n++) {
fprintf(fp3,"\t\t\t  %s %d %s ",m2resname[n], m2resnum[n],m2atname[n]);
fprintf(fp3, " %2.2f %2.2f   %2.2f\n", m2h2[n],m2c3[n],d); }}

   }
  }
```

C Software for Analysis of Structures and Contraint Data

dist.c

```
/*          dist.c      */
/* Program to calculate distances between two atoms */
/* for a family of structures  */
/* With a distance cutoff */
/* to compile use cc dist.c -lm (-lm to include math utility)*/
/* Input format: dist resN atmX resM atmY          */
/* for example : dist 22 HA 23 HN                  */
/* program then prompts user for prefix of structure files */
/* followed by number of structures */
/* for example : refine_          */
/*           : 30              */
/* so structure files must have the format prefixNUMBER.pdb */
/* output is sent to standard out (screen) and consists of  */
/* structure name resN atmX resM atmY distance(in angstroms) */
/* for all the structures specified along with the average */
/* and standard deviation */
/* This code forms the core of many other programs in this appendix */
/* such as hbond.c -FFD*/

/* Author: Jeff Pelton */

#include <math.h>
#include <stdio.h>
#include <string.h>

void main(int argc, char *argv[])
{
FILE *fp1, *fp3;
char *p;
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],filename1[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
/*char fileno[2][2]={"1","2"};*/
char
fileno[24][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24"};
/* change this to read in more structures along with size of array fileno */
if(argc != 5) {printf("Usage: dist res_number atname res_num atname");
exit(1); }
strcpy(pdb,".pdb");
resnotest1=atoi(argv[1]);
resnotest2=atoi(argv[3]);
strcpy(attest1,argv[2]);
```

```
strcpy(attest2,argv[4]);
printf("Looking for %d  %s  %d  %s\n",atoi(argv[1]),argv[2],atoi(argv[3]),argv[4]);

/*  **** Get basic information interactively  ***** */
printf("enter Base filename -ex. accept_  :\n");
/** **/
p=gets(s);
if(p) strcpy(filename,s);
strcpy(filename1,filename);
printf("enter number of files - default 24: \n");
p=gets(s);
if(p) no=atoi(s);
printf("number of files is: %d / Base file name is %s .\n",no,filename);
/***          Open file for output    ******************************** */
/*  if ((fp3=fopen("nhch3dist.output", "w")) == NULL) {
   printf("Error opening nhch3dist.output\n");
   exit(1); } */
/***************  Read in xyz coordinates for structures *******************/
for(k=0;k<no;k++) {
 /* add="1";*/
  strcat(filename,fileno[k]);
  strcat(filename,pdb);
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening file\n");
   exit(1); }
/*printf("Opened xyz coordinate file %s - reading coordinates \n",filename);*/
  strcpy(filename,filename1);
  i=-1;
  xyzc[k]=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc[k]=xyzc[k]+1;
   fscanf(fp1,"%s",&atomc);
  if(strcmp(atomc,"ATOM")==0) {
   fscanf(fp1, "%d", &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[k][i],
&resnamec[k][i],&resnumc[k][i],&xc[k][i],&yc[k][i],&zc[k][i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc,&c212c);
/*   printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[k][i], resnamec[k][i],
resnumc[k][i], xc[k][i], yc[k][i], zc[k][i]); */  }
else continue;
}
   fclose(fp1);
}
/********************Search for NH /Calc dist / print result ****************/
for(k=0;k<no;k++) {
 for (i=0; i<xyzc[k]; i++) {
     if(strcmp(atnamec[k][i],attest1)==0 && resnumc[k][i]==resnotest1) {
  /*   printf("found first match \n");*/
       for(j=0; j<xyzc[k]; j++){
    if(strcmp(atnamec[k][j],attest2)==0 && resnumc[k][j]==resnotest2) {
  /*    printf("found second match \n");*/
```

```
        a=(xc[k][i]-xc[k][j])*(xc[k][i]-xc[k][j]);
        b=(yc[k][i]-yc[k][j])*(yc[k][i]-yc[k][j]);
        cd=(zc[k][i]-zc[k][j])*(zc[k][i]-zc[k][j]);
        dd=a+b+cd;
   /* printf("Did basic calc up to dd \n");*/

        if(dd<=0.0) d[k]=0.0;
        else d[k]=sqrtf(a+b+cd);
        if(xc[k][i]==0.0 || xc[k][j]==0.0) d[k]=0.0;
/*    printf("Calc distance complete %d   %2.2f \n",k,d[k]);*/
/*if(d[k] < 7.0) { */
/*printf("start print of residue info");*/
printf(" %s %d %s",resnamec[k][i],resnumc[k][i],atnamec[k][i]);
printf("  %s %d %s    %2.2f",resnamec[k][j], resnumc[k][j],atnamec[k][j],d[k]);
printf(" mol no %d \n",k+1);
} }
}}
}
/*printf(" Start the sum process  \n"); */
sum=0.0;
for(j=0;j<no;j++) sum=sum+d[j];
/*printf(" Calc ave  \n");*/
ave=sum/no;
diff2=0.0;
for(k=0;k<no;k++) diff2=((ave-d[k])*(ave-d[k]))+diff2;
/*printf( "diff2= %2.2f  ave= %2.2f  \n",diff2,ave);*/
sd=sqrtf(diff2/(no-1));

printf(" ave = %2.2f sd=  %2.2f \n",ave,sd);

fclose(fp3);
}
```

hbond.c

```
/*              hbond.c            */
/*  Program to calculate angle and N-O H-O distances for a structure */
/*  Input file prompted for -  */
/*  The standard angle to measure to identify H-bonds is:    */
/*   O---H-N which ideally is linear (180 deg) but usually   */
/*   H-bonds are identified by an angle >= 120 deg          */
/*   Typical distance criteria between doner H and acceptor  */
/*   N or O is <= 3.3 A or 3.5 A                            */
/*   This program just outputs the dist and angle without    */
/*   Identifying pairs that satisfy the H-bond criteria      */
/*   hbondrev2.c applies the test for H-bonding and gives    */
/*   the percentage of structures where it is satisfied -FFD  */
/* to compile use cc dist.c -lm (-lm to include math utility)*/

/* AUTHOR: Jeff Pelton */

#include <math.h>
```

```
#include <stdio.h>
#include <string.h>

void main(int argc, char *argv[])
{
FILE *fp1, *fp3;
char *p;
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],filename1[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
float nx,ny,nz,hnx,hny,hnz,ox,oy,oz,hhnx,hhny,hhnz;
int check[50];
float hnox,hnoy,hnoz,nox,noy,noz,maghn[50],magho[50],magno[50];
float dot,theta[50];
float maghosum,magnosum,stdevno,stdevho,maghodiff2,magnodiff2,aveno,aveho;
/*char fileno[2][2]={"1","2"};*/
char
fileno[24][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24"};
if(argc != 3) {printf("Usage: dist res_number(for HN) res_num(for O) ");
exit(1); }
strcpy(pdb,".pdb");
resnotest1=atoi(argv[1]);
resnotest2=atoi(argv[2]);
printf("Looking for %d  HN  %d  O\n",atoi(argv[1]),atoi(argv[2]));

/*  **** Get basic information interactively  ***** */
printf("enter Base filename -ex. accept_  :\n");
/**  **/
p=gets(s);
if(p) strcpy(filename,s);
strcpy(filename1,filename);
printf("enter number of files - default 24: \n");
p=gets(s);
if(p) no=atoi(s);
printf("number of files is: %d / Base file name is %s  \n",no,filename);
/***          Open file for output    ******************************** */
/*  if ((fp3=fopen("nhch3dist.output", "w")) == NULL) {
   printf("Error opening nhch3dist.output\n");
   exit(1); } */
/***************  Read in xyz coordinates for structures ******************/
for(k=0;k<no;k++) {
 /* add="1";*/
  strcat(filename,fileno[k]);
  strcat(filename,pdb);
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening file\n");
   exit(1); }
```

```
/*printf("Opened xyz coordinate file %s - reading coordinates \n",filename);*/
  strcpy(filename,filename1);
  i=-1;
  xyzc[k]=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc[k]=xyzc[k]+1;
   fscanf(fp1,"%s",&atomc);
  if(strcmp(atomc,"ATOM")==0) {
   fscanf(fp1, "%d", &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[k][i],
&resnamec[k][i],&resnumc[k][i],&xc[k][i],&yc[k][i],&zc[k][i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc,&c212c);
/*   printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[k][i], resnamec[k][i],
resnumc[k][i], xc[k][i], yc[k][i], zc[k][i]); */  }
else continue;
}
   fclose(fp1);
}
/********************Search for NH /Calc dist / print result ****************/
magnosum=0.0;
maghosum=0.0;
for(k=0;k<no;k++) { check[k]=0;
/*  ***** Locate coordinates  ***** */
 for (i=0; i<xyzc[k]; i++) {
     if(strcmp(atnamec[k][i],"N")==0 && resnumc[k][i]==resnotest1) {
     nx=xc[k][i]; ny=yc[k][i]; nz=zc[k][i]; check[k]=check[k]+1;}
     if(strcmp(atnamec[k][i],"HN")==0 && resnumc[k][i]==resnotest1) {
     hnx=xc[k][i]; hny=yc[k][i]; hnz=zc[k][i]; check[k]=check[k]+1;}
     if(strcmp(atnamec[k][i],"O")==0 && resnumc[k][i]==resnotest2) {
     ox=xc[k][i]; oy=yc[k][i]; oz=zc[k][i]; check[k]=check[k]+1;}
  }
 if(check[k]==3) {
    /*  Calculate vectors if three matches H N O were found */
   /*  Calc H - N vector */
  hhnx=(hnx-nx); hhny=(hny-ny); hhnz=(hnz-nz);
   /*  Calc HN - O vector */
  hnox=(hnx-ox); hnoy=(hny-oy); hnoz=(hnz-oz);
   /*  Calc N - O vector */
  nox=(nx-ox); noy=(ny-oy); noz=(nz-oz);
  /* Calc magnitudes  */
  maghn[k]=sqrt((hhnx*hhnx) + (hhny*hhny) + (hhnz*hhnz));
  magho[k]=sqrt((hnox*hnox) + (hnoy*hnoy) + (hnoz*hnoz));
  magno[k]=sqrt((nox*nox + noy*noy + noz*noz));
  /* Calc dot product */
  dot=hhnx*hnox + hhny*hnoy + hhnz*hnoz;
  /* Calc theta angle between o - h - n */
  theta[k]=(360.0*0.1592)*(acos(dot/(maghn[k]*magho[k])));
  /* Sum distances to get an average */
  magnosum=magnosum + magno[k];
  maghosum=maghosum + magho[k];
}}
```

```
  /* Do calcs to get ave and std dev for H - O and N- O distances */
aveno=magnosum/no;
aveho=maghosum/no;
magnodiff2=0.0;
maghodiff2=0.0;
  for(k=0;k<no;k++) {

  magnodiff2=((aveno-magno[k])*(aveno-magno[k])) + magnodiff2;
  maghodiff2=((aveho-magho[k])*(aveho-magho[k])) + maghodiff2; }

  stdevno=sqrt(magnodiff2/(no-1)); stdevho=sqrt(maghodiff2/(no-1));

  printf(" Results for %d HN to %d O\n",resnotest1,resnotest2);
  for(k=0;k<no;k++){
   printf(" Angle\t %3.2f  \tH-O Dist\t %2.2f \tN-O Dist\t
%2.2f\n",theta[k],magho[k],magno[k]);
              }
   printf("Ave H-O dist %2.2f +/- %2.2f\n",aveho,stdevho);
   printf("Ave N-O dist %2.2f +/- %2.2f\n",aveno,stdevno);
}
```

hbondrev2.c

```
/*            hbondrev2.c        */
/* Program to calculate angle between the vectors N->H and C-O */
/* as well as the angle between the vectors H->N and O->N */
/* and N-O and H-O distances for a structure */
/* and determine number of structures satisfying H-bond criteria */
/* which are distN-O <=3.3 A and angleNH-CO >= 0.0 and angleHN-ON >=120 */
/* Input file prompted for - */
/* Output to file includes ave distance and percent H-bonded only */
/* Output to a file using the command >>filenam.out */
/* to compile use cc dist.c -lm (-lm to include math utility)*/

/* AUTHOR: Fred Damberger */

#include <math.h>
#include <stdio.h>
#include <string.h>

void main(int argc, char *argv[])
{
FILE *fp1, *fp3;
char *p;
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],filename1[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
float  nx,ny,nz,hnx,hny,hnz,ox,oy,oz,cx,cy,cz,hhnx,hhny,hhnz;
int check[50];
```

```
float
hnox,hnoy,hnoz,nox,noy,noz,cox,coy,coz,maghn[50],magho[50],magno[50],magco[50];
float dot,theta[50],thetb[50];
float
maghosum,magnosum,stdevno,stdevho,maghodiff2,magnodiff2,aveno,aveho,numhb,pct
hb,numstruc;
/*char fileno[2][2]={"1","2"};*/
char
fileno[40][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24","25","26","27","28","29","30","31","32","3
3","34","35","36","37","38","39","40"};
if(argc != 3) {printf("Usage: dist res_number(for HN) res_num(for O) ");
exit(1); }
strcpy(pdb,".pdb");
resnotest1=atoi(argv[1]);
resnotest2=atoi(argv[2]);
/* printf("Looking for %d  HN  %d  O\n",atoi(argv[1]),atoi(argv[2])); */

/*  **** Get basic information interactively  ***** */
/* printf("enter Base filename -ex. accept_  :\n"); */
/**  **/
p=gets(s);
if(p) strcpy(filename,s);
strcpy(filename1,filename);
/* printf("enter number of files - default 24: \n"); */
p=gets(s);
if(p) no=atoi(s);
/* printf("number of files is: %d / Base file name is %s  \n",no,filename); */
/***            Open file for output    ********************************  */
/*  if ((fp3=fopen("nhch3dist.output", "w")) == NULL) {
   printf("Error opening nhch3dist.output\n");
   exit(1); } */
/***************  Read in xyz coordinates for structures ******************/
for(k=0;k<no;k++) {
 /* add="l";*/
  strcat(filename,fileno[k]);
  strcat(filename,pdb);
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening file\n");
   printf(filename);
   exit(1); }
/*printf("Opened xyz coordinate file %s - reading coordinates \n",filename);*/
  strcpy(filename,filename1);
  i=-1;
  xyzc[k]=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc[k]=xyzc[k]+1;
   fscanf(fp1,"%s",&atomc);
  if(strcmp(atomc,"ATOM")==0) {
   fscanf(fp1, "%d", &atomnumc);
```

```
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[k][i],
&resnamec[k][i],&resnumc[k][i],&xc[k][i],&yc[k][i],&zc[k][i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc,&c212c);
/*    printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[k][i], resnamec[k][i],
resnumc[k][i], xc[k][i], yc[k][i], zc[k][i]); */  }
else continue;
}
   fclose(fp1);
}
/********************Search for NH /Calc dist / print result ****************/
magnosum=0.0;
maghosum=0.0;
numhb=0.0;
for(k=0;k<no;k++) { check[k]=0;
/*  ***** Locate coordinates  ***** */
 for (i=0; i<xyzc[k]; i++) {
     if(strcmp(atnamec[k][i],"N")==0 && resnumc[k][i]==resnotest1) {
     nx=xc[k][i]; ny=yc[k][i]; nz=zc[k][i]; check[k]=check[k]+1;}
     if(strcmp(atnamec[k][i],"HN")==0 && resnumc[k][i]==resnotest1) {
     hnx=xc[k][i]; hny=yc[k][i]; hnz=zc[k][i]; check[k]=check[k]+1;}
     if(strcmp(atnamec[k][i],"O")==0 && resnumc[k][i]==resnotest2) {
     ox=xc[k][i]; oy=yc[k][i]; oz=zc[k][i]; check[k]=check[k]+1;}
     if(strcmp(atnamec[k][i],"C")==0 && resnumc[k][i]==resnotest2) {
     cx=xc[k][i]; cy=yc[k][i]; cz=zc[k][i]; check[k]=check[k]+1;}
  }
 if(check[k]==4) {
    /*  Calculate vectors if four matches H N O C were found */
   /*  Calc H - N vector */
  hhnx=(hnx-nx); hhny=(hny-ny); hhnz=(hnz-nz);
   /*  Calc HN - O vector */
  hnox=(hnx-ox); hnoy=(hny-oy); hnoz=(hnz-oz);
   /*  Calc N - O vector */
  nox=(nx-ox); noy=(ny-oy); noz=(nz-oz);
   /*  Calc O - C vector */
  cox=(ox-cx); coy=(oy-cy); coz=(oz-cz);
  /* Calc magnitudes  */
  maghn[k]=sqrt((hhnx*hhnx) + (hhny*hhny) + (hhnz*hhnz));
  magho[k]=sqrt((hnox*hnox) + (hnoy*hnoy) + (hnoz*hnoz));
  magno[k]=sqrt((nox*nox + noy*noy + noz*noz));
  magco[k]=sqrt((cox*cox + coy*coy + coz*coz));
  /* Calc dot product */
  dot=hhnx*hnox + hhny*hnoy + hhnz*hnoz;
  /* Calc theta angle between o - h - n */
  theta[k]=(360.0*0.1592)*(acos(dot/(maghn[k]*magho[k])));
  /* Calc dot product */
  dot=hhnx*cox + hhny*coy + hhnz*coz;
  /* Calc theta angle between h->n and o->c */
  thetb[k]=(360.0*0.1592)*(acos(dot/(maghn[k]*magco[k])));
  /* Calc number of structures which have H-bond */
  /* Based on criteria of angle >=120 and dist N-O <=3.3 */
  if(magno[k]<=3.3 && theta[k]>=120.0 && thetb[k]>=0.0) numhb=numhb +1.0;
  /* Sum distances to get an average */
```

```
    magnosum=magnosum + magno[k];
    maghosum=maghosum + magho[k];
  }}
   /* Do calcs to get ave and std dev for H - O and N- O distances */
   /* and precentage of structures satisfying H-bond criteria */
aveno=magnosum/no;
aveho=maghosum/no;
pcthb=numhb*100.0/no;
numstruc=no;
magnodiff2=0.0;
maghodiff2=0.0;
  for(k=0;k<no;k++) {

  magnodiff2=((aveno-magno[k])*(aveno-magno[k])) + magnodiff2;
  maghodiff2=((aveho-magho[k])*(aveho-magho[k])) + maghodiff2; }

  stdevno=sqrt(magnodiff2/(no-1)); stdevho=sqrt(maghodiff2/(no-1));

  printf(" Results for %d HN to %d O\n",resnotest1,resnotest2);
/*  for(k=0;k<no;k++){
   printf(" HNO\t %3.1f \tHNOC\t %3.1f \tdHO \t %2.2f \dNO \t
%2.2f\n",theta[k],thetb[k],magho[k],magno[k]);
              } */
   printf("  \n");
   printf("Ave H-O dist %2.2f +/- %2.2f\n",aveho,stdevho);
   printf("Ave N-O dist %2.2f +/- %2.2f\n",aveno,stdevno);
   printf("Number H-bonded %2.0f of %3.0f\n",numhb,numstruc);
   printf("  \n");
   printf("Percent H-bonded %3.1f\n",pcthb);
   printf("                        \n");
   printf("_____________________________________\n");
}
```

hbondrev2.com

FILE: hbondrev2.com
a command file which executes
hbondrev2.c many times for all H-bond pairs
listed and writes the results to a file: hbondrev.out
run in directory containing both compiled version of hbondrev2.c (hbondrev2) and structure files (e.g. accept_) and
input file: input
Below is contents of input file:

---

accept_
33

Below is contents of hbondrev2.com:

```
hbondrev2 6 3 <input >>hbondrev.out
hbondrev2 7 3 <input >>hbondrev.out
hbondrev2 7 4 <input >>hbondrev.out
hbondrev2 8 4 <input >>hbondrev.out
hbondrev2 8 5 <input >>hbondrev.out
hbondrev2 9 5 <input >>hbondrev.out
hbondrev2 9 6 <input >>hbondrev.out
hbondrev2 13 9 <input >>hbondrev.out
hbondrev2 13 10 <input >>hbondrev.out
hbondrev2 14 10 <input >>hbondrev.out
hbondrev2 14 11 <input >>hbondrev.out
hbondrev2 15 11 <input >>hbondrev.out
hbondrev2 15 12 <input >>hbondrev.out
hbondrev2 17 13 <input >>hbondrev.out
hbondrev2 17 14 <input >>hbondrev.out
hbondrev2 19 15 <input >>hbondrev.out
hbondrev2 19 16 <input >>hbondrev.out
hbondrev2 21 17 <input >>hbondrev.out
hbondrev2 21 18 <input >>hbondrev.out
hbondrev2 25 30 <input >>hbondrev.out
hbondrev2 29 88 <input >>hbondrev.out
hbondrev2 31 86 <input >>hbondrev.out
hbondrev2 32 23 <input >>hbondrev.out
hbondrev2 33 84 <input >>hbondrev.out
hbondrev2 38 34 <input >>hbondrev.out
hbondrev2 38 35 <input >>hbondrev.out
hbondrev2 39 35 <input >>hbondrev.out
hbondrev2 39 36 <input >>hbondrev.out
hbondrev2 42 38 <input >>hbondrev.out
hbondrev2 42 39 <input >>hbondrev.out
hbondrev2 43 39 <input >>hbondrev.out
hbondrev2 43 40 <input >>hbondrev.out
hbondrev2 44 39 <input >>hbondrev.out
hbondrev2 44 40 <input >>hbondrev.out
hbondrev2 44 41 <input >>hbondrev.out
hbondrev2 46 42 <input >>hbondrev.out
hbondrev2 46 43 <input >>hbondrev.out
hbondrev2 47 43 <input >>hbondrev.out
hbondrev2 47 44 <input >>hbondrev.out
hbondrev2 48 44 <input >>hbondrev.out
hbondrev2 48 45 <input >>hbondrev.out
hbondrev2 54 49 <input >>hbondrev.out
hbondrev2 54 50 <input >>hbondrev.out
hbondrev2 54 51 <input >>hbondrev.out
hbondrev2 54 52 <input >>hbondrev.out
hbondrev2 55 51 <input >>hbondrev.out
hbondrev2 55 52 <input >>hbondrev.out
hbondrev2 56 52 <input >>hbondrev.out
hbondrev2 56 53 <input >>hbondrev.out
hbondrev2 57 53 <input >>hbondrev.out
hbondrev2 57 54 <input >>hbondrev.out
hbondrev2 58 54 <input >>hbondrev.out
hbondrev2 58 55 <input >>hbondrev.out
hbondrev2 59 55 <input >>hbondrev.out
hbondrev2 59 56 <input >>hbondrev.out
hbondrev2 60 56 <input >>hbondrev.out
hbondrev2 60 57 <input >>hbondrev.out
hbondrev2 61 57 <input >>hbondrev.out
hbondrev2 61 58 <input >>hbondrev.out
hbondrev2 62 58 <input >>hbondrev.out
hbondrev2 62 59 <input >>hbondrev.out
hbondrev2 63 59 <input >>hbondrev.out
hbondrev2 63 60 <input >>hbondrev.out
hbondrev2 64 60 <input >>hbondrev.out
hbondrev2 64 61 <input >>hbondrev.out
hbondrev2 65 60 <input >>hbondrev.out
hbondrev2 65 61 <input >>hbondrev.out
hbondrev2 65 62 <input >>hbondrev.out
hbondrev2 66 87 <input >>hbondrev.out
hbondrev2 68 85 <input >>hbondrev.out
hbondrev2 84 33 <input >>hbondrev.out
hbondrev2 86 31 <input >>hbondrev.out
hbondrev2 87 66 <input >>hbondrev.out
```

Below is contents of output file (hbondrev.out):

Results for 6 HN to 3 O

Ave H-O dist 2.78 +/- 0.30
Ave N-O dist 3.21 +/- 0.22
Number H-bonded 5 of 33

Percent H-bonded 15.2

Results for 7 HN to 3 O

Ave H-O dist 2.29 +/- 0.07
Ave N-O dist 3.22 +/- 0.09
Number H-bonded 27 of 33

Percent H-bonded 81.8

Results for 7 HN to 4 O

Ave H-O dist 2.76 +/- 0.36
Ave N-O dist 3.26 +/- 0.32
Number H-bonded 6 of 33

Percent H-bonded 18.2

Results for 8 HN to 4 O

Ave H-O dist 2.15 +/- 0.07
Ave N-O dist 3.05 +/- 0.12
Number H-bonded 32 of 33

Percent H-bonded 97.0

Results for 8 HN to 5 O

Ave H-O dist 2.50 +/- 0.35
Ave N-O dist 2.99 +/- 0.25
Number H-bonded 5 of 33

Percent H-bonded 15.2

Results for 9 HN to 5 O

Ave H-O dist 2.19 +/- 0.08
Ave N-O dist 2.94 +/- 0.10
Number H-bonded 31 of 33

Percent H-bonded 93.9

Results for 9 HN to 6 O

Ave H-O dist 2.32 +/- 0.18
Ave N-O dist 3.00 +/- 0.19
Number H-bonded 25 of 33

Percent H-bonded 75.8

Results for 13 HN to 9 O

Ave H-O dist 2.12 +/- 0.08
Ave N-O dist 2.98 +/- 0.10
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 13 HN to 10 O

Ave H-O dist 4.03 +/- 0.57
Ave N-O dist 4.26 +/- 0.55
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 14 HN to 10 O

Ave H-O dist 2.62 +/- 0.44
Ave N-O dist 3.28 +/- 0.40
Number H-bonded 14 of 33

Percent H-bonded 42.4

Results for 14 HN to 11 O

Ave H-O dist 3.51 +/- 0.53
Ave N-O dist 4.16 +/- 0.43
Number H-bonded 1 of 33

Percent H-bonded 3.0

Results for 15 HN to 11 O

Ave H-O dist 4.16 +/- 0.70
Ave N-O dist 4.67 +/- 0.79
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 15 HN to 12 O

Ave H-O dist 2.60 +/- 0.51
Ave N-O dist 3.43 +/- 0.45
Number H-bonded 13 of 33

Percent H-bonded 39.4

Results for 17 HN to 13 O

Ave H-O dist 7.04 +/- 0.59
Ave N-O dist 7.56 +/- 0.50
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 17 HN to 14 O

Ave H-O dist 5.73 +/- 0.67
Ave N-O dist 6.17 +/- 0.58
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 19 HN to 15 O

Ave H-O dist 3.32 +/- 0.55
Ave N-O dist 3.37 +/- 0.54
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 19 HN to 16 O

Ave H-O dist 3.31 +/- 0.73
Ave N-O dist 4.11 +/- 0.78
Number H-bonded 1 of 33

Percent H-bonded 3.0

Results for 21 HN to 17 O

Ave H-O dist 5.85 +/- 0.60
Ave N-O dist 6.52 +/- 0.58
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 21 HN to 18 O

Ave H-O dist 2.21 +/- 0.08
Ave N-O dist 3.07 +/- 0.13
Number H-bonded 30 of 33

Percent H-bonded 90.9

Results for 25 HN to 30 O

Ave H-O dist 2.19 +/- 0.08
Ave N-O dist 2.98 +/- 0.12
Number H-bonded 30 of 33

Percent H-bonded 90.9

Results for 29 HN to 88 O

Ave H-O dist 8.51 +/- 1.66
Ave N-O dist 8.50 +/- 1.51
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 31 HN to 86 O

Ave H-O dist 2.24 +/- 0.09
Ave N-O dist 3.12 +/- 0.12
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 32 HN to 23 O

Ave H-O dist 2.22 +/- 0.08
Ave N-O dist 3.14 +/- 0.09
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 33 HN to 84 O

Ave H-O dist 2.22 +/- 0.09
Ave N-O dist 3.03 +/- 0.16
Number H-bonded 28 of 33

Percent H-bonded 84.8

Results for 38 HN to 34 O

Ave H-O dist 6.61 +/- 0.36
Ave N-O dist 7.47 +/- 0.27
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 38 HN to 35 O

Ave H-O dist 3.03 +/- 0.61
Ave N-O dist 3.63 +/- 0.37
Number H-bonded 7 of 33

Percent H-bonded 21.2

Results for 39 HN to 35 O

Ave H-O dist 2.42 +/- 0.24
Ave N-O dist 3.31 +/- 0.23
Number H-bonded 15 of 33

Percent H-bonded 45.5

Results for 39 HN to 36 O

Ave H-O dist 4.01 +/- 0.46
Ave N-O dist 4.16 +/- 0.47
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 42 HN to 38 O

Ave H-O dist 2.31 +/- 0.04
Ave N-O dist 2.84 +/- 0.10
Number H-bonded 5 of 33

Percent H-bonded 15.2

Results for 42 HN to 39 O

Ave H-O dist 2.74 +/- 0.26
Ave N-O dist 3.62 +/- 0.25
Number H-bonded 6 of 33

Percent H-bonded 18.2

Results for 43 HN to 39 O

Ave H-O dist 2.07 +/- 0.02
Ave N-O dist 2.85 +/- 0.07
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 43 HN to 40 O

Ave H-O dist 3.67 +/- 0.24
Ave N-O dist 4.00 +/- 0.06
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 44 HN to 39 O

Ave H-O dist 2.44 +/- 0.12
Ave N-O dist 3.11 +/- 0.11
Number H-bonded 24 of 33

Percent H-bonded 72.7

Results for 44 HN to 40 O

Ave H-O dist 2.36 +/- 0.02
Ave N-O dist 3.11 +/- 0.03
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 44 HN to 41 O

Ave H-O dist 5.02 +/- 0.23

Ave N-O dist 5.78 +/- 0.26
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 46 HN to 42 O

Ave H-O dist 3.71 +/- 0.50
Ave N-O dist 4.21 +/- 0.49
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 46 HN to 43 O

Ave H-O dist 2.29 +/- 0.05
Ave N-O dist 3.12 +/- 0.08
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 47 HN to 43 O

Ave H-O dist 3.28 +/- 0.35
Ave N-O dist 3.73 +/- 0.23
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 47 HN to 44 O

Ave H-O dist 2.54 +/- 0.23
Ave N-O dist 3.45 +/- 0.23
Number H-bonded 8 of 33

Percent H-bonded 24.2

Results for 48 HN to 44 O

Ave H-O dist 2.18 +/- 0.10
Ave N-O dist 3.08 +/- 0.12
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 48 HN to 45 O

Ave H-O dist 4.01 +/- 0.26
Ave N-O dist 4.45 +/- 0.19
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 54 HN to 49 O

Ave H-O dist 10.94 +/- 0.72
Ave N-O dist 11.25 +/- 0.67
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 54 HN to 50 O

Ave H-O dist 8.88 +/- 0.79
Ave N-O dist 9.20 +/- 0.63
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 54 HN to 51 O

Ave H-O dist 6.11 +/- 0.79
Ave N-O dist 6.45 +/- 0.89
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 54 HN to 52 O

Ave H-O dist 3.30 +/- 0.61
Ave N-O dist 3.14 +/- 0.53
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 55 HN to 51 O

Ave H-O dist 5.48 +/- 1.22
Ave N-O dist 6.26 +/- 1.23
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 55 HN to 52 O

Ave H-O dist 2.65 +/- 0.86
Ave N-O dist 3.19 +/- 0.72
Number H-bonded 11 of 33

Percent H-bonded 33.3

Results for 56 HN to 52 O

Ave H-O dist 2.52 +/- 0.36
Ave N-O dist 3.40 +/- 0.30
Number H-bonded 15 of 33

Percent H-bonded 45.5

Results for 56 HN to 53 O

Ave H-O dist 2.76 +/- 0.42
Ave N-O dist 3.12 +/- 0.34
Number H-bonded 1 of 33

Percent H-bonded 3.0

Results for 57 HN to 53 O

Ave H-O dist 2.16 +/- 0.10
Ave N-O dist 3.10 +/- 0.12
Number H-bonded 30 of 33

Percent H-bonded 90.9

Results for 57 HN to 54 O

Ave H-O dist 2.86 +/- 0.44
Ave N-O dist 3.31 +/- 0.34
Number H-bonded 2 of 33

Percent H-bonded 6.1

Results for 58 HN to 54 O

Ave H-O dist 2.16 +/- 0.08
Ave N-O dist 3.07 +/- 0.10
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 58 HN to 55 O

Ave H-O dist 2.46 +/- 0.30
Ave N-O dist 2.95 +/- 0.28
Number H-bonded 4 of 33

Percent H-bonded 12.1

Results for 59 HN to 55 O

Ave H-O dist 2.14 +/- 0.07
Ave N-O dist 3.08 +/- 0.07
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 59 HN to 56 O

Ave H-O dist 2.88 +/- 0.31
Ave N-O dist 3.32 +/- 0.24
Number H-bonded 1 of 33

Percent H-bonded 3.0

Results for 60 HN to 56 O

Ave H-O dist 2.26 +/- 0.08
Ave N-O dist 3.22 +/- 0.08
Number H-bonded 31 of 33

Percent H-bonded 93.9

Results for 60 HN to 57 O

Ave H-O dist 2.72 +/- 0.32
Ave N-O dist 3.14 +/- 0.29
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 61 HN to 57 O

Ave H-O dist 2.14 +/- 0.06
Ave N-O dist 3.05 +/- 0.07

Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 61 HN to 58 O

Ave H-O dist 2.42 +/- 0.24
Ave N-O dist 2.97 +/- 0.21
Number H-bonded 5 of 33

Percent H-bonded 15.2

Results for 62 HN to 58 O

Ave H-O dist 2.22 +/- 0.09
Ave N-O dist 3.03 +/- 0.10
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 62 HN to 59 O

Ave H-O dist 2.23 +/- 0.20
Ave N-O dist 2.91 +/- 0.13
Number H-bonded 25 of 33

Percent H-bonded 75.8

Results for 63 HN to 59 O

Ave H-O dist 2.60 +/- 0.33
Ave N-O dist 3.20 +/- 0.21
Number H-bonded 16 of 33

Percent H-bonded 48.5

Results for 63 HN to 60 O

Ave H-O dist 2.34 +/- 0.20
Ave N-O dist 3.19 +/- 0.19
Number H-bonded 22 of 33

Percent H-bonded 66.7

Results for 64 HN to 60 O

Ave H-O dist 2.26 +/- 0.23
Ave N-O dist 3.06 +/- 0.14
Number H-bonded 28 of 33

Percent H-bonded 84.8

Results for 64 HN to 61 O

Ave H-O dist 3.11 +/- 0.48
Ave N-O dist 3.76 +/- 0.36
Number H-bonded 5 of 33

Percent H-bonded 15.2

Results for 65 HN to 60 O

Ave H-O dist 2.64 +/- 0.39
Ave N-O dist 3.03 +/- 0.20
Number H-bonded 6 of 33

Percent H-bonded 18.2

Results for 65 HN to 61 O

Ave H-O dist 4.53 +/- 0.74
Ave N-O dist 4.99 +/- 0.41
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 65 HN to 62 O

Ave H-O dist 6.24 +/- 0.41
Ave N-O dist 6.79 +/- 0.31
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 66 HN to 87 O

Ave H-O dist 2.24 +/- 0.08
Ave N-O dist 3.09 +/- 0.14
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 68 HN to 85 O

Ave H-O dist 4.06 +/- 0.95
Ave N-O dist 4.51 +/- 0.64
Number H-bonded 0 of 33

Percent H-bonded 0.0

Results for 84 HN to 33 O

Ave H-O dist 2.67 +/- 0.44
Ave N-O dist 3.45 +/- 0.30
Number H-bonded 12 of 33

Percent H-bonded 36.4

Results for 86 HN to 31 O

Ave H-O dist 2.20 +/- 0.09
Ave N-O dist 3.06 +/- 0.14
Number H-bonded 33 of 33

Percent H-bonded 100.0

Results for 87 HN to 66 O

Ave H-O dist 2.50 +/- 0.48
Ave N-O dist 3.27 +/- 0.46
Number H-bonded 18 of 33

Percent H-bonded 54.5

---

energyrmsd.c

```
/*energyrmsd.c*/
/*  Program to calculate  ave + sd for energy bond angle etc */
/*  for set of structures */
/*  The program reads the headers for structure files generated */
/*  by X-PLOR which contain refinement statistics for each structure */
/*  Input: program prompts user for file prefix and number of structures */
/*        (see dist.c for description of format) */
/* to compile use cc dist.c -lm (-lm to include math utility)*/


/* AUTHOR: Jeff Pelton */


#include <math.h>
#include <stdio.h>
#include <string.h>

void main()
{
FILE *fp1, *fp3;
char *p;
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5],string[1000];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],filename1[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
float bond[50],angle[50],improper[50],vdw[50],noe[50],cdih[50];
float rbond[50],rangle[50],rimproper[50],rnoe[50],rcdih[50],tot[50];
int noeviol[50],diheviol[50];
float sbond,sangle,simproper,svdw,snoe,scdih;
float srbond,srangle,srimproper,srnoe,srcdih,stot;
float abond,aangle,aimproper,avdw,anoe,acdih;
float arbond,arangle,arimproper,arnoe,arcdih,atot;
float dbond,dangle,dimproper,dvdw,dnoe,dcdih;
float drbond,drangle,drimproper,drnoe,drcdih,dtot;
float sdbond,sdangle,sdimproper,sdvdw,sdnoe,sdcdih;
float sdrbond,sdrangle,sdrimproper,sdrnoe,sdrcdih,sdtot;
/*char fileno[2][2]={"1","2"};*/
char
fileno[40][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24","25","26","27","28","29","30","31","32","3
3","34","35","36","37","38","39","40"};
strcpy(pdb,".pdb");

/*  **** Get basic information interactively  ***** */
printf("enter Base filename -ex. accept_  :\n");
/**  **/
```

```
p=gets(s);
if(p) strcpy(filename,s);
strcpy(filename1,filename);
printf("enter number of files - default 24: \n");
p=gets(s);
if(p) no=atoi(s);
printf("number of files is: %d / Base file name is %s  \n",no,filename);
/***            Open file for output    ******************************** */
   if ((fp3=fopen("energyrmsd.output", "w")) == NULL) {
    printf("Error opening energyrmsd.output\n");
    exit(1); }
/*************** Read in headers ex. accept structures ******************/
for(k=0;k<no;k++) {
  /* add="1";*/
   strcat(filename,fileno[k]);
   strcat(filename,pdb);
   if ((fp1=fopen(filename, "r")) == NULL) {
    printf("Error opening file\n");
    exit(1); }
/*printf("Opened xyz coordinate file %s - reading coordinates \n",filename);*/
   strcpy(filename,filename1);
   i=-1;
   xyzc[k]=-1;
 while(i<10) {
    xyzc[k]=xyzc[k]+1;
    fscanf(fp1,"%s",string);
 if(strcmp(string,"energies:")==0) {
    fscanf(fp1, "%f", &tot[k]);
    fscanf(fp1, "%f%f%f%f%f%f", &bond[k],
&angle[k],&improper[k],&vdw[k],&noe[k],&cdih[k]); }
  if(strcmp(string,"rms-d:")==0) {

fscanf(fp1,"%f%f%f%f%f%f",&rbond[k],&rangle[k],&rimproper[k],&rnoe[k],&rcdih[k]
); }
  if(strcmp(string,"violations.:")==0) { i=11;
   fscanf(fp1,"%d%d",&noeviol[k],&diheviol[k]); fclose(fp1);
 fprintf(fp3,"Structure %d\n",k+1);
 fprintf(fp3,"Energies - tot, bond, angle, improper, vdw, noe, cdih\n");
  fprintf(fp3,"rmsds - bond, angle, improper, noe, cdih\n\n");
  fprintf(fp3,"%2.3f \t%2.3f %2.3f %2.3f %2.3f \t%2.3f   %2.3f \t%d  \t %d\n\n", tot[k],
bond[k], angle[k], improper[k], vdw[k], noe[k], cdih[k],noeviol[k],diheviol[k]);
  fprintf(fp3,"%2.4f %2.4f %2.4f %2.4f
%2.4f\n\n",rbond[k],rangle[k],rimproper[k],rnoe[k],rcdih[k]);

} }

}
/********************Calc ave + rmsds - print result ****************/
/* Zero terms */
sbond=0.0; sangle=0.0; simproper=0.0; svdw=0.0; snoe=0.0; scdih=0.0;
srbond=0.0; srangle=0.0; srimproper=0.0; srnoe=0.0; srcdih=0.0;
/* Sum terms - calc average */
```

```
for(k=0;k<no;k++) {  stot=stot+tot[k];
sbond=sbond+bond[k]; sangle=sangle+angle[k]; simproper=simproper+improper[k];
svdw=svdw+vdw[k]; snoe=snoe+noe[k]; scdih=scdih+cdih[k];
srbond=srbond+rbond[k]; srangle=srangle+rangle[k];
srimproper=srimproper+rimproper[k]; srnoe=srnoe+rnoe[k]; srcdih=srcdih+rcdih[k];
            }
/*printf(" Calc ave  \n");*/
atot=stot/no;
abond=sbond/no; aangle=sangle/no; aimproper=simproper/no; avdw=svdw/no;
anoe=snoe/no; acdih=scdih/no; arbond=srbond/no; arangle=srangle/no;
arimproper=srimproper/no; arnoe=srnoe/no; arcdih=srcdih/no;

/*  Calc diff squared */
for(k=0;k<no;k++) { dtot=pow((atot-tot[k]),2)+dtot;
dbond=pow((abond-bond[k]),2)+dbond; dangle=pow((aangle-angle[k]),2)+dangle;
dimproper=pow((aimproper-improper[k]),2)+dimproper;
dvdw=pow((avdw-vdw[k]),2)+dvdw; dnoe=pow((anoe-noe[k]),2)+dnoe;
dcdih=pow((acdih-cdih[k]),2)+dcdih; drbond=pow((arbond-rbond[k]),2)+drbond;
drangle=pow((arangle-rangle[k]),2)+drangle;
drimproper=pow((arimproper-rimproper[k]),2)+drimproper;
drnoe=pow((arnoe-rnoe[k]),2)+drnoe; drcdih=pow((arcdih-rcdih[k]),2)+drcdih;
            }
/*  Calc standard deviations */
sdtot=sqrtf(dtot/(no-1));
sdbond=sqrtf(dbond/(no-1));
sdangle=sqrtf(dangle/(no-1));
sdimproper=sqrtf(dimproper/(no-1));
sdvdw=sqrtf(dvdw/(no-1));
sdnoe=sqrtf(dnoe/(no-1));
sdcdih=sqrtf(dcdih/(no-1));
sdrbond=sqrtf(drbond/(no-1)); sdrangle=sqrtf(drangle/(no-1));
sdrimproper=sqrtf(drimproper/(no-1)); sdrnoe=sqrtf(drnoe/(no-1));
sdrcdih=sqrtf(drcdih/(no-1));
/*  Print average and standard deviation */
fprintf(fp3,"Averages - Energies\n");
fprintf(fp3,"TOTAL \t\t %2.5f +/- %2.5f\n",atot,sdtot);
fprintf(fp3,"BOND \t\t %2.5f +/- %2.5f\n",abond,sdbond);
fprintf(fp3,"ANGLE \t\t %2.5f +/- %2.5f\n",aangle,sdangle);
fprintf(fp3,"IMPROPER \t %2.5f +/- %2.5f\n",aimproper,sdimproper);
fprintf(fp3,"VDW \t\t %2.5f +/- %2.5f\n",avdw,sdvdw);
fprintf(fp3,"NOE \t\t %2.5f +/- %2.5f\n", anoe,sdnoe);
fprintf(fp3,"CDIH \t\t %2.5f +/- %2.5f\n", acdih,sdcdih);

fprintf(fp3,"Averages - RMSDs\n");
fprintf(fp3,"BOND \t\t %2.5f +/- %2.5f\n", arbond,sdrbond);
fprintf(fp3,"ANGLE \t\t %2.5f +/- %2.5f\n", arangle,sdrangle);
fprintf(fp3,"IMPROPER \t %2.5f +/- %2.5f\n", arimproper,sdrimproper);
fprintf(fp3,"NOE \t\t %2.5f +/- %2.5f\n",arnoe ,sdrnoe);
fprintf(fp3,"CDIH \t\t %2.5f +/- %2.5f\n",arcdih,sdrcdih);
printf("Output file - energyrmsd.output\n");
}
```

standev.c

```
/* Program to calculate standard deviation and average */
/* RMSD for a set of structures using ave.inp or xrayrms.inp */
/* Modified to that the line rms* 0.63 appears on a single line */
/* for each structure - by using display rms $result */
/* all that is required for this program is ave.log the output */
/* log file generated by X-PLOR com file ave.com (or any of its */
/* derivatives) */

/* AUTHOR: Jeff Pelton */

#include <math.h>
#include <stdio.h>
#include <string.h>
main()
{
FILE *fp,*fp1;
char rms[80],result[80],display[80],count[10];
float rmsnum[50],rmstot,ave_rms,rmsdiff2,standdev,rmst;
int i,rmsdc;
float rmsd;

/*  Open file ave.log to look for RMS deviations */
if((fp=fopen("ave.log", "r"))==NULL) {
  printf("Error opening ave.log");
  exit(1); }
/***** Find RMSDs  ******* */
i=-1;
rmsdc=-1;
while(!feof(fp)){
fscanf(fp,"%s",&rms);
if(strcmp(rms,"rms9")==0) {
i=i+1;
fscanf(fp,"%s%s%f",&result,&rms,&rmsnum[i]);
rmsdc=rmsdc+1;
    }
    }
printf("value of rmsdc is %d\n",rmsdc);
/*  Calc ave RMSD and standard deviation */
/* Calc ave RMSD */
rmstot=0.0;
for(i=0;i<rmsdc;i++) rmstot=rmstot+rmsnum[i];
 ave_rms=rmstot/rmsdc;
printf("rmstot is %f, ave_rms is %f\n",rmstot,ave_rms);
 /* Calc standard deviation */
rmsdiff2=0.0;
for(i=0;i<rmsdc;i++) rmsdiff2=rmsdiff2+((ave_rms-rmsnum[i])*(ave_rms-rmsnum[i]));
 if(rmsdiff2==0 || rmsdc==0) {
                    printf("RMSDIFF2 or RMSD ==0");
                    exit(1);}
standdev=sqrt(rmsdiff2/(rmsdc-1));
```

```
printf("The RMS deviation of %d structures is:", rmsdc);
printf("  %2.2f, +/-  %2.2f\n", ave_rms,standdev);
}
```

noeclass.c

```
/* noeclass.c */
/* Program to add up number of NOEs in each class; intra, seq, med, long */
/* and report them */
/* Input file is an noe table with H-bond constraints and commented out
/* lines removed. To do the latter open a COPY of noe_tbl with vi and type */
/* :g/!/d1 [RET]  */
/* Then write to a new file name */
/* Input file noe_tbl.1  -> output file noeclas.output  */
/*  Add EOF at end of newly created constraint list to get proper #s */

/* AUTHOR: Fred Damberger */

#include <math.h>
#include <stdio.h>
#include <string.h>

void main()
{
FILE *fp1, *fp3;
char *p;
int i, j, k, resnum1c[2000],resnum2c[2000];
float noelo[2000], noemid[2000], noehi[2000];
int noeintra, noeseq, noemed, noelong,noetot;
int noeintrabb, noeseqbb, noemedbb, noelongbb;
char assignc[10],resid[10],and[10],name[10],filename[20];
char atname1c[2000][10],atname2c[2000][10],s[80];
int count;
/*** Get name of noe_tbl file interactively ***/
printf("Enter name of noe_tbl file:\n");
p=gets(s);
if (p) strcpy(filename,s);
printf(" \n");
printf(" \n");
printf("Looking for NOEs in %s\n",filename);
/**************Read in NOEs from filename*************/
/*strcpy(filename,argv[1]);*/
if ((fp1=fopen(filename, "r")) == NULL) {
  printf("Error opening filename\n");
  exit(1); }
/* printf("Opened filename: %s - reading NOEs\n",filename); */
noetot=0;
i=-1;
while(!feof(fp1)) {
 i=i+1;
 count=count+1;
 fscanf(fp1,"%s",&assignc);
```

```
  if(strcmp(assignc,"assign")==0) {
   noetot=noetot+1;
   fscanf(fp1, "%s%d%s%s%s",&resid,&resnum1c[i],&and,&name,&atname1c[i]);
   fscanf(fp1, "%s%d%s%s%s",&resid,&resnum2c[i],&and,&name,&atname2c[i]);
   fscanf(fp1, "%f%f%f",&noelo[i],&noemid[i],&noehi[i]);
/*   printf("%d %d\n",resnum1c[i],resnum2c[i]); */ }
  else continue;
  }
fclose(fp1);
/****************** Do NOE sums *************************/
noeintra=0;
noeseq=0;
noemed=0;
noelong=0;
for (i=0; i<noetot; i++) {
  printf("%d %d\n",resnum1c[i],resnum2c[i]);
  if (resnum2c[i]-resnum1c[i]<0) printf("Wrong order for %d %s %d
%s\n",resnum1c[i],atname1c[i],resnum2c[i],atname2c[i]);
  if (resnum2c[i]-resnum1c[i]==0) noeintra=noeintra+1;
  if (resnum2c[i]-resnum1c[i]==1) noeseq=noeseq+1;
  if (resnum2c[i]-resnum1c[i]>1 && resnum2c[i]-resnum1c[i]<5) noemed=noemed+1;
  if (resnum2c[i]-resnum1c[i]>4) noelong=noelong+1;
}
/*** report results ***/
printf("Total number of NOEs        is: %d\n",noetot);
printf("_____________________________________\n");
printf("Total number of intrares NOEs is: %d\n",noeintra);
printf("Total number of sequential NOES : %d\n",noeseq);
printf("Total number of med NOEs 1<i-j<5: %d\n",noemed);
printf("Total number of long NOEs i-j>4 : %d\n",noelong);
printf(" \n");
printf(" \n");
}
```

noeclass_writeall.c

```
/* noeclass_writeall.c */
/* by FFD */
/* Program to add up number of NOEs in each class; intra,seq,med,long*/
/* This one also writes out the NOEs into separate files according to*/
/* their class */
/* and report them */
/* Input file is an noe table with H-bond constraints and commented out*/
/* lines removed. To do the latter open a COPY of noe_tbl with vi and type */
/* :g/!/d1 [RET]  */
/* Then write to a new file name */
/* Input file noe_tbl.1  -> output file noeclas.output  */

/* AUTHOR: Fred Damberger */

#include <math.h>
#include <stdio.h>
#include <string.h>
```

```
void main()
{
FILE *fp1, *fp3, *fp4,*fp5, *fp6;
char *p;
int i, j, k, resnum1c[2000],resnum2c[2000];
float noelo[2000], noemid[2000], noehi[2000];
int noeintra, noeseq, noemed, noelong,noetot;
int noeintrabb, noeseqbb, noemedbb, noelongbb;
char assignc[10],resid[10],and[10],name[10],filename[20],outfile[20];
char atname1c[2000][10],atname2c[2000][10],s[80];
char file3[10],file4[10],file5[10],file6[10];
char answ3[10],answ4[10],answ5[10],answ6[10];
int count;
/*** Get name of noe_tbl file interactively ***/
printf("Enter name of noe_tbl file:\n");
p=gets(s);
printf("I am here\n");
if (p) strcpy(filename,s);
printf("Enter name of noe output file for noe_intra:\n");
p=gets(s);
if (p) strcpy(answ3,s);
printf("Enter name of noe output file for noe_seq:\n");
p=gets(s);
if (p) strcpy(answ4,s);
printf("Enter name of noe output file for noe_med:\n");
p=gets(s);
if (p) strcpy(answ5,s);
printf("Enter name of noe output file for noe_long:\n");
p=gets(s);
if (p) strcpy(answ6,s);
printf("Output file will be %s\n",answ3);
printf("Output file will be %s\n",answ4);
printf("Output file will be %s\n",answ5);
printf("Output file will be %s\n",answ6);
strcpy(file3,answ3);
strcpy(file4,answ4);
strcpy(file5,answ5);
strcpy(file6,answ6);
printf(" \n");
printf(" \n");
printf("Looking for NOEs in %s\n",filename);
/***************Read in NOEs from filename*************/
/*strcpy(filename,argv[1]);*/
if ((fp1=fopen(filename, "r")) == NULL) {
  printf("Error opening filename\n");
  exit(1); }
/*printf("Opening outputfiles %s\n",noeintra);*/

if ((fp3=fopen(file3,"w"))==NULL) {
  printf("Error opening outputfile\n");
  exit(3); }
printf("Test writing to outputfile\n");
```

```
if ((fp4=fopen(file4,"w"))==NULL) {
  printf("Error opening outputfile\n");
  exit(3); }
printf("Test writing to outputfile\n");
if ((fp5=fopen(file5,"w"))==NULL) {
  printf("Error opening outputfile\n");
  exit(3); }
printf("Test writing to outputfile\n");
if ((fp6=fopen(file6,"w"))==NULL) {
  printf("Error opening outputfile\n");
  exit(3); }
printf("Test writing to outputfile\n");
/*fprintf(fp3,"just a quick hello\n");*/
/* printf("Opened filename: %s - reading NOEs\n",filename); */
noetot=0;
i=-1;
while(!feof(fp1)) {
 i=i+1;
 count=count+1;
 fscanf(fp1,"%s",&assignc);
 if(strcmp(assignc,"assign")==0) {
  noetot=noetot+1;
  fscanf(fp1, "%s%d%s%s%s",&resid,&resnum1c[i],&and,&name,&atname1c[i]);
  fscanf(fp1, "%s%d%s%s%s",&resid,&resnum2c[i],&and,&name,&atname2c[i]);
  fscanf(fp1, "%f%f%f",&noelo[i],&noemid[i],&noehi[i]);
/*   printf("%d %d\n",resnum1c[i],resnum2c[i]);*/}
 else continue;
 }
fclose(fp1);
/****************** Do NOE sums *************************/
noeintra=0;
noeseq=0;
noemed=0;
noelong=0;
for (i=0; i<noetot; i++) {
  if (resnum2c[i]-resnum1c[i]<0) printf("Wrong order for %d %s %d
%s\n",resnum1c[i],atname1c[i],resnum2c[i],atname2c[i]);
  if (resnum2c[i]-resnum1c[i]==0) {noeintra=noeintra+1;fprintf(fp3,"assign (resid %d and
name %s (resid %d and name %s %2.1f %2.1f
%2.1f\n",resnum1c[i],atname1c[i],resnum2c[i],atname2c[i],noelo[i],noemid[i],noehi[i]);}

  if (resnum2c[i]-resnum1c[i]==1) {noeseq=noeseq+1;fprintf(fp4,"assign (resid %d and
name %s (resid %d and name %s %2.1f %2.1f
%2.1f\n",resnum1c[i],atname1c[i],resnum2c[i],atname2c[i],noelo[i],noemid[i],noehi[i]);}
  if (resnum2c[i]-resnum1c[i]>1 && resnum2c[i]-resnum1c[i]<5)
{noemed=noemed+1;fprintf(fp5,"assign (resid %d and name %s (resid %d and name %s
%2.1f %2.1f
%2.1f\n",resnum1c[i],atname1c[i],resnum2c[i],atname2c[i],noelo[i],noemid[i],noehi[i]);}
  if (resnum2c[i]-resnum1c[i]>4) {noelong=noelong+1;fprintf(fp6,"assign (resid %d and
name %s (resid %d and name %s %2.1f %2.1f
%2.1f\n",resnum1c[i],atname1c[i],resnum2c[i],atname2c[i],noelo[i],noemid[i],noehi[i]);}
}
```

```
fclose(fp3);
fclose(fp4);
fclose(fp5);
fclose(fp6);

/*** report results ***/
printf("Total number of NOEs        is: %d\n",noetot);
printf("_____________________________________________\n");
printf("Total number of intrares NOEs is: %d\n",noeintra);
printf("Total number of sequential NOES : %d\n",noeseq);
printf("Total number of med NOEs 1<i-j<5: %d\n",noemed);
printf("Total number of long NOEs i-j>4 : %d\n",noelong);
printf(" \n");
printf(" \n");
}
```

ptnoe.c

```
/* ptnoe.c */
/* by FFD */
/* Program to add up number of NOEs at each residue position */
/* creates a file containing three columns of integers*/
/* res# total-noes long-range-noes */
/* you can ftp this file to a mac and read it directly into */
/* your favorite graphing program/spreadsheet */
/* input file required: noe_tbl.1 with all comment lines removed */
/* this is easily accomplished on a COPY of the noe_tbl.1 */
/* with the vi command: */
/* :g/!/d1 [RET]  */
/* Then write to a new file name */
/* Input file: edited noe_tbl.1  -> output file whatever_you_call_it  */

/* AUTHOR: Fred Damberger */

#include <math.h>
#include <stdio.h>
#include <string.h>

void main()
{
FILE *fp1, *fp3;
char *p;
int i, j, k, resnum1c[2000],resnum2c[2000],noept[250],noelongpt[250];
float noelo[2000], noemid[2000], noehi[2000];
int noeintra, noeseq, noemed, noelong,noetot,resnum;
/*int noeintrabb, noeseqbb, noemedbb, noelongbb;*/
char assignc[10],resid[10],and[10],name[10],filename[20],outfile[20];
char atname1c[2000][10],atname2c[2000][10],s[80];
/*** Get name of noe_tbl file interactively ***/
printf("Enter name of edited noe_tbl file:\n");
p=gets(s);
if (p) strcpy(filename,s);
printf("Enter name of noetable output file:\n");
p=gets(s);
```

```
if (p) strcpy(outfile,s);
printf("Output file will be %s\n",outfile);
printf("Enter number of residues (dont exceed 250 please):");
p=gets(s);
if (p) resnum=atoi(s);
printf(" \n");
printf(" \n");
printf("Looking for NOEs in %s\n",filename);
/***************Read in NOEs from filename*************/
/*strcpy(filename,argv[1]);*/
if ((fp1=fopen(filename, "r")) == NULL) {
  printf("Error opening filename\n");
  exit(1); }
printf("Opening outputfile %s\n",outfile);
if ((fp3=fopen(outfile,"w"))==NULL) {
  printf("Error opening outputfile\n");
  exit(3); }
/*printf("Test writing to outputfile\n");*/
/*fprintf(fp3,"just a quick hello\n");*/
/* printf("Opened filename: %s - reading NOEs\n",filename); */
noetot=-1;
i=-1;
while(!feof(fp1)) {
 i=i+1;
 noetot=noetot+1;
 fscanf(fp1,"%s",&assignc);
 if(strcmp(assignc,"assign")==0) {
  fscanf(fp1, "%s%d%s%s%s",&resid,&resnum1c[i],&and,&name,&atname1c[i]);
  fscanf(fp1, "%s%d%s%s%s",&resid,&resnum2c[i],&and,&name,&atname2c[i]);
  fscanf(fp1, "%f%f%f",&noelo[i],&noemid[i],&noehi[i]);
/*   printf("%d %d\n",resnum1c[i],resnum2c[i]);*/}
 else continue;
 }
fclose(fp1);
/******************* Do NOE sums *************************/
noeintra=0;
noeseq=0;
noemed=0;
noelong=0;
for (j=0; j<resnum; j++) {noept[j]=0;noelongpt[j]=0;}
for (j=0; j<resnum; j++) {
  for (i=0; i<noetot; i++) {
   if (resnum1c[i]==j) {noept[j]=noept[j]+1;
    if ((resnum1c[i]-resnum2c[i]>5) || (resnum2c[i]-resnum1c[i]>5))
noelongpt[j]=noelongpt[j]+1;
   }
   if (resnum2c[i]==j) {noept[j]=noept[j]+1;
    if ((resnum1c[i]-resnum2c[i]>5) || (resnum2c[i]-resnum1c[i]>5))
noelongpt[j]=noelongpt[j]+1;
   }
  }
}
```

```
/*** report results ***/
i=0;
for (j=0; j<resnum; j++) {i=i+1;fprintf(fp3,"%d %d %d\n",i,noept[i],noelongpt[i]);}
fclose(fp3);
}
```

ptbkrmsd.c

```
/* ptbkrmsd.c */
/*  Program to find and list by residue backbone N,CA,C,O rmsds for plotting */

/* Input: obtains structure filename and number of structures interactively */
/*       from user */
/* Output: ptbkrmsd.output */

/* AUTHOR: Jeff Pelton */

#include <math.h>
#include <stdio.h>
#include <string.h>

main()
{
FILE *fp1, *fp3;
char *p;
int i, j, xyzc,atomnumc,resnumc[3000],n;
float xc[3000],yc[3000],zc[3000];
char atomc[10],atnamec[3000][10],resnamec[3000][5];
float c1,bfc[5000], a, b, cd, d, dd, sum;
char c212c[10];
char filename[30];
char s[30];
printf("\n Output filename is ptbkrmsd.output \n\n");
printf("Must delete header remarks in input file before running\n");

printf("Enter input filename\n");
p=gets(s);
if(p) strcpy(filename,s);
/***          Open file for output    ******************************** */
  if ((fp3=fopen("ptbkrmsd.output", "w")) == NULL) {
   printf("Error opening ptbkrmsd.output\n");
   exit(1); }
/*************** Read in xyz coordinates for a structure ******************/
/* From X-PLOR with bfactor column containing rmsds for that atom */
/* Must remove header format and remarks statements before running */
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening %s\n",filename);
   exit(1); }
printf("Opened xyz coordinate file xyz.d - reading coordinates \n");
  i=-1;
  xyzc=-1;
```

```
 while(!feof(fp1)) {
   i=i+1;
   xyzc=xyzc+1;
   fscanf(fp1, "%s%d", &atomc, &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[i],
&resnamec[i],&resnumc[i],&xc[i],&yc[i],&zc[i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc[i],c212c);
/*   printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[i], resnamec[i], resnumc[i], xc[i],
yc[i], zc[i]);  */
   }
/** Search the data for N,C,CA,O ** Data */
/* Do this - for each residuue HSF 1-92, search coordinates for CA,C,N,O */
/* Sum rmsds for these atoms (# = xyzc- divide by the number found (n) */
for(i=1;i<93;i++) {  n=0; sum=0;
 for(j=0;j<xyzc;j++) {
if(resnumc[j]==i && (strcmp(atnamec[j],"N")==0 || strcmp(atnamec[j],"C")==0
|| strcmp(atnamec[j],"CA")==0 || strcmp(atnamec[j],"O")==0 ||
  strcmp(atnamec[j],"OT1")==0)) { n=n+1;
  sum=sum+bfc[j]; }
              }
/*  If atoms were found for residue i - take the average and print result */
  if(n > 0) {
sum=sum/n;
/*  Make average rmsd for backbone atoms print out */
/*  print out residue number and rmsd for that CA */
     fprintf(fp3,"%d\t%f\n",i,sum);
        }
          }
fclose(fp1);
}
```

ptnonhrmsd.c

```
/* ptnonhrmsd.c */
/*  Program to find and list pointwise backbone+side chain rmsds for plotting */
/*  Using average.pdb files from X-PLOR  with either ave.com or xrayrms.com */

/* Input : prefixname for structure files and number of files interactively */
/*       from user and needs ave.pdb generated by appropriate ave.inp     */
/* Output: ptnonrmsd.output */

/* AUTHOR: Jeff Pelton */

#include <math.h>
#include <stdio.h>
#include <string.h>

main()
{
FILE *fp1, *fp3;
char *p;
int i, j, xyzc,atomnumc,resnumc[3000],n;
```

```
float xc[3000],yc[3000],zc[3000];
char atomc[10],atnamec[3000][10],resnamec[3000][5];
float c1,bfc[5000], a, b, cd, d, dd, sum;
char c212c[10];
char filename[30],s[30];
printf(" Enter pdb file name for ave structure from X-PLOR\n");
p=gets(s);
if(p) strcpy(filename,s);
printf("N.B.  Must delete remarks at top of file before using program\n\n");
printf("Output file is ptnonhrmsd.output\n");
/***          Open file for output    ******************************** */
  if ((fp3=fopen("ptnonhrmsd.output", "w")) == NULL) {
   printf("Error opening ptrmsd.output\n");
   exit(1); }
/*************** Read in xyz coordinates for a structure *******************/
/* From X-PLOR with bfactor column containing rmsds for that atom */
/* Must remove header format and remarks statements before running */
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening %s \n",filename);
   exit(1); }
printf("Opened xyz coordinates - reading coordinates \n");
  i=-1;
  xyzc=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc=xyzc+1;
   fscanf(fp1, "%s%d", &atomc, &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[i],
&resnamec[i],&resnumc[i],&xc[i],&yc[i],&zc[i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc[i],c212c);
/*   printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[i], resnamec[i], resnumc[i], xc[i],
yc[i], zc[i]);  */
  }
/** Search the data for N,C,CA,O ** Data */
/* Do this - for each residuue HSF 1-92, search coordinates for CA,C,N,O */
/* Sum rmsds for these atoms (# = xyzc- divide by the number found (n) */
for(i=1;i<93;i++) {  n=0; sum=0;
 for(j=0;j<xyzc;j++) {
if(resnumc[j]==i && (strcmp(atnamec[j],"N")==0 || strcmp(atnamec[j],"C")==0 \
|| strcmp(atnamec[j],"CA")==0 || strcmp(atnamec[j],"O")==0 ||
  strcmp(atnamec[j],"OT1")==0 || strcmp(atnamec[j],"CB")==0 || \
  strcmp(atnamec[j],"CG")==0 || strcmp(atnamec[j],"CG1")==0 || \
  strcmp(atnamec[j],"CG2")==0 || strcmp(atnamec[j],"OG")==0 || \
  strcmp(atnamec[j],"SG")==0 || strcmp(atnamec[j],"CD")==0 || \
  strcmp(atnamec[j],"CD1")==0 || strcmp(atnamec[j],"CD2")==0 || \
  strcmp(atnamec[j],"SD")==0 || strcmp(atnamec[j],"OD1")==0 || \
  strcmp(atnamec[j],"ND2")==0 || strcmp(atnamec[j],"OD2")==0 || \
  strcmp(atnamec[j],"CE")==0 || strcmp(atnamec[j],"CE1")==0 || \
  strcmp(atnamec[j],"CE2")==0 || strcmp(atnamec[j],"NE")==0 || \
  strcmp(atnamec[j],"NE1")==0 || strcmp(atnamec[j],"NE2")==0 || \
  strcmp(atnamec[j],"CE3")==0 || strcmp(atnamec[j],"OE1")==0 || \
  strcmp(atnamec[j],"OE2")==0 || strcmp(atnamec[j],"OE2")==0 || \
```

```
  strcmp(atnamec[j],"CZ")==0 || strcmp(atnamec[j],"CZ2")==0 || \
  strcmp(atnamec[j],"CZ3")==0 || strcmp(atnamec[j],"NZ")==0 || \
  strcmp(atnamec[j],"NH1")==0 || strcmp(atnamec[j],"NH2")==0 || \
  strcmp(atnamec[j],"CH2")==0 || strcmp(atnamec[j],"OH")==0
                                )) { n=n+1;
  sum=sum+bfc[j]; }
              }
/* If atoms were found for residue i - take the average and print result */
  if(n > 0) {
sum=sum/n;
/* Make average rmsd for backbone atoms print out */
/* print out residue number and rmsd for that CA */
      fprintf(fp3,"%d\t%f\n",i,sum);
        }
          }
fclose(fp1);
}
```

phicalc_s.c

```
/*     phicalc_s.c     */
/* Program to calculate PHI angles and order param for fam of structure */
/* Asks for base file name - ex. accept_ and number of structures */
/* Input file accept_*.pdb - output file redirect using pipe */
/* to compile use cc phicalc_s.c -lm (-lm to include math utility)*/
/* Below is an overview of the program written for chi1. This program */
/* works in an analogous fashion for phi angles */

/* Overview of program: - uses xyz coordinates for n,ca,cb,cg from */
/* X-PLOR pdb file - 1) calculate nca,cacb,cbca,cgcb vectors using */
/* function vector. 2) calculate two cross products A) ncacbca and*/
/* B) cgcbcacb using function crossx,crossy,crossz. 3) Calculate
angle between the two resultant vectors via a dot product. 4) Worry
about the sign of the dihedral angle by calculating the determinant
of ncacbca , cacb, and cgcbcacb. Finally sum and calculate vectors
for order parameter calculation - print out results */

/* AUTHOR: Jeff Pelton */

#include <math.h>
#include <stdio.h>
#include <string.h>

float vector(float s1, float s2);
float mag(double x1, double y1, double z1);
float crossx(double x1, double y1, double z1, double x2, double y2, double z2);
float crossy(double x1, double y1, double z1, double x2, double y2, double z2);
float crossz(double x1, double y1, double z1, double x2, double y2, double z2);

void main(void)
{
FILE *fp,*fp1, *fp3;
char *p;
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
float nx[50][100],ny[50][100],nz[50][100];
float cax[50][100],cay[50][100],caz[50][100];
float cox[50][100],coy[50][100],coz[50][100];
float cx[50][3000],cy[50][3000],cz[50][3000];
float conx[50][100],cony[50][100],conz[50][100];
float ncax[50][100],ncay[50][100],ncaz[50][100];
float canx[50][100],cany[50][100],canz[50][100];
float ccax[50][100],ccay[50][100],ccaz[50][100];
float concanx[50][100],concany[50][100],concanz[50][100];
```

```
float ccancax[50][100],ccancay[50][100],ccancaz[50][100];
float theta[50][100],x,y,xave,yave;
float r,dot[50][100],mag1[50][100],mag2[50][100];
int residue;
char filename1[50];
float det[50][100],detx[50][100],dety[50][100],detz[50][100];
float phi[50][100];
float totx[100],toty[100],angave,angtot;
/*char fileno[2][2]={"1","2"};*/
char
fileno[40][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24","25","26","27","28","29","30","31","32","3
3","34","35","36","37","38","39","40"};
char resname[100][5];
float tempx,tempy;
int reno;
float ang360[50][100];
int outp;

strcpy(pdb,".pdb");

printf("Program phicalc_s  -  used to calculate phi angles and order\n");
printf("parameters for a family of structures\n");
printf("output consists of phi for one structure and order parameter \n");
printf("for the family of structures\n");
printf("\n");
printf("Warning - average angle is only accurate if all angles have the ");
printf("same sign");
printf("\n");
printf("Output file - phicalc_s.output");
printf("\n\n\n");

/*  **** Get basic information interactively  ***** */
/* base filename, number of structures and number of residues */
/* no = number of structures,reno - number of residues */
/* filename - filename1 - base filename */
/* Long or short output */
printf("enter Base filename -ex. refine_  :\n");
/**  **/
p=gets(s);
if(p) {strcpy(filename,s); strcpy(filename1,s);}
printf("enter number of files - default 30: \n");
p=gets(s);
if(p) no=atoi(s);
printf("enter number of residues :\n");
p=gets(s);
if(p) reno=atoi(s);
printf("number of files is: %d / Base file name is %s  \n",no,filename);
printf("Long or short output : 0 for short, 1 for long\n");
p=gets(s);
if(p) outp=atoi(s);
/****** Open output file *******/
```

```
if((fp=fopen("phicalc_s.output","w"))== NULL) {
  printf("Error opening file phicalc_s.output");
  exit(1); }
fprintf(fp, "Warning - average angle is only accurate if all angles have the ");
fprintf(fp, " same sign \n");
/*************** Read in xyz coordinates for structures ******************/
for(k=0;k<no;k++) {
 /* add="l";*/
  strcat(filename,fileno[k]);
  strcat(filename,pdb);
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening file %s\n",filename);
   exit(1); }
/*printf("Opened xyz coordinate file %s - reading coordinates \n",filename);*/
  strcpy(filename,filename1);
  i=-1;
  xyzc[k]=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc[k]=xyzc[k]+1;
   fscanf(fp1,"%s",&atomc);
  if(strcmp(atomc,"ATOM")==0) {
   fscanf(fp1, "%d", &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[k][i],
&resnamec[k][i],&resnumc[k][i],&xc[k][i],&yc[k][i],&zc[k][i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc,&c212c);
/* printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[k][i], resnamec[k][i],
resnumc[k][i], xc[k][i], yc[k][i], zc[k][i]); */  }
else continue;
}
  fclose(fp1);
}
/******************** Search for atoms to calculate phi ****************/
  for(i=0;i<xyzc[k];i++) { cox[k][i]=0.0; coy[k][i]=0.0; coz[k][i]=0.0;
                    cx[k][i]=0.0; cy[k][i]=0.0; cz[k][i]=0.0;
                   /* printf(" residue # %d\n",resnumc[k][i]);*/}


/*  Pick out CO,N,CA,C */
for(k=0;k<no;k++) {
for(i=0; i<xyzc[k]; i++) {residue=resnumc[k][i];
  strcpy(resname[residue],resnamec[k][i]);
  if(strcmp(atnamec[k][i],"C")==0) {
       cx[k][(resnumc[k][i])]=xc[k][i];
       cy[k][(resnumc[k][i])]=yc[k][i];
       cz[k][(resnumc[k][i])]=zc[k][i];
/*printf("resid %d cx %2.3f cy %2.3f cz
%2.3f\n",resnumc[k][i],cx[k][residue],cy[k][residue],cz[k][residue]); */ }
  if(strcmp(atnamec[k][i],"N")==0) {
       nx[k][resnumc[k][i]]=xc[k][i];
       ny[k][resnumc[k][i]]=yc[k][i];
       nz[k][resnumc[k][i]]=zc[k][i];
```

```
/*printf("resid %d nx %2.3f ny %2.3f nz %2.3f\n",residue,
nx[k][residue],ny[k][residue],nz[k][residue]); */}
  if(strcmp(atnamec[k][i],"CA")==0) {
       cax[k][resnumc[k][i]]=xc[k][i];
       cay[k][resnumc[k][i]]=yc[k][i];
       caz[k][resnumc[k][i]]=zc[k][i];
}
} }
/** Calculate vectors **/
for(j=1;j<reno+1;j++) {
 for(k=0;k<no;k++) {
              if(cx[k][j]==0.0 && cy[k][j]==0.0) continue;

        conx[k][j]=vector(cx[k][j-1],nx[k][j]);
        cony[k][j]=vector(cy[k][j-1],ny[k][j]);
        conz[k][j]=vector(cz[k][j-1],nz[k][j]);
/*printf(" resid %d conx %2.3f cony %2.3f conz %2.3f
\n",j,conx[k][j],cony[k][j],conz[k][j]); */

        canx[k][j]=vector(cax[k][j],nx[k][j]);
        cany[k][j]=vector(cay[k][j],ny[k][j]);
        canz[k][j]=vector(caz[k][j],nz[k][j]);
/*printf("resid %d canx %2.3f cany %2.3f canz %2.3f
\n",j,canx[k][j],cany[k][j],canz[k][j]); */

        ncax[k][j]=vector(nx[k][j],cax[k][j]);
        ncay[k][j]=vector(ny[k][j],cay[k][j]);
        ncaz[k][j]=vector(nz[k][j],caz[k][j]);

        ccax[k][j]=vector(cx[k][j],cax[k][j]);
        ccay[k][j]=vector(cy[k][j],cay[k][j]);
        ccaz[k][j]=vector(cz[k][j],caz[k][j]);
/*printf(" %2.3f  %2.3f  %2.3f\n",ccax[k][j],ccay[k][j],ccaz[k][j]);*/
/** Calc cross products **/

 concanx[k][j]=crossx(conx[k][j],cony[k][j],conz[k][j],canx[k][j],cany[k][j],canz[k][j]);
 concany[k][j]=crossy(conx[k][j],cony[k][j],conz[k][j],canx[k][j],cany[k][j],canz[k][j]);
 concanz[k][j]=crossz(conx[k][j],cony[k][j],conz[k][j],canx[k][j],cany[k][j],canz[k][j]);
/*printf("resid %d cross concanx %2.3f %2.3f
%2.3f\n",j,concanx[k][j],concany[k][j],concanz[k][j]); */
 ccancax[k][j]=crossx(ccax[k][j],ccay[k][j],ccaz[k][j],ncax[k][j],ncay[k][j],ncaz[k][j]);
 ccancay[k][j]=crossy(ccax[k][j],ccay[k][j],ccaz[k][j],ncax[k][j],ncay[k][j],ncaz[k][j]);
 ccancaz[k][j]=crossz(ccax[k][j],ccay[k][j],ccaz[k][j],ncax[k][j],ncay[k][j],ncaz[k][j]);
/*printf("resid %d cross ccancax %2.3f %2.3f
%2.3f\n",j,ccancax[k][j],ccancay[k][j],ccancaz[k][j]);*/
/** Calc magnitude of each **/

  mag1[k][j]=mag(ccancax[k][j],ccancay[k][j],ccancaz[k][j]);
/*printf("resid %d mag1= %2.3f\n",j,mag1[k][j]);*/
  mag2[k][j]=mag(concanx[k][j],concany[k][j],concanz[k][j]);
/*printf("resid %d mag2= %2.3f\n",j,mag2[k][j]);*/
```

```
/** Calc dot product + angle **/

dot[k][j]=ccancax[k][j]*concanx[k][j]+ccancay[k][j]*concany[k][j]+ccancaz[k][j]*concanz
[k][j];
/*printf("resid %d dot= %2.3f\n",j,dot[k][j]);*/
theta[k][j]=acos(dot[k][j]/(mag1[k][j]*mag2[k][j]));

/** Determine the sign of the angle using determinant -*/
/* Determinant of concan/nca/ccanca */

detx[k][j]=concanx[k][j]*crossx(ncax[k][j],ncay[k][j],ncaz[k][j],\
  ccancax[k][j],ccancay[k][j],ccancaz[k][j]);
dety[k][j]=1.0*(concany[k][j]*crossy(ncax[k][j],ncay[k][j],ncaz[k][j],\
  ccancax[k][j],ccancay[k][j],ccancaz[k][j]));
detz[k][j]=concanz[k][j]*crossz(ncax[k][j],ncay[k][j],ncaz[k][j],\
  ccancax[k][j],ccancay[k][j],ccancaz[k][j]);
det[k][j]=detx[k][j]+dety[k][j]+detz[k][j];
if(det[k][j]>0.0) phi[k][j]=(theta[k][j]*57.2958)-180.0;
if(det[k][j]<0.0) phi[k][j]=180.0-(theta[k][j]*57.2958);



/*printf("resid %d detx %2.3f %2.3f %2.3f\n",j,detx[k][j],dety[k][j],detz[k][j]);*/
/*printf("resid %d det %2.3f theta %2.3f chi1 %2.3f\n",j,det[k][j],theta[k][j],chi1[k][j]);
*/
   } }

/** Calculate phi order parameter for each residue using the family of structures **/

for(j=1;j<reno+1;j++) { i=0; x=0.0; y=0.0; xave=0.0; yave=0.0; angtot=0.0;
                angave=0.0;
  for(k=0;k<no;k++) {
angtot=phi[k][j]+angtot;
tempx=cos(phi[k][j]*0.017453);
tempy=sin(phi[k][j]*0.017453);
i=i+1;
y=y+tempy;
x=x+tempx;
/*printf("x(%d),y(%d)= %2.3f  %2.3f tempx = %2.3f  %2.3f phi =
%2.3f\n",i,j,j,x,y,tempx,tempy,phi[k][j]);*/ }
xave=x/no;
yave=y/no;
angave=angtot/no;
if(angave >180.0) angave=angave-360.0;
r=sqrt(xave*xave+yave*yave);

/* print the results short or long */
/*printf("start print of residue info");*/
if(outp==1) {
if(cx[1][j]==0.0 || phi[1][j] == 0.0) r=0.0;
fprintf(fp,"%s \t %d ",resname[j],j);
```

```
fprintf(fp,"\tphi(deg)\t %2.3f \tS(phi)\t %2.2f \t angle ave %2.2f \n",phi[1][j],r,angave);
if(r>0.9) { for(i=0;i<no;i++) fprintf(fp,"\t %2.3f \n",phi[i][j]); } }
else { fprintf(fp,"%d   %2.3f\n", j,r); }
}

/* Print out phi for residue 2 for all structures */
if(r>0.9) {
for(k=0;k<no;k++) printf(" phi 2 struct %d is : %2.3f\n",k,phi[k][2]);}
}

float vector(float s1, float s2) {
        return s1-s2; }
float mag(double x1, double y1, double z1) {
        return sqrt(x1*x1+y1*y1+z1*z1); }
float crossx(double x1, double y1, double z1, double x2, double y2, double z2) {
        return y1*z2-y2*z1; }
float crossy(double x1, double y1, double z1, double x2, double y2, double z2) {
        return -1.0*(x1*z2-x2*z1); }
float crossz(double x1, double y1, double z1, double x2, double y2, double z2) {
        return x1*y2-x2*y1; }
```

psicalc_s.c

```
/*psicalc_s.c*/
/*  Program to calculate PSI angles and order param for fam of structure */
/*  Reads in N CA C N coordinates to calc PSI angle */
/*  Derived from phicalc_s - so vector names etc don't quite coorespond */
/* Asks for base file name - ex. accept_ and number of structures */
/*  Input file Accept_*.pdb - output file redirect using pipe */
/* to compile use cc psicalc_s.c -lm (-lm to include math utility)*/
/* Overview of program: - uses xyz coordinates for n,ca,c n from */
/* X-PLOR pdb file - 1) calculate nca,cacb,cbca,cgcb vectors using */
/* function vector. 2) calculate two cross products A) ncacbca and*/
/* B) cgcbcacb using function crossx,crossy,crossz. 3) Calculate
angle between the two resultant vectors via a dot product. 4)  Worry
about the sign of the dihedral angle by calculating the determinant
of ncacbca , cacb, and cgcbcacb.  Finally sum and calculate vectors
for order parameter calculation - print out results */

/* AUTHOR: Jeff Pelton */

#include <math.h>
#include <stdio.h>
#include <string.h>

float vector(float s1, float s2);
float mag(double x1, double y1, double z1);
float crossx(double x1, double y1, double z1, double x2, double y2, double z2);
float crossy(double x1, double y1, double z1, double x2, double y2, double z2);
float crossz(double x1, double y1, double z1, double x2, double y2, double z2);

void main(void)
{
FILE *fp,*fp1, *fp3;
char *p;
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
float nx[50][100],ny[50][100],nz[50][100];
float cax[50][100],cay[50][100],caz[50][100];
float cox[50][100],coy[50][100],coz[50][100];
float cx[50][3000],cy[50][3000],cz[50][3000];
float conx[50][100],cony[50][100],conz[50][100];
float ncax[50][100],ncay[50][100],ncaz[50][100];
float canx[50][100],cany[50][100],canz[50][100];
float ccax[50][100],ccay[50][100],ccaz[50][100];
float concanx[50][100],concany[50][100],concanz[50][100];
float ccancax[50][100],ccancay[50][100],ccancaz[50][100];
float theta[50][100],x,y,xave,yave;
```

```
float r,dot[50][100],mag1[50][100],mag2[50][100];
int residue;
char filename1[50];
float det[50][100],detx[50][100],dety[50][100],detz[50][100];
float phi[50][100];
float totx[100],toty[100],angave,angtot;
/*char fileno[2][2]={"1","2"};*/
char
fileno[30][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24","25","26","27","28","29","30"};
char resname[100][5];
float tempx,tempy;
int reno;
float ang360[50][100];
int outp;

strcpy(pdb,".pdb");

printf("Program psicalc_s  -  used to calculate psi angles and order\n");
printf("parameters for a family of structures\n");
printf("output consists of psi for one structure and order parameter \n");
printf("for the family of structures\n");
printf("\n");
printf("Warning - average angle is only accurate if all angles have the ");
printf("same sign");
printf("\n");
printf("Output file - psicalc_s.output");
printf("\n\n\n");

/*  **** Get basic information interactively  ***** */
/* base filename, number of structures and number of residues */
/* no = number of structures,reno - number of residues */
/* filename - filename1 - base filename */
/* Long or short output */
printf("enter Base filename -ex. refine_  :\n");
/**  **/
p=gets(s);
if(p) {strcpy(filename,s); strcpy(filename1,s);}
printf("enter number of files - default 30: \n");
p=gets(s);
if(p) no=atoi(s);
printf("enter number of residues :\n");
p=gets(s);
if(p) reno=atoi(s);
printf("number of files is: %d / Base file name is %s  \n",no,filename);
printf("Long or short output : 0 for short, 1 for long\n");
p=gets(s);
if(p) outp=atoi(s);
/****** Open output file *******/
if((fp=fopen("psicalc_s.output","w"))== NULL) {
  printf("Error opening file psicalc_s.output");
  exit(1); }
```

```
fprintf(fp,"*************************** \n");
fprintf(fp, "Warning - average angle is only accurate if all angles have the ");
fprintf(fp, " same sign \n");
fprintf(fp, "*************************** \n");
/*************** Read in xyz coordinates for structures ******************/
for(k=0;k<no;k++) {
 /* add="l";*/
  strcat(filename,fileno[k]);
  strcat(filename,pdb);
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening file %s\n",filename);
   exit(1); }
/*printf("Opened xyz coordinate file %s - reading coordinates \n",filename);*/
  strcpy(filename,filename1);
  i=-1;
  xyzc[k]=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc[k]=xyzc[k]+1;
   fscanf(fp1,"%s",&atomc);
  if(strcmp(atomc,"ATOM")==0) {
   fscanf(fp1, "%d", &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[k][i],
&resnamec[k][i],&resnumc[k][i],&xc[k][i],&yc[k][i],&zc[k][i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc,&c212c);
/* printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[k][i], resnamec[k][i],
resnumc[k][i], xc[k][i], yc[k][i], zc[k][i]); */  }
else continue;
}
   fclose(fp1);
}
/******************** Search for atoms to calculate psi  ****************/
  for(i=0;i<xyzc[k];i++) { cox[k][i]=0.0; coy[k][i]=0.0; coz[k][i]=0.0;
                    cx[k][i]=0.0; cy[k][i]=0.0; cz[k][i]=0.0;
                   /* printf(" residue # %d\n",resnumc[k][i]);*/}


/*  Pick out N,CA,C N*/
for(k=0;k<no;k++) {
for(i=0; i<xyzc[k]; i++) {residue=resnumc[k][i];
  strcpy(resname[residue],resnamec[k][i]);
  if(strcmp(atnamec[k][i],"N")==0) {
       cx[k][(resnumc[k][i])]=xc[k][i];
       cy[k][(resnumc[k][i])]=yc[k][i];
       cz[k][(resnumc[k][i])]=zc[k][i];
/* printf("resid %d cx %2.3f cy %2.3f cz
%2.3f\n",resnumc[k][i],cx[k][residue],cy[k][residue],cz[k][residue]); */ }
  if(strcmp(atnamec[k][i],"CA")==0) {
       nx[k][resnumc[k][i]]=xc[k][i];
       ny[k][resnumc[k][i]]=yc[k][i];
       nz[k][resnumc[k][i]]=zc[k][i];
```

```
/*printf("resid %d nx %2.3f ny %2.3f nz %2.3f\n",residue,
nx[k][residue],ny[k][residue],nz[k][residue]); */}
  if(strcmp(atnamec[k][i],"C")==0) {
        cax[k][resnumc[k][i]]=xc[k][i];
        cay[k][resnumc[k][i]]=yc[k][i];
        caz[k][resnumc[k][i]]=zc[k][i];
}
} }
/** Calculate vectors **/
for(j=1;j<reno+1;j++) {
 for(k=0;k<no;k++) {
                if(cx[k][j]==0.0 && cy[k][j]==0.0) continue;

          conx[k][j]=vector(cx[k][j],nx[k][j]);
          cony[k][j]=vector(cy[k][j],ny[k][j]);
          conz[k][j]=vector(cz[k][j],nz[k][j]);
/* printf(" resid %d conx %2.3f cony %2.3f conz %2.3f
\n",j,conx[k][j],cony[k][j],conz[k][j]);
*/
          canx[k][j]=vector(cax[k][j],nx[k][j]);
          cany[k][j]=vector(cay[k][j],ny[k][j]);
          canz[k][j]=vector(caz[k][j],nz[k][j]);
/*printf("resid %d canx %2.3f cany %2.3f canz %2.3f
\n",j,canx[k][j],cany[k][j],canz[k][j]);
*/
          ncax[k][j]=vector(nx[k][j],cax[k][j]);
          ncay[k][j]=vector(ny[k][j],cay[k][j]);
          ncaz[k][j]=vector(nz[k][j],caz[k][j]);

          ccax[k][j]=vector(cx[k][j+1],cax[k][j]);
          ccay[k][j]=vector(cy[k][j+1],cay[k][j]);
          ccaz[k][j]=vector(cz[k][j+1],caz[k][j]);
/*printf(" %2.3f  %2.3f  %2.3f\n",ccax[k][j],ccay[k][j],ccaz[k][j]);*/
/** Calc cross products **/

 concanx[k][j]=crossx(conx[k][j],cony[k][j],conz[k][j],canx[k][j],cany[k][j],canz[k][j]);
 concany[k][j]=crossy(conx[k][j],cony[k][j],conz[k][j],canx[k][j],cany[k][j],canz[k][j]);
 concanz[k][j]=crossz(conx[k][j],cony[k][j],conz[k][j],canx[k][j],cany[k][j],canz[k][j]);
/*printf("resid %d cross concanx %2.3f %2.3f
%2.3f\n",j,concanx[k][j],concany[k][j],concanz[k][j]); */
 ccancax[k][j]=crossx(ccax[k][j],ccay[k][j],ccaz[k][j],ncax[k][j],ncay[k][j],ncaz[k][j]);
 ccancay[k][j]=crossy(ccax[k][j],ccay[k][j],ccaz[k][j],ncax[k][j],ncay[k][j],ncaz[k][j]);
 ccancaz[k][j]=crossz(ccax[k][j],ccay[k][j],ccaz[k][j],ncax[k][j],ncay[k][j],ncaz[k][j]);
/*printf("resid %d cross ccancax %2.3f %2.3f
%2.3f\n",j,ccancax[k][j],ccancay[k][j],ccancaz[k][j]);*/
/** Calc magnitude of each **/

  mag1[k][j]=mag(ccancax[k][j],ccancay[k][j],ccancaz[k][j]);
/*printf("resid %d mag1= %2.3f\n",j,mag1[k][j]);*/
  mag2[k][j]=mag(concanx[k][j],concany[k][j],concanz[k][j]);
/*printf("resid %d mag2= %2.3f\n",j,mag2[k][j]);*/
```

```
/** Calc dot product + angle **/

dot[k][j]=ccancax[k][j]*concanx[k][j]+ccancay[k][j]*concany[k][j]+ccancaz[k][j]*concanz
[k][j];
/*printf("resid %d dot= %2.3f\n",j,dot[k][j]);*/
theta[k][j]=acos(dot[k][j]/(mag1[k][j]*mag2[k][j]));

/** Determine the sign of the angle using determinant -*/
/* Determinant of concan/nca/ccanca */

detx[k][j]=concanx[k][j]*crossx(ncax[k][j],ncay[k][j],ncaz[k][j],\
  ccancax[k][j],ccancay[k][j],ccancaz[k][j]);
dety[k][j]=1.0*(concany[k][j]*crossy(ncax[k][j],ncay[k][j],ncaz[k][j],\
  ccancax[k][j],ccancay[k][j],ccancaz[k][j]));
detz[k][j]=concanz[k][j]*crossz(ncax[k][j],ncay[k][j],ncaz[k][j],\
  ccancax[k][j],ccancay[k][j],ccancaz[k][j]);
det[k][j]=detx[k][j]+dety[k][j]+detz[k][j];
if(det[k][j]>0.0) phi[k][j]=(theta[k][j]*57.2958)-180.0;
if(det[k][j]<0.0) phi[k][j]=180.0-(theta[k][j]*57.2958);



/*printf("resid %d detx %2.3f %2.3f %2.3f\n",j,detx[k][j],dety[k][j],detz[k][j]);*/
/*printf("resid %d det %2.3f theta %2.3f chi1 %2.3f\n",j,det[k][j],theta[k][j],chi1[k][j]);
*/
   } }

/** Calculate psi order parameter for each residue using the family of structures **/

for(j=1;j<reno+1;j++) { i=0; x=0.0; y=0.0; xave=0.0; yave=0.0; angtot=0.0;
                angave=0.0;
  for(k=0;k<no;k++) {
angtot=phi[k][j]+angtot;
tempx=cos(phi[k][j]*0.017453);
tempy=sin(phi[k][j]*0.017453);
i=i+1;
y=y+tempy;
x=x+tempx;
/*printf("x(%d),y(%d)= %2.3f  %2.3f tempx = %2.3f  %2.3f psi =
%2.3f\n",i,j,j,x,y,tempx,tempy,psi[k][j]);*/ }
xave=x/no;
yave=y/no;
angave=angtot/no;
if(angave >180.0) angave=angave-360.0;
r=sqrt(xave*xave+yave*yave);

/* print the results short or long */
/*printf("start print of residue info");*/
if(outp==1) {
if(cx[1][j]==0.0 || phi[1][j] == 0.0) r=0.0;
fprintf(fp,"%s  %d ",resname[j],j);
```

```
fprintf(fp,"\tpsi(deg)\t %2.3f \tS(psi) %2.2f \t angle ave %2.2f \n",phi[1][j],r,angave);
if(r>0.9) { for(i=0;i<no;i++) fprintf(fp,"\t %2.3f \n",phi[i][j]); } }
else { fprintf(fp,"%d   %2.3f\n", j,r); }
}

/* Print out psi for residue 2 for all structures */
/* if(r>0.9) {
for(k=0;k<no;k++) printf(" phi 2 struct %d is : %2.3f\n",k,phi[k][2]);} */
}


float vector(float s1, float s2) {
        return s1-s2; }
float mag(double x1, double y1, double z1) {
        return sqrt(x1*x1+y1*y1+z1*z1); }
float crossx(double x1, double y1, double z1, double x2, double y2, double z2) {
        return y1*z2-y2*z1; }
float crossy(double x1, double y1, double z1, double x2, double y2, double z2) {
        return -1.0*(x1*z2-x2*z1); }
float crossz(double x1, double y1, double z1, double x2, double y2, double z2) {
        return x1*y2-x2*y1; }
```

chi1_s.c

```
/*chi1_s.c*/
/* Program to calculate chi1 angles and order param for fam of structure */
/* Asks for base file name - ex. accept_ and number of structures */
/* Input file Accept_*.pdb - output file redirect using pipe */
/* to compile use cc chi1_s.c -lm (-lm to include math utility)*/
/* Overview of program: - uses xyz coordinates for n,ca,cb,cg from */
/* X-PLOR pdb file - 1) calculate nca,cacb,cbca,cgcb vectors using */
/* function vector. 2) calculate two cross products A) ncacbca and*/
/* B) cgcbcacb using function crossx,crossy,crossz. 3) Calculate
angle between the two resultant vectors via a dot product. 4) Worry
about the sign of the dihedral angle by calculating the determinant
of ncacbca , cacb, and cgcbcacb. Finally sum and calculate vectors
for order parameter calculation - print out results */

/* AUTHOR: Jeff Pelton */

/*modified to recognize SER OG and THR OG1 as CD*/
/*by FFD 3/25/95 */

#include <math.h>
#include <stdio.h>
#include <string.h>

float vector(float s1, float s2);
float mag(double x1, double y1, double z1);
float crossx(double x1, double y1, double z1, double x2, double y2, double z2);
```

```
float crossy(double x1, double y1, double z1, double x2, double y2, double z2);
float crossz(double x1, double y1, double z1, double x2, double y2, double z2);

void main(void)
{
FILE *fp,*fp1, *fp3;
char *p;
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
float nx[50][100],ny[50][100],nz[50][100];
float cax[50][100],cay[50][100],caz[50][100];
float cbx[50][100],cby[50][100],cbz[50][100];
float cgx[50][3000],cgy[50][3000],cgz[50][3000];
float ncax[50][100],ncay[50][100],ncaz[50][100];
float cacbx[50][100],cacby[50][100],cacbz[50][100];
float cbcax[50][100],cbcay[50][100],cbcaz[50][100];
float cgcbx[50][100],cgcby[50][100],cgcbz[50][100];
float ncacbcax[50][100],ncacbcay[50][100],ncacbcaz[50][100];
float cgcbcacbx[50][100],cgcbcacby[50][100],cgcbcacbz[50][100];
float theta[50][100],x,y,xave,yave;
float r,dot[50][100],mag1[50][100],mag2[50][100];
int residue;
char filename1[50];
float det[50][100],detx[50][100],dety[50][100],detz[50][100];
float chi1[50][100];
float totx[100],toty[100],angave,angtot;
/*char fileno[2][2]={"1","2"};*/
char
fileno[40][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24","25","26","27","28","29","30","31","32","3
3","34","35","36","37","38","39","40"};
char resname[100][5];
float tempx,tempy;
int reno;
float ang360[50][100];
int outp;

strcpy(pdb,".pdb");

printf("Program chi1_s - used to calculate chi1 angles and order\n");
printf("parameters for a family of structures\n");
printf("output consists of chi1 for one structure and order parameter \n");
printf("for the family of structures\n");
printf("\n");
printf("Output file - chi1_s.output");
printf("\n\n\n");
```

```
/*  **** Get basic information interactively  ***** */
/* base filename, number of structures and number of residues */
/* no = number of structures,reno - number of residues */
/* filename - filename1 - base filename */
printf("enter Base filename -ex. refine_  :\n");
/**  **/
p=gets(s);
if(p) {strcpy(filename,s); strcpy(filename1,s);}
printf("enter number of files - default 24: \n");
p=gets(s);
if(p) no=atoi(s);
printf("enter number of residues :\n");
p=gets(s);
if(p) reno=atoi(s);
printf("number of files is: %d / Base file name is %s  \n",no,filename);
printf("Output format - 0 for short 1 for long\n");
p=gets(s);
if(p) outp=atoi(s);
/****** Open output file *******/
if((fp=fopen("chi1_s.output","w"))== NULL) {
  printf("Error opening file chi1_s.output");
  exit(1); }
/***************  Read in xyz coordinates for structures ******************/
for(k=0;k<no;k++) {
 /* add="l";*/
  strcat(filename,fileno[k]);
  strcat(filename,pdb);
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening file %s\n",filename);
   exit(1); }
printf("Opened xyz coordinate file %s - reading coordinates \n",filename);
  strcpy(filename,filename1);
  i=-1;
  xyzc[k]=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc[k]=xyzc[k]+1;
   fscanf(fp1,"%s",&atomc);
  if(strcmp(atomc,"ATOM")==0) {
   fscanf(fp1, "%d", &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[k][i],
&resnamec[k][i],&resnumc[k][i],&xc[k][i],&yc[k][i],&zc[k][i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc,&c212c);
/* printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[k][i], resnamec[k][i],
resnumc[k][i], xc[k][i], yc[k][i], zc[k][i]); */  }
else continue;
}
  fclose(fp1);
}
/******************** Search for atoms to calculate chi_1  ****************/
/*  Set cg's  to zero */
for(k=0;k<no;k++) {
```

```
  for(i=0;i<xyzc[k];i++) { cgx[k][i]=0.0; cgy[k][i]=0.0; cgz[k][i]=0.0;
                    /* printf(" residue # %d\n",resnumc[k][i]);*/}
 }

 /*  Pick out N,CA,CB,CG */
 for(k=0;k<no;k++) {
 for (i=0; i<xyzc[k]; i++) {residue=resnumc[k][i];
   strcpy(resname[residue],resnamec[k][i]);
   if(strcmp(atnamec[k][i],"N")==0) {
        nx[k][resnumc[k][i]]=xc[k][i];
        ny[k][resnumc[k][i]]=yc[k][i];
        nz[k][resnumc[k][i]]=zc[k][i];
 /*printf("resid %d nx %2.3f ny %2.3f nz
 %2.3f\n",resnumc[k][i],nx[k][residue],ny[k][residue],nz[k][residue]); */ }
   if(strcmp(atnamec[k][i],"CA")==0) {
        cax[k][resnumc[k][i]]=xc[k][i];
        cay[k][resnumc[k][i]]=yc[k][i];
        caz[k][resnumc[k][i]]=zc[k][i];
 /*printf("resid %d cax %2.3f cay %2.3f caz %2.3f\n",residue,
 cax[k][residue],cay[k][residue],caz[k][residue]); */}
   if(strcmp(atnamec[k][i],"CB")==0) {
        cbx[k][resnumc[k][i]]=xc[k][i];
        cby[k][resnumc[k][i]]=yc[k][i];
        cbz[k][resnumc[k][i]]=zc[k][i];
 /*printf("resid %d cbx %2.3f cby %2.3f cbz %2.3f\n",residue,
 cbx[k][residue],cby[k][residue],cbz[k][residue]); */}
   if(strcmp(atnamec[k][i],"CG")==0 || strcmp(atnamec[k][i],"CG1")==0
 || strcmp(atnamec[k][i],"OG")==0 || strcmp(atnamec[k][i],"OG1")==0 ) {
        cgx[k][resnumc[k][i]]=xc[k][i];
        cgy[k][resnumc[k][i]]=yc[k][i];
        cgz[k][resnumc[k][i]]=zc[k][i];
 /*printf("resid %d cgx %2.3f cgy %2.3f cg
 %2.3f\n",residue,cgx[k][residue],cgy[k][residue],cgz[k][residue]); */}
 } }
 /**  Calculate vectors **/
 for(j=1;j<reno+1;j++) {
  for(k=0;k<no;k++) {
                if(cgx[k][j]==0.0 && cgy[k][j]==0.0) continue;

          ncax[k][j]=vector(nx[k][j],cax[k][j]);
          ncay[k][j]=vector(ny[k][j],cay[k][j]);
          ncaz[k][j]=vector(nz[k][j],caz[k][j]);
 /*printf(" resid %d ncax %2.3f ncay %2.3f ncaz %2.3f
 \n",j,ncax[k][j],ncay[k][j],ncaz[k][j]); */

          cbcax[k][j]=vector(cbx[k][j],cax[k][j]);
          cbcay[k][j]=vector(cby[k][j],cay[k][j]);
          cbcaz[k][j]=vector(cbz[k][j],caz[k][j]);
 /*printf("resid %d cbcax %2.3f cbcay %2.3f cbcaz %2.3f
 \n",j,cbcax[k][j],cbcay[k][j],cbcaz[k][j]); */

          cacbx[k][j]=vector(cax[k][j],cbx[k][j]);
```

```
        cacby[k][j]=vector(cay[k][j],cby[k][j]);
        cacbz[k][j]=vector(caz[k][j],cbz[k][j]);

        cgcbx[k][j]=vector(cgx[k][j],cbx[k][j]);
        cgcby[k][j]=vector(cgy[k][j],cby[k][j]);
        cgcbz[k][j]=vector(cgz[k][j],cbz[k][j]);
/*printf(" %2.3f  %2.3f  %2.3f\n",ncax[k][j],ncay[k][j],ncaz[k][j]);*/
/** Calc cross products **/


ncacbcax[k][j]=crossx(ncax[k][j],ncay[k][j],ncaz[k][j],cbcax[k][j],cbcay[k][j],cbcaz[k][j]);

ncacbcay[k][j]=crossy(ncax[k][j],ncay[k][j],ncaz[k][j],cbcax[k][j],cbcay[k][j],cbcaz[k][j]);

ncacbcaz[k][j]=crossz(ncax[k][j],ncay[k][j],ncaz[k][j],cbcax[k][j],cbcay[k][j],cbcaz[k][j]);
/*printf("resid %d cross ncacbcax %2.3f %2.3f
%2.3f\n",j,ncacbcax[k][j],ncacbcay[k][j],ncacbcaz[k][j]); */

cgcbcacbx[k][j]=crossx(cgcbx[k][j],cgcby[k][j],cgcbz[k][j],cacbx[k][j],cacby[k][j],cacbz[
k][j]);

cgcbcacby[k][j]=crossy(cgcbx[k][j],cgcby[k][j],cgcbz[k][j],cacbx[k][j],cacby[k][j],cacbz[
k][j]);

cgcbcacbz[k][j]=crossz(cgcbx[k][j],cgcby[k][j],cgcbz[k][j],cacbx[k][j],cacby[k][j],cacbz[k
][j]);
/*printf("resid %d cross cgcbcacbx %2.3f %2.3f
%2.3f\n",j,cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);*/
/** Calc magnitude of each **/

  mag1[k][j]=mag(cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);
/*printf("resid %d mag1= %2.3f\n",j,mag1[k][j]);*/
  mag2[k][j]=mag(ncacbcax[k][j],ncacbcay[k][j],ncacbcaz[k][j]);
/*printf("resid %d mag2= %2.3f\n",j,mag2[k][j]);*/


/** Calc dot product + angle **/

dot[k][j]=cgcbcacbx[k][j]*ncacbcax[k][j]+cgcbcacby[k][j]*ncacbcay[k][j]+cgcbcacbz[k][j]
*ncacbcaz[k][j];
/*printf("resid %d dot= %2.3f\n",j,dot[k][j]);*/
theta[k][j]=acos(dot[k][j]/(mag1[k][j]*mag2[k][j]));

/**  Determine the sign of the angle using determinant -*/
/*  Determinant of ncacbca/cacb/cgcbcacb  */

detx[k][j]=ncacbcax[k][j]*crossx(cacbx[k][j],cacby[k][j],cacbz[k][j],\
  cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);
dety[k][j]=1.0*(ncacbcay[k][j]*crossy(cacbx[k][j],cacby[k][j],cacbz[k][j],\
  cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]));
detz[k][j]=ncacbcaz[k][j]*crossz(cacbx[k][j],cacby[k][j],cacbz[k][j],\
  cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);
```

```
det[k][j]=detx[k][j]+dety[k][j]+detz[k][j];
if(det[k][j]>0.0) chi1[k][j]=(theta[k][j]*57.2958)-180.0;
if(det[k][j]<0.0) chi1[k][j]=180.0-(theta[k][j]*57.2958);



/*printf("resid %d detx %2.3f %2.3f %2.3f\n",j,detx[k][j],dety[k][j],detz[k][j]);*/
/*printf("resid %d det %2.3f theta %2.3f chi1 %2.3f\n",j,det[k][j],theta[k][j],chi1[k][j]);
*/
   } }

/** Calculate chi_1 order parameter for each residue using the family of structures **/

for(j=1;j<reno+1;j++) { i=0; x=0.0; y=0.0; xave=0.0; yave=0.0; angtot=0.0;
                angave=0.0;
  for(k=0;k<no;k++) {
angtot=ang360[k][j]+angtot;
tempx=cos(chi1[k][j]*0.017453);
tempy=sin(chi1[k][j]*0.017453);
i=i+1;
y=y+tempy;
x=x+tempx;
/*printf("x(%d),y(%d)= %2.3f  %2.3f tempx = %2.3f  %2.3f chi1 =
%2.3f\n",i,j,j,x,y,tempx,tempy,chi1[k][j]);*/ }
xave=x/no;
yave=y/no;
angave=angtot/no;
if(angave >180.0) angave=angave-360.0;
r=sqrt(xave*xave+yave*yave);

/* print the results long or short */
/*printf("start print of residue info");*/
if(cgx[1][j]==0.0 || chi1[1][j] == 0.0) r=0.0;
if(outp==1) {
fprintf(fp,"%s \t %d ",resname[j],j);
fprintf(fp,"\tave_chi1(deg)\t %2.3f \tS(chi1)\t %2.2f \n",chi1[1][j],r);
if(r>0.9) { for(i=0;i<no;i++) fprintf(fp,"\t %2.3f \n",chi1[i][j]); } }
else { fprintf(fp,"%d   %2.3f\n",j,r); }
}

/*  Print out ch1 for residue 2 for all structures */
/*if(r>0.9) {
for(k=0;k<no;k++) printf(" Chi1 2 struct %d is : %2.3f\n",k,chi1[k][2]);*/
}


float vector(float s1, float s2) {

        return s1-s2; }

float mag(double x1, double y1, double z1) {
```

```
        return sqrt(x1*x1+y1*y1+z1*z1); }

float crossx(double x1, double y1, double z1, double x2, double y2, double z2) {

        return y1*z2-y2*z1; }

float crossy(double x1, double y1, double z1, double x2, double y2, double z2) {

        return -1.0*(x1*z2-x2*z1); }

float crossz(double x1, double y1, double z1, double x2, double y2, double z2) {

        return x1*y2-x2*y1; }
```

chi2_s.c

```
/*chi2_s.c*/
/*  Program to calculate chi2 angles and order param for fam of structure */
/* Asks for base file name - ex. accept_ and number of structures */
/*  Input file Accept_*.pdb - output file redirect using pipe */
/* to compile use cc chi2_s.c -lm (-lm to include math utility)*/
/* Overview of program: - uses xyz coordinates for n,ca,cb,cg from */
/* X-PLOR pdb file - 1) calculate nca,cacb,cbca,cgcb vectors using */
/* function vector. 2) calculate two cross products A) ncacbca and*/
/* B) cgcbcacb using function crossx,crossy,crossz. 3) Calculate
angle between the two resultant vectors via a dot product. 4)  Worry
about the sign of the dihedral angle by calculating the determinant
of ncacbca , cacb, and cgcbcacb.  Finally sum and calculate vectors
for order parameter calculation - print out results */

/* AUTHOR: Jeff Pelton */

/* modified to recognize MET SD, ASN OD1, and HIS ND1 as CD */
/* also modified to set S=0 in cases where CD (or equiv) does not exist*/
/* on the short list version */
/* -FFD 3/25/95 */

#include <math.h>
#include <stdio.h>
#include <string.h>

float vector(float s1, float s2);
float mag(double x1, double y1, double z1);
float crossx(double x1, double y1, double z1, double x2, double y2, double z2);
float crossy(double x1, double y1, double z1, double x2, double y2, double z2);
float crossz(double x1, double y1, double z1, double x2, double y2, double z2);

void main(void)
{
FILE *fp,*fp1, *fp3;
char *p;
```

```
int i, j, k, xyzc[50],atomnumc,resnumc[50][3000];
float diff2,xc[50][3000],yc[50][3000],zc[50][3000];
char atomc[10],atnamec[50][3000][10],resnamec[50][3000][5];
float c1,bfc, a, b, cd, d[50], dd,sum,ave,sd,l;
char c212c[10],filename[20],s[80];
char attest1[5],attest2[5],add[5],pdb[5],file[20];
int resnotest1, resnotest2,no;
float nx[50][100],ny[50][100],nz[50][100];
float cax[50][100],cay[50][100],caz[50][100];
float cbx[50][100],cby[50][100],cbz[50][100];
float cgx[50][3000],cgy[50][3000],cgz[50][3000];
float ncax[50][100],ncay[50][100],ncaz[50][100];
float cacbx[50][100],cacby[50][100],cacbz[50][100];
float cbcax[50][100],cbcay[50][100],cbcaz[50][100];
float cgcbx[50][100],cgcby[50][100],cgcbz[50][100];
float ncacbcax[50][100],ncacbcay[50][100],ncacbcaz[50][100];
float cgcbcacbx[50][100],cgcbcacby[50][100],cgcbcacbz[50][100];
float theta[50][100],x,y,xave,yave;
float r,dot[50][100],mag1[50][100],mag2[50][100];
int residue;
char filename1[50];
float det[50][100],detx[50][100],dety[50][100],detz[50][100];
float chi1[50][100];
float totx[100],toty[100],angave,angtot;
/*char fileno[2][2]={"1","2"};*/
char
fileno[40][3]={"1","2","3","4","5","6","7","8","9","10","11","12","13","14","15","16"
,"17","18","19","20","21","22","23","24","25","26","27","28","29","30","31","32","3
3","34","35","36","37","38","39","40"};
char resname[100][5];
float tempx,tempy;
int reno;
float ang360[50][100];
int outp;

strcpy(pdb,".pdb");

printf("Program chi2_s  -  used to calculate chi2 angles and order\n");
printf("parameters for a family of structures\n");
printf("output consists of chi2 for one structure and order parameter \n");
printf("for the family of structures\n");
printf("\n");
printf("Output file - chi2_s.output");
printf("\n\n\n");

/*  **** Get basic information interactively  ***** */
/* base filename, number of structures and number of residues */
/* no = number of structures,reno - number of residues */
/* filename - filename1 - base filename */
printf("enter Base filename -ex. refine_  :\n");
/**  **/
p=gets(s);
```

```
if(p) {strcpy(filename,s); strcpy(filename1,s);}
printf("enter number of files - default 24: \n");
p=gets(s);
if(p) no=atoi(s);
printf("enter number of residues :\n");
p=gets(s);
if(p) reno=atoi(s);
printf("number of files is: %d / Base file name is %s  \n",no,filename);
printf("Output 0 for short and 1 for long\n");
p=gets(s);
if(p) outp=atoi(s);
/****** Open output file *******/
if((fp=fopen("chi2_s.output","w"))== NULL) {
  printf("Error opening file chi2_s.output");
  exit(1); }
/*************** Read in xyz coordinates for structures *******************/
for(k=0;k<no;k++) {
 /* add="1";*/
  strcat(filename,fileno[k]);
  strcat(filename,pdb);
  if ((fp1=fopen(filename, "r")) == NULL) {
   printf("Error opening file %s\n",filename);
   exit(1); }
/*printf("Opened xyz coordinate file %s - reading coordinates \n",filename);*/
  strcpy(filename,filename1);
  i=-1;
  xyzc[k]=-1;
 while(!feof(fp1)) {
   i=i+1;
   xyzc[k]=xyzc[k]+1;
   fscanf(fp1,"%s",&atomc);
  if(strcmp(atomc,"ATOM")==0) {
   fscanf(fp1, "%d", &atomnumc);
   fscanf(fp1, "%s%s%d%f%f%f", &atnamec[k][i],
&resnamec[k][i],&resnumc[k][i],&xc[k][i],&yc[k][i],&zc[k][i]);
   fscanf(fp1, "%f%f%s", &c1, &bfc,&c212c);
/* printf("%s %s %d %2.3f %2.3f %2.3f\n", atnamec[k][i], resnamec[k][i],
resnumc[k][i], xc[k][i], yc[k][i], zc[k][i]); */  }
else continue;
}
  fclose(fp1);
}
/******************** Search for atoms to calculate chi_1  ****************/
/*  Set cg's  to zero */
for(k=0;k<no;k++) {
  for(i=0;i<xyzc[k];i++) { cgx[k][i]=0.0; cgy[k][i]=0.0; cgz[k][i]=0.0;
                    /* printf(" residue # %d\n",resnumc[k][i]);*/}
}

/*  Pick out N,CA,CB,CG */
for(k=0;k<no;k++) {
for (i=0; i<xyzc[k]; i++) {residue=resnumc[k][i];
```

```
  strcpy(resname[residue],resnamec[k][i]);
  if(strcmp(atnamec[k][i],"CA")==0) {
        nx[k][resnumc[k][i]]=xc[k][i];
        ny[k][resnumc[k][i]]=yc[k][i];
        nz[k][resnumc[k][i]]=zc[k][i];
/*printf("resid %d nx %2.3f ny %2.3f nz
%2.3f\n",resnumc[k][i],nx[k][residue],ny[k][residue],nz[k][residue]); */ }
  if(strcmp(atnamec[k][i],"CB")==0) {
        cax[k][resnumc[k][i]]=xc[k][i];
        cay[k][resnumc[k][i]]=yc[k][i];
        caz[k][resnumc[k][i]]=zc[k][i];
/*printf("resid %d cax %2.3f cay %2.3f caz %2.3f\n",residue,
cax[k][residue],cay[k][residue],caz[k][residue]); */}
  if(strcmp(atnamec[k][i],"CG")==0 || strcmp(atnamec[k][i],"CG1")==0) {
        cbx[k][resnumc[k][i]]=xc[k][i];
        cby[k][resnumc[k][i]]=yc[k][i];
        cbz[k][resnumc[k][i]]=zc[k][i];
/*printf("resid %d cbx %2.3f cby %2.3f cbz %2.3f\n",residue,
cbx[k][residue],cby[k][residue],cbz[k][residue]); */}
  if(strcmp(atnamec[k][i],"CD")==0 || strcmp(atnamec[k][i],"CD1")==0
|| strcmp(atnamec[k][i],"OD1")==0 || strcmp(atnamec[k][i],"ND1")==0
|| strcmp(atnamec[k][i],"SD")==0 ) {
        cgx[k][resnumc[k][i]]=xc[k][i];
        cgy[k][resnumc[k][i]]=yc[k][i];
        cgz[k][resnumc[k][i]]=zc[k][i];
/*printf("resid %d cgx %2.3f cgy %2.3f cg
%2.3f\n",residue,cgx[k][residue],cgy[k][residue],cgz[k][residue]); */}
} }
/** Calculate vectors **/
for(j=1;j<reno+1;j++) {
 for(k=0;k<no;k++) {
                if(cgx[k][j]==0.0 && cgy[k][j]==0.0) continue;

          ncax[k][j]=vector(nx[k][j],cax[k][j]);
          ncay[k][j]=vector(ny[k][j],cay[k][j]);
          ncaz[k][j]=vector(nz[k][j],caz[k][j]);
/*printf(" resid %d ncax %2.3f ncay %2.3f ncaz %2.3f
\n",j,ncax[k][j],ncay[k][j],ncaz[k][j]); */

          cbcax[k][j]=vector(cbx[k][j],cax[k][j]);
          cbcay[k][j]=vector(cby[k][j],cay[k][j]);
          cbcaz[k][j]=vector(cbz[k][j],caz[k][j]);
/*printf("resid %d cbcax %2.3f cbcay %2.3f cbcaz %2.3f
\n",j,cbcax[k][j],cbcay[k][j],cbcaz[k][j]); */

          cacbx[k][j]=vector(cax[k][j],cbx[k][j]);
          cacby[k][j]=vector(cay[k][j],cby[k][j]);
          cacbz[k][j]=vector(caz[k][j],cbz[k][j]);

          cgcbx[k][j]=vector(cgx[k][j],cbx[k][j]);
          cgcby[k][j]=vector(cgy[k][j],cby[k][j]);
          cgcbz[k][j]=vector(cgz[k][j],cbz[k][j]);
```

```
/*printf(" %2.3f  %2.3f  %2.3f\n",ncax[k][j],ncay[k][j],ncaz[k][j]);*/
/** Calc cross products **/


ncacbcax[k][j]=crossx(ncax[k][j],ncay[k][j],ncaz[k][j],cbcax[k][j],cbcay[k][j],cbcaz[k][j]);

ncacbcay[k][j]=crossy(ncax[k][j],ncay[k][j],ncaz[k][j],cbcax[k][j],cbcay[k][j],cbcaz[k][j]);

ncacbcaz[k][j]=crossz(ncax[k][j],ncay[k][j],ncaz[k][j],cbcax[k][j],cbcay[k][j],cbcaz[k][j]);
/*printf("resid %d cross ncacbcax %2.3f %2.3f
%2.3f\n",j,ncacbcax[k][j],ncacbcay[k][j],ncacbcaz[k][j]); */

cgcbcacbx[k][j]=crossx(cgcbx[k][j],cgcby[k][j],cgcbz[k][j],cacbx[k][j],cacby[k][j],cacbz[
k][j]);

cgcbcacby[k][j]=crossy(cgcbx[k][j],cgcby[k][j],cgcbz[k][j],cacbx[k][j],cacby[k][j],cacbz[
k][j]);

cgcbcacbz[k][j]=crossz(cgcbx[k][j],cgcby[k][j],cgcbz[k][j],cacbx[k][j],cacby[k][j],cacbz[k
][j]);
/*printf("resid %d cross cgcbcacbx %2.3f %2.3f
%2.3f\n",j,cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);*/
/** Calc magnitude of each **/

 mag1[k][j]=mag(cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);
/*printf("resid %d mag1= %2.3f\n",j,mag1[k][j]);*/
 mag2[k][j]=mag(ncacbcax[k][j],ncacbcay[k][j],ncacbcaz[k][j]);
/*printf("resid %d mag2= %2.3f\n",j,mag2[k][j]);*/


/** Calc dot product + angle **/

dot[k][j]=cgcbcacbx[k][j]*ncacbcax[k][j]+cgcbcacby[k][j]*ncacbcay[k][j]+cgcbcacbz[k][j]
*ncacbcaz[k][j];
/*printf("resid %d dot= %2.3f\n",j,dot[k][j]);*/
theta[k][j]=acos(dot[k][j]/(mag1[k][j]*mag2[k][j]));

/**  Determine the sign of the angle using determinant -*/
/*  Determinant of ncacbca/cacb/cgcbcacb  */

detx[k][j]=ncacbcax[k][j]*crossx(cacbx[k][j],cacby[k][j],cacbz[k][j],\
  cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);
dety[k][j]=1.0*(ncacbcay[k][j]*crossy(cacbx[k][j],cacby[k][j],cacbz[k][j],\
  cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]));
detz[k][j]=ncacbcaz[k][j]*crossz(cacbx[k][j],cacby[k][j],cacbz[k][j],\
  cgcbcacbx[k][j],cgcbcacby[k][j],cgcbcacbz[k][j]);
det[k][j]=detx[k][j]+dety[k][j]+detz[k][j];
if(det[k][j]>0.0) chi1[k][j]=(theta[k][j]*57.2958)-180.0;
if(det[k][j]<0.0) chi1[k][j]=180.0-(theta[k][j]*57.2958);

/*printf("resid %d detx %2.3f %2.3f %2.3f\n",j,detx[k][j],dety[k][j],detz[k][j]);*/
```

```
/*printf("resid %d det %2.3f theta %2.3f chi1 %2.3f\n",j,det[k][j],theta[k][j],chi1[k][j]);
*/
   } }

/** Calculate chi_2 order parameter for each residue using the family of structures **/

for(j=1;j<reno+1;j++) { i=0; x=0.0; y=0.0; xave=0.0; yave=0.0; angtot=0.0;
                angave=0.0;
  for(k=0;k<no;k++) {
angtot=ang360[k][j]+angtot;
tempx=cos(chi1[k][j]*0.017453);
tempy=sin(chi1[k][j]*0.017453);
i=i+1;
y=y+tempy;
x=x+tempx;
/*printf("x(%d),y(%d)= %2.3f  %2.3f tempx = %2.3f  %2.3f chi1 =
%2.3f\n",i,j,j,x,y,tempx,tempy,chi1[k][j]);*/ }
xave=x/no;
yave=y/no;
angave=angtot/no;
if(angave >180.0) angave=angave-360.0;
r=sqrt(xave*xave+yave*yave);

/* print the results long or short form */
/*printf("start print of residue info");*/
if(cgx[1][j]==0.0 || chi1[1][j] == 0.0) r=0.0;
if(outp==1) {
fprintf(fp,"%s \t %d ",resname[j],j);
fprintf(fp,"\tave_chi2(deg)\t %2.3f \tS(chi2)\t %2.2f \n",chi1[1][j],r);
if(r>0.9) { for(i=0;i<no;i++) fprintf(fp,"\t %2.3f \n",chi1[i][j]); }}
else {fprintf(fp,"%d   %2.3f\n",j,r); }
}

/*  Print out chi2 for residue 2 for all structures */
/*if(r>0.9) {
for(k=0;k<no;k++) printf(" Chi2 2 struct %d is : %2.3f\n",k,chi1[k][2]);*/
}


float vector(float s1, float s2) {

        return s1-s2; }

float mag(double x1, double y1, double z1) {

        return sqrt(x1*x1+y1*y1+z1*z1); }

float crossx(double x1, double y1, double z1, double x2, double y2, double z2) {

       return y1*z2-y2*z1; }

float crossy(double x1, double y1, double z1, double x2, double y2, double z2) {
```

```
        return -1.0*(x1*z2-x2*z1); }
float crossz(double x1, double y1, double z1, double x2, double y2, double z2) {
        return x1*y2-x2*y1; }
```

X-PLOR macros for structure analysis

X-PLOR version 3.1 (Brünger, 1992) was used for all calculations on UNIX-based workstations (Silicon Graphics). Structure generation and refinement were accomplished using X-PLOR script files: generate.inp, generate_template.inp, dg_sub_embed.inp, dgsa.inp, refine.inp, and accept.inp except that 2000 steps (instead of 1000 steps) of cooling were applied during the slow cooling phase of dgsa refinement to improve convergence. During interative refinement, typically 15-20 structures were calculated in overnight runs for structural comparison and analysis of recurrent NOE violations. Early rounds of refinement had a convergence rate (no NOE violatios >5 Å, no dihedral violations >5°) of ~50%. Later rounds of refinement at higher resolution produced over 75% converged structures (no NOE violations >3 Å, no dihedral violations >3°). Structures reported in Chapter 3 showed significant numbers of non-bonded contacts using the PROCHECK program (Morris, et al., 1992) within CCP4 and had positive Lennard-Jones energies. These problems were reduced significantly in structures calculated after the completion of this dissertation when the fin_rad parameter in refine.inp was set to 0.80 (final Van der Waals radii 0.80 times their values in the CHARMM energy potential (Brooks, et al., 1983)) rather than the default value of 0.75 with near average numbers of non-bonded contacts and negative calculated Lennard-Jones energies.

ave.com

```
#
/usr1/scratch/somoza/xplor31/xplor31.exe < ave.inp > ave.log
```

ave.inp

```
remarks file  nmr/average.inp
remarks Computes the average structure, atomic rms differences from the
remarks mean for a family of structures, and average overall rms
remarks difference between the family and the mean structure.
remarks user comments in atom selection of interest from
remarks choices below
remarks frequently edited sections are high-lighted with bold
remarks text. These are changed depending on the application

{===>}
structure @/nmr1/wemmer/pelton/xplor/generatec212.psf end  {*Read the structure file.*}


{===>}      {*"Backbone" selection--structured regions.*}
{* commented in for ave_1.inp *}
vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 4:25 or resid 31:68
 or  resid 84:86 ))
vector idend ( store8 )
((name ca or name n or name c or name o) and (resid 4:25 or resid 31:68
 or  resid 84:86 ))

{===>}      {*"Backbone" selection--secondary structure.*}
```

```
{* commented in for ave_2.inp *}
{vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 4:14 or resid 22:24
 or resid 31:34 or resid 36:48 or resid 54:63 or resid 65:67 or  resid 84:86 ))
vector idend ( store8 )
(( resid 4:14 or resid 22:24
 or resid 31:34 or resid 36:48 or resid 54:63 or resid 65:67 or  resid 84:86 ))
}

{====>}      {*"Backbone" selection--helices.*}
{* commented in for ave_3.inp *}
{vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 4:14 or resid 36:48
or resid 54:63 ))
vector idend ( store8 )
(( (resid 4:14 or resid 36:48
or resid 54:63 ))
}
{====>}      {*"Backbone" selection--beta sheet.*}
{* commented in for ave_4.inp *}
{vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 22:24
 or resid 31:34 or resid 65:67 or  resid 84:86 ))
vector idend ( store8 )
((name ca or name n or name c or name o) and (resid 22:24
 or resid 31:34 or resid 65:67 or  resid 84:86 ))
}

{===>}  {*Backbone selection - structured regions except b3 (65-67)*}
{* commented in for ave_5.inp *}
{vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 4:25 or resid 31:65
or resid 84:86))
vector idend ( store8 )
((name ca or name n or name c or name o) and (resid 4:25 or resid 31:65
or resid 84:86))
}

{===>}  {*Backbone selection - b sheet except sheet b3 *}
{* commented in for ave_6.inp *}
{vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 22:24 or resid 31:34 or
resid 84:86))
vector idend ( store8 )
((name ca or name n or name c or name o) and (resid 22:24 or resid 31:34 or
resid 84:86))
}

{===>} {*Backbone selection - secondary structure except b 3}
{* commented in for ave_7.inp *}
{vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 4:14 or resid 22:24 or
```

```
resid 31:34 or resid 36:48 or resid 54:63 or resid 84:86))
vector idend ( store8 )
((name ca or name n or name c or name o) and (resid 4:14 or resid 22:24 or
resid 31:34 or resid 36:48 or resid 54:63 or resid 84:86))
}
{====>}      {*"Backbone" selection--helix-turn-helix *}
{* commented in for ave_8.inp *}
{vector idend ( store9 )
((name ca or name n or name c or name o) and (resid 36:63))
vector idend ( store8 )
((name ca or name n or name c or name o) and (resid 36:63))
}
{====>}      {*"<20% exposed core" selection--.*}
{* commented in for ave_9.inp *}
{vector idend ( store9 )
 ((not hydrogen) and (resid 5 or resid 9 or resid 12 or resid 13 or resid 18
 or resid 21:22 or resid 25 or resid 30:33 or resid 39 or resid 43:44
 or resid 47:48 or resid 53 or resid 56:57 or resid 60 or resid 64:65
 or resid 84 or resid 86 ))

vector idend ( store8 )
 ((not hydrogen) and (resid 5 or resid 9 or resid 12 or resid 13 or resid 18
 or resid 21:22 or resid 25 or resid 30:33 or resid 39 or resid 43:44
 or resid 47:48 or resid 53 or resid 56:57 or resid 60 or resid 64:65
 or resid 84 or resid 86 ))
}

{====>}      {*"<40% exposed core" selection--.*}
{* commented in for ave_10.inp *}
{vector idend ( store9 )
 ((not hydrogen) and (resid 3 or resid 5:6 or resid 9 or resid 12:13
 or resid 18 or resid 21:25 or resid 30:36 or resid 39:40 or resid 42:44
 or resid 47:48 or resid 51 or resid 53 or resid 56:57 or resid 59:60
 or resid 63:67 or resid 82 or resid 84:86 or resid 88))

vector idend ( store8 )
 ((not hydrogen) and (resid 3 or resid 5:6 or resid 9 or resid 12:13
 or resid 18 or resid 21:25 or resid 30:36 or resid 39:40 or resid 42:44
 or resid 47:48 or resid 51 or resid 53 or resid 56:57 or resid 59:60
 or resid 63:67 or resid 82 or resid 84:86 or resid 88))
}

  {*============ The first stage consists of computing the mean structure.*}

{====>}              {*Loop through the family of 10 accepted structures.*}
evaluate ($end_count=33) {*change for the number of structures to be used *}

eval ($nfile=0)
vector do (store1=0) (all)
vector do (store2=0) (all)
vector do (store3=0) (all)
vector do (store4=0) (all)
```

```
evaluate ($count = 0)
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)

{====>}                    {*This is the name of the family of structures.*}
   evaluate ($filename="accept_"+encode($count)+".pdb")
!   evaluate ($filename="refine_"+encode($count)+".pdb")
  coor @@$filename

  if ($count=1) then
    coor copy end            {*Store first structure in comparison set.*}
  end if
  coor sele=( recall9 ) fit end
  vector do (store1=store1+x) (all)
  vector do (store2=store2+y) (all)
  vector do (store3=store3+z) (all)
  vector do (store4=store4+x*x+y*y+z*z) (all)
  eval ($nfile=$nfile+1)
end loop main

vector do (x = store1 / $nfile) (all)
vector do (y = store2 / $nfile) (all)
vector do (z = store3 / $nfile) (all)
vector do (bcomp=sqrt(max(0,store4/$nfile-(x**2+y**2+z**2)))) (all)


     {*The second stage consists of computing an overall rms difference.*}

evaluate ($ave_rmsd_all=0.)
evaluate ($ave_rmsd_back=0.)

coor copy end

evaluate ($count = 0)
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)


{====>}                    {*This is the name of the family of structures.*}
  evaluate ($filename="accept_"+encode($count)+".pdb")
!  evaluate ($filename="refine_"+encode($count)+".pdb")

  coor @@$filename
  coor fit sele=( recall9 ) end
  coor rms selection=( recall9 )end
  evaluate ($ave_rmsd_back=$ave_rmsd_back + $result)
  display rms9 $result
  coor rms selection=( recall8 )end
  evaluate ($ave_rmsd_all =$ave_rmsd_all + $result)
end loop main
```

```
evaluate ($ave_rmsd_back=$ave_rmsd_back / $nfile)
evaluate ($ave_rmsd_all =$ave_rmsd_all  / $nfile)
display ave. rms diff. to the mean struct. for non-h atoms= $ave_rmsd_all
display ave. rms diff. to the mean struct. for the backbone= $ave_rmsd_back

  {*===== Finally, the average structure and RMSDs are written to a file.*}
coor swap end
vector do (b=bcomp) ( all )

remarks unminimized average over $nfile files
remarks ave. rms diff. to the mean struct. for non-h atoms= $ave_rmsd_all
remarks ave. rms diff. to the mean struct. for the backbone= $ave_rmsd_back
remarks b array (last column) is the rms difference from the mean

{====>}          {*Write average coordinates and RMSDs to specified file.*}
write coordinates output=ave.pdb end

stop
```

avecycle.com

```
# cycle through a series of X-PLOR calculations
# to determine <rmsd> values for a family of structures
# in different regions of the structure
#
# requires files ave_1.inp through ave_10.inp
# these files are versions of ave.inp with different
# atom selections commented in (see comment lines in
# ave.inp).
# requires structure files (e.g. accept_1.pdb -accept_33.pdb)
# requires files comment_1 - comment_10
# example given below
# output of standev is sent to rmsdall.log
# example of rmsdall.log file is also given below
set i = 1
while ($i <= 10)
cat comment_{$i} >>rmsdall.log
/usr1/scratch/somoza/xplor31/xplor31.exe < ave_{$i}.inp > ave.log
standev >>rmsdall.log
cat blank >>rmsdall.log
#echo $i "complete"
@ i++
end
```

contents of comment_1 file:

```
backbone-structured regions <rmsd>
```

---

contents of rmsdall.log file:

```
backbone-structured regions <rmsd>
```

---

value of rmsdc is 33
rmstot is 26.682022, ave_rms is 0.808546
The RMS deviation of 33 structures is: 0.81, +/- 0.16

secondary-structured regions <rmsd>

---

value of rmsdc is 33
rmstot is 20.259554, ave_rms is 0.613926
The RMS deviation of 33 structures is: 0.61, +/- 0.14

alpha helical bundle <rmsd>

---

value of rmsdc is 33
rmstot is 17.481466, ave_rms is 0.529741
The RMS deviation of 33 structures is: 0.53, +/- 0.13

beta sheet <rmsd>

---

value of rmsdc is 33
rmstot is 18.692917, ave_rms is 0.566452
The RMS deviation of 33 structures is: 0.57, +/- 0.15

backbone structured regions - B3 <rmsd>

---

value of rmsdc is 33
rmstot is 23.781012, ave_rms is 0.720637
The RMS deviation of 33 structures is: 0.72, +/- 0.15

beta sheet - B3 <rmsd>

---

value of rmsdc is 33
rmstot is 12.506274, ave_rms is 0.378978
The RMS deviation of 33 structures is: 0.38, +/- 0.07

backbone secondary structured regions - B3 <rmsd>

---

value of rmsdc is 33
rmstot is 18.367388, ave_rms is 0.556588
The RMS deviation of 33 structures is: 0.56, +/- 0.12

Helix-turn-helix region <rmsd>

---

value of rmsdc is 33
rmstot is 22.207224, ave_rms is 0.672946
The RMS deviation of 33 structures is: 0.67, +/- 0.22

20% exposed in X-ray heavy atoms <rmsd>

---

value of rmsdc is 33
rmstot is 27.610207, ave_rms is 0.836673
The RMS deviation of 33 structures is: 0.84, +/- 0.15

40% exposed in X-ray heavy atoms <rmsd>

---

value of rmsdc is 33
rmstot is 36.994728, ave_rms is 1.121052
The RMS deviation of 33 structures is: 1.12, +/- 0.18

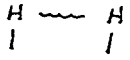
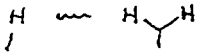
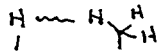
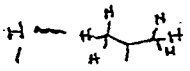
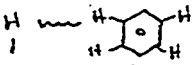
Appendix E Restraints for Structure Calculations

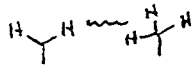
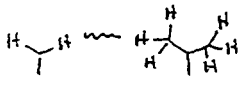
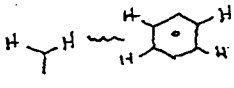
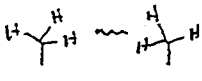
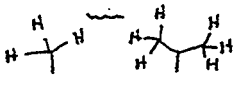
Distance Constraints were derived from NOEs observed in experiments described in the text. NOEs were qualitatively identified as strong (s), medium (m), or weak (w). These intensities were converted into distance constraints (X-PLOR format) using the following table designed by Jeff Pelton. The table is derived from papers reporting the use of NOEs as distance constraints (Williamson, et al., 1985; Clore, et al., 1986). The table includes corrections for pseudoatoms and the larger intensity of NOEs to methyl groups (correction +0.5 Å) (Wüthrich, et al., 1983; Clore, et al., 1986). A second table (Intraresidue) is used for certain cases of NOEs between protons within the same residue. Although not used in the calculations described here, a +0.2 Å correction is often applied to the upper bound of constraints involving NH protons. Thus one would add +0.2 Å to the numbers in the third column for the affected restraints.

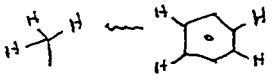
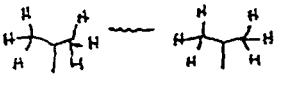
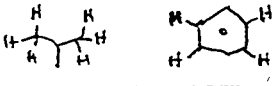
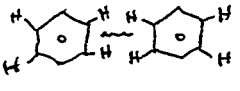
Table E.1 Pseudoatom corrections to NOEs

Pseudoatom corrections

(assuming center averaging and a square-well potential)

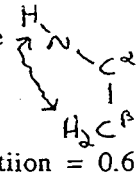
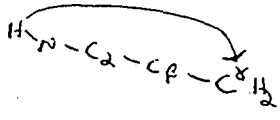
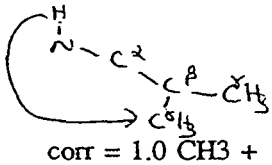
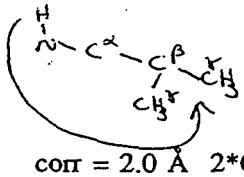
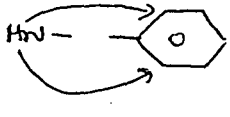
Type	Class				Range
	strong	2.4	0.6	0.3	1.8 - 2.7 Å
correctiion = none	medium	3.3	1.5	0.2	1.8 - 3.5
	weak	5.0	3.2	0.0	1.8 - 5.0
	-				
	s	3.4	1.6	0.3	1.8 - 3.7
correction = 1.0 Å	m	4.3	2.5	0.2	1.8 - 4.5
	w	6.0	4.2	0.0	1.8 - 6.0
	-				
	s	3.9	2.1	0.3	1.8 - 4.2
corr = 1.0 CH3 + 0.5 for intensity	m	4.8	3.0	0.2	1.8 - 5.0
total 1.5 Å	w	6.5	4.7	0.0	1.8 - 6.5
	-				
	s	5.3	3.5	0.3	1.8 - 5.6
corr = 2.4Å 2*CH3 val or leu + 0.5 inten	m	6.2	4.4	0.2	1.8 - 6.4
total 2.9 Å	w	7.9	6.1	0.0	1.8 - 7.9
	-				
	s	4.4	2.6	0.3	1.8 - 4.7
corr = 2.0 Å	m	5.3	3.5	0.2	1.8 - 5.5
	w	7.0	5.2	0.0	1.8 - 7.0
	-				

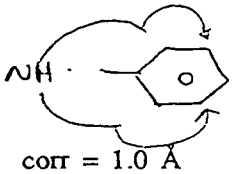
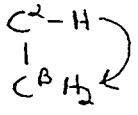
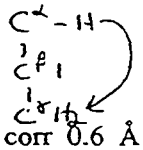
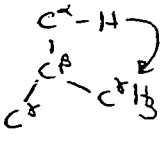
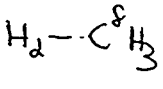
	s	4.4	2.6	0.3	1.8 - 4.7
corr = 2.0 for 2*CH2	m	5.3	3.5	0.2	1.8 - 5.5
	w	7.0	5.2	0.0	1.8 - 7.0
	-				
	s	4.9	3.1	0.3	1.8 - 5.2
corr - 1.0 CH2 + 1.0 CH3 + 0.5 Int total 2.5 Å	m	5.8	4.0	0.2	1.8 - 5.5
	w	7.5	5.7	0.0	1.8 - 7.5
	-				
	s	6.3	4.5	0.3	1.8 - 6.6
corr 1.0 CH2 + 2.4 2*CH3 + 0.5 inten total 3.9 Å	m	7.2	5.4	0.2	1.8 - 7.4
	w	8.9	7.1	0.0	1.8 - 8.9
	-				
	s	5.4	3.6	0.3	1.8 - 5.7
corr 1.0 CH2 + 2.0 Aro ring total 3.0 Å	m	6.3	4.5	0.2	1.8 - 6.5
	w	8.0	6.2	0.0	1.8 - 8.0
	-				
	s	4.9	3.1	0.3	1.8 - 5.2
corr 2.0 for 2*CH3 + 0.5 inten total 2.5 Å	m	5.8	4.0	0.2	1.8 - 6.0
	w	7.5	5.7	0.0	1.8 - 7.5
	-				
	s	6.3	4.5	0.3	1.8 - 6.6
corr = 1.0 CH3 + 2.4 2*CH3 + 0.5 inten total 3.9 Å	m	7.2	5.4	0.2	1.8 - 7.4
	w	8.9	7.1	0.0	1.8 - 8.9
	-				

	s	5.9	4.1	0.3	1.8 - 6.2
corr = 1.0 CH3 + 2.0 Aro ring + 0.5 inten	m	6.8	5.0	0.2	1.8 - 7.0
total 3.5	w	8.5	6.7	0.0	1.8 - 8.5
	-				
	s	7.7	5.9	0.3	1.8 - 8.0
corr = 2.4*2 for 2* (2*CH3) + 0.5 inten	m	8.6	6.8	0.2	1.8 - 8.8
total 5.3 Å	w	10.3	8.5	0.0	1.8 - 10.3
	-				
	s	7.3	5.5	0.3	1.8 - 7.6
corr = 2.4 2*CH3 + 2.0 Aro + 0.5 inten	m	8.2	6.4	0.2	1.8 - 8.4
total 4.9 Å	w	9.9	8.1	0.0	1.8 - 9.9
	-				
	s	6.4	4.6	0.3	1.8 - 6.7
corr = 4.0 for 2*Aro ring	m	7.3	5.5	0.2	1.8 - 7.4
total 4.0 Å	w	9.0	7.2	0.0	1.8 - 9.0
H-bonding					
R_{NH-O}		2.1	0.3	0.2	1.8 - 2.3
R_{N-O}		3.1	0.3	0.2	2.8 - 3.3

Pseudoatom corrections - Intraresidue

(assuming center averaging and a square-well potential)

Type	Class				Range
	strong	3.0	1.2	0.3	1.8 - 3.3 Å
correctiion = 0.6 Å	medium	3.9	2.1	0.2	1.8 - 4.1
	weak	5.6	3.8	0.0	1.8 - 5.6
	-				
	s	3.4	1.6	0.3	1.8 - 3.7
correction = 1.0 Å	m	4.3	2.5	0.2	1.8 - 4.5
	w	6.0	4.2	0.0	1.8 - 6.0
	-				
	s	3.9	2.1	0.3	1.8 - 4.2
corr = 1.0 CH3 + 0.5 for intensity	m	4.8	3.0	0.2	1.8 - 5.0
total 1.5 Å	w	6.5	4.7	0.0	1.8 - 6.5
	-				
	s	4.9	3.1	0.3	1.8 - 5.2
corr = 2.0 Å 2*CH3 val or leu + 0.5 inten	m	5.8	4.0	0.2	1.8 - 6.0
total 2.5 Å	w	7.5	5.7	0.0	1.8 - 7.5
	-				
	s	4.4	2.6	0.3	1.8 - 4.7
corr = 2.0 Å	m	5.3	3.5	0.2	1.8 - 5.5
	w	7.0	5.2	0.0	1.8 - 7.0

	s	3.4	1.6	0.3	1.8 - 3.7
corr = 1.0 Å	m	4.3	2.5	0.2	1.8 - 4.5
	w	6.0	4.2	0.0	1.8 - 6.0
	-				
	s	3.0	1.2	0.3	1.8 - 3.3
corr - 0.6 Å	m	3.9	2.1	0.2	1.8 - 4.1
	w	5.6	3.8	0.0	1.8 - 5.6
	-				
	s	3.0	1.2	0.3	1.8 - 3.3
corr 0.6 Å	m	3.9	2.1	0.2	1.8 - 4.1
	w	5.6	3.8	0.0	1.8 - 5.6
	-				
	s	3.5	1.7	0.3	1.8 - 3.8
corr 1.1 Å	m	4.4	2.6	0.2	1.8 - 4.6
	w	6.1	4.3	0.0	1.8 - 6.1
	-				
	s	3.5	1.7	0.3	1.8 - 3.8
corr 0.6 +	m	4.4	2.6	0.2	1.8 - 4.6
+ 0.5 inten CH3					
total 1.1 Å	w	6.1	4.3	0.0	1.8 - 6.1

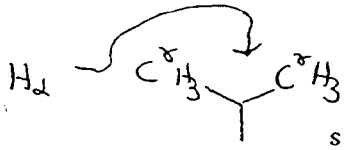
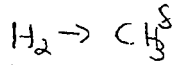
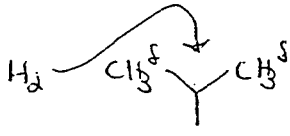
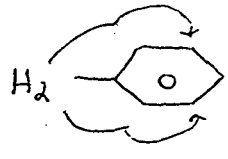
	s	3.5	1.7	0.3	1.8 - 3.8
corr 1.1 Å	m	4.4	2.6	0.2	1.8 - 4.6
	w	6.1	4.3	0.0	1.8 - 6.1
	s	3.9	2.1	0.3	1.8 - 4.2
corr = 1.0 + 0.5 inten	m	4.8	3.0	0.2	1.8 - 5.0
total 1.5 Å	w	6.5	4.7	0.0	1.8 - 6.5
	s	4.9	3.1	0.3	1.8 - 5.2
corr = 2.0 0.5 inten CH3	m	5.8	4.0	0.2	1.8 - 5.5
total 2.5 Å	w	7.5	5.7	0.0	1.8 - 7.5
	s	3.4	1.6	0.3	1.8 - 3.7
corr = 1.0	m	4.3	2.5	0.2	1.8 - 4.5
	w	6.0	4.2	0.0	1.8 - 6.0
	s	3.0	1.2	0.3	1.8 - 3.3
corr = 0.6	m	3.9	2.1	0.2	1.8 - 4.1
	w	5.6	3.8	0.0	1.8 - 5.6

Table E.2 NOE and H-bond restraints for Chapter 2

restraints

```
!noe_tbl.1
!NOE Constraints
! used for structures
! described in Chapter 2
! these constraints are an
! abbreviation of X-PLOR
! format which is:
! (resid N and name hX)
! (resid M and name hY)
! 4.3 2.5 0.2
!
!Residue 1 Residue 2
Constraint
!Constraint=NOE
        (+dist) (-dist)
! residue 1A

! residue 2R
2ha 3hd* 4.3 2.5 0.2
2 hn 3hd* 4.3 2.5 0.2

! residue 3P
3 hg* 3 hd* 5.3 3.5 0.2
3 hb* 3 hg* 5.3 3.5 0.2
3ha 4 hn 2.4 0.6 0.3
3ha 5 hn 5.0 3.2 0.0
3 hd* 6 hb* 4.3 2.5 0.2
3 hg* 6 hg2* 5.8 4.0 0.2
3 hd* 6 hg2* 5.8 4.0 0.2
3 hb* 63 he* 8.0 6.2 0.0
3 hg* 63 he* 6.3 4.5 0.2
3 hg* 63 hd* 8.0 6.2 0.0

! residue 4A
4 ha 4 hb* 3.9 2.1 0.3
4ha 5 hn 3.3 1.5 0.2
4 hn 5 hn 3.3 1.5 0.2
4hb* 5 hn 4.8 3.0 0.2
4ha 7 hn 5.0 3.2 0.0
4ha 7hb* 4.3 2.5 0.2

! residue 5F
5  hn 5 hd* 5.3 3.5 0.2
5 ha 5 hb* 4.3 2.5 0.2
5 ha 5 hd* 5.3 3.5 0.2
5 hb* 5 hd* 6.3 4.5 0.2
5ha 6 hn 5.0 3.2 0.0
5 hn 6 hn 3.3 1.5 0.2
5hb* 6 hn 4.3 2.5 0.2
5ha 8hb* 4.3 2.5 0.2
5 ha 48 he* 7.0 5.2 0.0
! residue6V
6 ha 6 hb 5.0 3.2 0.0
6 ha 6 hg1* 4.8 3.0 0.2
6 ha 6 hg2* 4.8 3.0 0.2
6 hb 6 hg2* 3.9 2.1 0.3
6  hn 5 hd* 7.0 5.2 0.0
6ha 7 hn 5.0 3.2 0.0
6 hn 7 hn 3.3 1.5 0.2
6hb* 7 hn 6.0 4.2 0.0
6ha 9hb1 3.3 1.5 0.2
6 ha 9 hb2 3.3 1.5 0.2
6 ha 9 hd* 6.2 4.4 0.2
6 hg2* 63 hd* 6.8 5.0 0.2

! residue 7N
7 ha 7 hb1 3.3 1.5 0.2
7 ha 7 hb2 3.3 1.5 0.2
7ha 8 hn 5.0 3.2 0.0
7 hn 8 hn 3.3 1.5 0.2
7hb* 8 hn 4.3 2.5 0.2
7ha 9 hn 5.0 3.2 0.0
7ha 10 hn 3.3 1.5 0.2
7ha 10hb* 4.3 2.5 0.2

! residue 8K
8 ha 8 hb* 4.3 2.5 0.2
8 ha 8 hg* 4.3 2.5 0.2
8ha 9 hn 5.0 3.2 0.0
8 hn 9 hn 3.3 1.5 0.2
8hb* 9 hn 6.0 4.2 0.0
8 hb* 47 ha 4.3 2.5 0.2
8hb* 47hd* 6.3 4.5 0.2
8 hb* 47 he* 8.0 6.2 0.0
8 hb* 48 he* 8.0 6.2 0.0
8 hd* 59 hg* 7.0 5.2 0.0

! residue 9L
9 ha 9 hb* 6.0 4.2 0.0
9 ha 9 hg 3.3 1.5 0.2
9 hb* 9 hg 4.3 2.5 0.2
9 ha 9 hd* 6.2 4.4 0.2
9 hb* 9 hd* 7.2 5.4 0.3
9 hg 9 hd* 6.2 4.4 0.
9ha 10 hn 5.0 3.2 0.0
9 hn 10 hn 3.3 1.5 0.2
9hb* 10 hn 4.3 2.5 0.2
9 hd* 10  hn 7.9 6.1 0.0
9ha 12hb* 6.0 4.2 0.0
9 hd* 12 he* 7.2 5.4 0.2
9 hd* 31 hd* 7.2 5.4 0.2
9 hg 10  hn 5.0 3.2 0.0
9 hd* 53 he* 9.9 8.1 0.0
9 hd* 56 he* 8.2 6.4 0.2
9 hd* 56 hz 6.2 4.4 0.2
9 hd* 60 hd* 8.6 6.8 0.2

! residue 10W
10 ha 10 hb* 4.3 2.5 0.2
10 ha 10 he3 3.3 1.5 0.2
10  hn 10 hd1 5.0 3.2 0.0
10 hn 11 hn 3.3 1.5 0.2
10hb* 11 hn 4.3 2.5 0.2
10ha 13 hn 5.0 3.2 0.0
10ha 13hb 4.3 2.5 0.2
10 ha 13 hg1* 6.5 4.7 0.0
10 ha 13 hg2* 4.8 3.0 0.2
10 hd1 10  hn 5.0 3.2 0.0
10 hd1 10 ha 5.0 3.2 0.0
10 hd1 10 hb* 4.3 2.5 0.2
10 ha 24 hz2 5.0 3.2 0.0

! residue 11S
11ha 12 hn 3.3 1.5 0.2
11ha 14 hn 5.0 3.2 0.0

! residue 12M
12ha 13 hn 5.0 3.2 0.0
12 hn 13 hn 5.0 3.2 0.0
12hb* 13 hn 4.3 2.5 0.2
12ha 14 hn 5.0 3.2 0.0
12ha 15 hn 5.0 3.2 0.0
12 he* 44 ha 4.8 3.0 0.2
12 hb* 47 hd* 8.0 6.2 0.0
12 hb* 47 he* 8.0 6.2 0.0
12 hg* 47 he* 8.0 6.2 0.0

! residue 13V
13 ha 13 hb 3.3 1.5 0.2
13 ha 13 hg1* 4.8 3.0 0.2
13 hb 13 hg1* 4.8 3.0 0.2
13 hb 13 hg2* 4.8 3.0 0.2
13ha 14 hn 5.0 3.2 0.0
13 hn 14 hn 3.3 1.5 0.2
13hb* 14 hn 3.3 1.5 0.2
13 ha 22 hg2* 4.8 3.0 0.2
13 hg1* 22 hg2* 5.8 4.0
0.2
```

13 hg1* 22 hd1* 7.5 5.7
0.0
13 hg2* 22 hg2* 5.8 4.0
0.2
13 hg2* 22 hd* 5.8 4.0 0.2
13 hg1* 23 ha 4.8 3.0 0.2
13 hg1* 24 hn 4.8 3.0 0.2
13 hb(resid 24 he1 5.0 3.2
0.0
13 hg2* 24 he1 4.8 3.0 0.2
13 hg1* 31 hg2* 7.5 5.7
0.0
13 hg1* 31 hd* 5.8 4.0 0.2
13 hg1* 32 hn 6.5 4.7 0.0

! residue 14N
14 ha 14 hb1 2.4 0.6 0.3
14 ha 14 hb2 2.4 0.6 0.3
14 hn 15 hn 3.3 1.5 0.2
14hb* 15 hn 4.3 2.5 0.2
14 ha 24 he1 3.3 1.5 0.2

! residue 15D
15ha 16 hn 3.3 1.5 0.2
15 hb* 22 hg2* 5.8 4.0 0.2

! residue 16K
16 ha 16 hb* 3.4 1.6 0.3
16 ha 16 hg* 4.3 2.5 0.2
16 ha 16 hd* 6.0 4.2 0.0
16 hg* 16 hd* 4.4 2.6 0.3
16 hg* 16 he* 5.3 3.5 0.2
16 hb* 16 hg* 4.4 2.6 0.3
16 hd* 16 he* 4.4 2.6 0.3
16ha 17 hn 3.3 1.5 0.2
16 hn 17 hn 3.3 1.5 0.2
16hb* 17 hn 4.3 2.5 0.2
16ha 19 hn 5.0 3.2 0.0
16ha 19hb1 5.0 3.2 0.0
16 ha 19 hb2 5.0 3.2 0.0

! residue 17S
17 ha 17 hb1 2.4 0.6 0.3
17 ha 17 hb2 2.4 0.6 0.3
17hb* 18 hn 3.3 1.5 0.2
17ha 19 hn 5.0 3.2 0.0

! residue 18N
18ha 19 hn 3.3 1.5 0.2
18ha 20 hn 5.0 3.2 0.0
18ha 21 hn 5.0 3.2 0.0
18ha 21hd* 5.3 3.5 0.2
18 hb* 39 hd* 8.0 6.2 0.0
18 ha 43 hg* 6.2 4.4 0.2

! residue 19E
19ha 20 hn 3.3 1.5 0.2
19 hn 20 hn 5.0 3.2 0.0
19hb* 20 hn 4.3 2.5 0.2
! residue20K
20 ha 20 hb* 3.4 1.6 0.3
20 ha(resid 20 hd* 6.0 4.2
0.0
20 hd* 20 he* 4.4 2.6 0.3
20 hb* 20 hd* 5.3 3.5 0.2
20ha 21 hn 5.0 3.2 0.0
20 hn 21 hn 3.3 1.5 0.2
20 hb* 21 hn 6.0 4.2 0.0

! residue 21F
21 hn 21 hd* 5.3 3.5 0.2
21 ha 21 hd* 5.3 3.5 0.2
21 ha 21 hb* 6.0 4.2 0.0
21 hb* 21 hd* 6.3 4.5 0.2
21ha 22 hn 5.0 3.2 0.0
21hb* 22 hn 4.3 2.5 0.2
21 ha 34 hd* 6.0 4.2 0.0
21 hb* 34 hg* 5.3 3.5 0.2
21 hd* 39 ha 5.3 3.5 0.2
21 hd* 43 hg* 9.9 8.1 0.0

! residue 22I
22 ha 22 hb 3.3 1.5 0.2
22 ha 22 hg2* 4.8 3.0 0.2
22 ha 22 hd* 4.8 3.0 0.2
22 hb 22 hd* 4.8 3.0 0.2
22 hb 22 hg2* 3.9 2.1 0.3
22 hg2* 22 hd* 4.9 3.1 0.3
22 hn 21 hd* 7.0 5.2 0.0
22 ha 23 hn 2.4 0.6 0.3
22 hb 23 hn 5.0 3.2 0.0
22 hd* 23 hn 6.5 4.7 0.0
22 ha 31 hg2* 4.8 3.0 0.2
22 hd* 31 hd* 5.8 4.0 0.2
22 ha 33ha 3.3 1.5 0.2
22 ha 33 hg2* 6.5 4.7 0.0
22 hb 33 hg2* 6.5 4.7 0.0
22 hd* 33 hg2* 4.9 3.1 0.3
22 ha 34 hd* 4.3 2.5 0.2
22 hd* 39 hd* 8.5 6.7 0.0

! residue 23H
23 ha 24 hn 2.4 0.6 0.3
23 hb* 24 hn 4.3 2.5 0.2
23 hn 32 hn 3.3 1.5 0.2
23 hn 33 ha 3.3 1.5 0.2

! residue 24W
24 hn 24 hd1 3.3 1.5 0.2
24 hn 24 hd1 5.0 3.2 0.0
24 ha 24 hb1 3.3 1.5 0.2
24 ha 24 hb2 3.3 1.5 0.2
24 ha 24 he3 3.3 1.5 0.2
24 hb1 24 hd1 3.3 1.5 0.2
24 hb2 24 hd1 3.3 1.5 0.2
24 ha 25 hn 2.4 0.6 0.3
24 hb* 25 hn 4.3 2.5 0.2
24 ha 32 hn 5.0 3.2 0.0
24 ha 31ha 3.3 1.5 0.2

! residue 25S
25 ha 25 hb* 4.3 2.5 0.2
25 ha 26 hn 3.3 1.5 0.2
25 hn 26 hn 5.0 3.2 0.0
25 hb* 26 hn 6.0 4.2 0.0
25 ha 32 hg2* 4.8 3.0 0.2
25 hb1 32 hg2* 4.8 3.0 0.2
25 hb2 32 hg2* 4.8 3.0 0.2
25 hn 31 ha 5.0 3.2 0.0

! residue 26T
26 hg* 27 hn 6.5 4.7 0.0

! residue 27S

! residue 28G
28 ha* 29 hn 5.0 3.2 0.0

! residue 29E
29 ha 29 hg* 3.4 1.6 0.3
29 hb* 29 hg* 4.4 2.6 0.3
29 ha 30 hn 5.0 3.2 0.0
29 hn 30 hn 3.3 1.5 0.2

! residue 30S
30 ha 31 hn 2.4 0.6 0.3
30hb* 31 hn 6.0 4.2 0.0
30 ha 87ha 3.3 1.5 0.2
30 hb* 87 ha 4.3 2.5 0.2
30 hb* 87hg* 7.0 5.2 0.0

! residue 31I
31 ha 31 hb* 6.0 4.2 0.0
31 ha 31 hg2* 4.8 3.0 0.2
31 hb 31 hd* 4.8 3.0 0.2
31 hb 31 hg2* 4.8 3.0 0.2

31 hg2* 31 hd* 4.9 3.1 0.3
31ha 32 hn 2.4 0.6 0.3
31 hg2* 32 hn 4.8 3.0 0.2
31 hd* 60 hd* 8.9 7.1 0.0
31 hd* 65 he1 6.5 4.7 0.0
31 hn 86 hn 3.3 1.5 0.2
31 hb* 86 hb* 5.3 3.5 0.2
31 hb* 86 hd* 6.3 4.5 0.2
31 hn 87 ha 5.0 3.2 0.0

! residue 32V
32 ha 32 hb 3.3 1.5 0.2
32 ha 32 hg1* 3.9 2.1 0.3
32 ha 32 hg2* 3.9 2.1 0.3
32 hb 32 hg1* 3.9 2.1 0.3
32 hb 32 hg2* 3.9 2.1 0.3
32ha 33 hn 2.4 0.6 0.3
32 hg1* 34 ha 4.8 3.0 0.2
32 hg1* 83 hd1* 5.8 4.0
0.2
32ha 85ha 3.3 1.5 0.2
32 hg1* 34 ha 6.5 4.7 0.0
32 hg1* 34 hd* 7.5 5.7 0.0
32 hg1* 34 hg* 5.8 4.0 0.2
32 hg1* 33 hn 6.5 4.7 0.0
32 hg1* 34 ha 4.8 3.0 0.2
32 hg2* 85 ha 4.8 3.0 0.2
32 hg2* 85 ha 6.5 4.7 0.0

! residue 33V
33 ha 33 hg1* 4.8 3.0 0.2
33 ha 33 hg2* 4.8 3.0 0.2
33 hb 33 hg1* 4.8 3.0 0.2
33 hb 33 hg2* 3.9 2.1 0.3
33ha 34hd1 3.3 1.5 0.2
33 ha 34 hd2 3.3 1.5 0.2
33 ha 34 hg* 4.3 2.5 0.2
33 hg1* 35 hn 4.8 3.0 0.2
33 hg1* 36 hn 6.5 4.7 0.0
33 hg1* 36 ha 4.8 3.0 0.2
33 hb 84 hn 3.3 1.5 0.2
33 hb 84 hb* 4.3 2.5 0.2
33 hn 85 ha 3.3 1.5 0.2
33 hg2* 86 hd* 8.5 6.7 0.0
33 hg1* 37 hn 4.8 3.0 0.2
33 hg1* 84 he1 6.5 4.7 0.0
33 hg2* 84 hd1 6.5 4.7 0.0
33 hg1* 39 hn 6.5 4.7 0.0
33 hg1* 39 hd* 8.5 6.7 0.0

! residue 34P
34 hg* 34 ha 6.0 4.2 0.0
34 hg* 34 hd1 4.3 2.5 0.2
34 hg* 34 hd2 4.3 2.5 0.2
34hd* 35 hn 6.0 4.2 0.0
34hb* 35 hn 4.3 2.5 0.2
34 ha 83 ha 5.0 3.2 0.0

! residue 35N
35 ha 35 hb* 4.3 2.5 0.2
35ha 36 hn 2.4 0.6 0.3
35hb* 36 hn 6.0 4.2 0.0

! residue 36R
36 hg1 36 ha 5.0 3.2 0.0
36 hg2 36 ha 5.0 3.2 0.0
36 hn 37 hn 5.0 3.2 0.0
36hb* 37 hn 3.4 1.6 0.3
36ha 39 hn 5.0 3.2 0.0
36 hn 82 ha 5.0 3.2 0.0
36 hn 82 hb* 6.0 4.2 0.0
36 hb* 82 ha 6.0 4.2 0.0
36 hd* 84 he1 4.3 2.5 0.2
36 hb* 84 he1 3.4 1.6 0.3
36 hg* 40 hn 6.0 4.2 0.0

! residue 37E
37 ha 37 hb* 4.3 2.5 0.2
37 hb1 37 hg* 3.4 1.6 0.3
37 hb2 37 hg* 3.4 1.6 0.3
37ha 38 hn 3.3 1.5 0.2
37 hn 38 hn 3.3 1.5 0.2
37hb* 38 hn 4.3 2.5 0.2
37ha 40 hb 5.0 3.2 0.0
37 hn 82 hb* 6.0 4.2 0.0

! residue 38R
38 hn 39 hn 3.3 1.5 0.2
38hb* 39 hn 4.3 2.5 0.2
38ha 41hb* 4.3 2.5 0.2
38 ha 42 hn 3.3 1.5 0.2

! residue 39F
39 hn 39 hd* 5.3 3.5 0.2
39 ha 39 hd* 5.3 3.5 0.2
39 hb* 39 hd* 6.3 4.5 0.2
39ha 40 hn 3.3 1.5 0.2
39 hn 40 hn 3.3 1.5 0.2
39hb* 40 hn 4.3 2.5 0.2
39 hb* 40 hg1* 7.5 5.7 0.0
39 hd* 40 hg1* 8.5 6.7 0.0
39 hd* 40 hg2* 6.8 5.0 0.2
39ha 42hb* 6.0 4.2 0.0
39 hd* 44 hd* 9.9 8.1 0.0

! residue 40V

40 ha 40 hb 3.3 1.5 0.2
40 ha 40 hg1* 3.9 2.1 0.3
40 ha 40 hg2* 3.9 2.1 0.3
40 hb 40 hg1* 3.9 2.1 0.3
40 hb 40 hg2* 3.9 2.1 0.3
40 hg2* 37 ha 4.8 3.0 0.2
40 hn 39 hd* 7.0 5.2 0.0
40 hn 41 hn 3.3 1.5 0.2
40 hb 41 hn 4.3 2.5 0.2
40 ha 43hb 5.0 3.2 0.0
40 hg1* 41 hn 6.5 4.7 0.0
40 ha 44 hn 3.3 1.5 0.2
40 ha 44 hb1 3.3 1.5 0.2
40 ha 44 hb2 3.3 1.5 0.2
40 ha 44 hd* 6.2 4.4 0.2
40 hg1* 44 hd* 8.9 7.1 0.0
40 hg2* 44 hd* 7.2 5.4 0.2
40 hg1* 41 ha 4.8 3.0 0.2
40 hg2* 41 hn 6.5 4.7 0.0
40 ha 53 hd* 7.0 5.2 0.0
40 hg2* 53 ha 6.5 4.7 0.0
40 hg2* 56 hb* 7.5 5.7 0.0

! residue 41Q
41 ha 41 hg1 3.3 1.5 0.2
41 ha 41 hg2 3.3 1.5 0.2
41 hb* 41 hg* 4.4 2.6 0.3
41 hn 39 hd* 7.0 5.2 0.0
41ha 42 hn 3.3 1.5 0.2
41hb* 42 hn 4.3 2.5 0.2

! residue 42E
42 ha 42 hb1 3.3 1.5 0.2
42 ha 42 hb2 3.3 1.5 0.2
42 ha 42 hg* 3.4 1.6 0.3
42 hb* 42 hg* 5.2 3.5 0.2
42 hn 43 hn 2.4 0.6 0.3
42 hb* 43 ha 4.3 2.5 0.2
42 hb* 43 hg1* 5.8 4.0 0.2

! residue 43V
43 ha 43 hb 5.0 3.2 0.0
43 ha 43 hg* 6.2 4.4 0.2
43 hb 43 hg2* 3.9 2.1 0.3
43ha 44 hn 3.3 1.5 0.2
43 hn 44 hn 2.4 0.6 0.3
43 hg* 44 hn 7.9 6.1 0.0
43ha 45hd* 6.0 4.2 0.0
43ha 46 hn 3.3 1.5 0.2
43ha 46hb* 6.0 4.2 0.0

43hg2* 47hd* 5.9 4.1 0.3

! residue 44L
44 ha 44 hb1 3.3 1.5 0.2
44 ha 44 hb2 3.3 1.5 0.2
44 ha 44 hg 5.0 3.2 0.0
44 ha 44 hd* 7.9 6.1 0.0
44 hb* 44 hd* 7.2 5.4 0.3
44 hg 44 hd* 6.2 4.4 0.2
44 hn 45hd* 3.4 1.6 0.3
44hb* 45hd1 4.3 2.5 0.2
44 hb* 45 hd2 4.3 2.5 0.2
44ha 47hd* 4.4 2.6 0.3
44 hd* 51 hb* 7.2 5.4 0.2
44 hd* 52 ha 7.9 6.1 0.0
44 hd* 52 hb* 8.9 7.1 0.0
44 hd* 53 ha 7.9 6.1 0.0
44 hd* 53 hb* 8.9 7.1 0.0
44 hd* 53 hn 7.9 6.1 0.0
44 hd* 56 hb* 7.2 5.4 0.2
44 hd* 56 hd* 9.9 8.1 0.0
44 hd* 48 hb* 8.9 7.1 0.0
44 ha 48 hd* 7.0 5.2 0.0
44 hd* 48 hb* 8.9 7.1 0.0
44 hd* 48 hd* 9.9 8.1 0.0

! residue 45P
45ha 46 hn 5.0 3.2 0.0
45hd* 46 hn 3.4 1.6 0.3
45ha 48 hn 3.3 1.5 0.2
45 ha 49 ha 3.3 1.5 0.2
! residue46K
46 ha 46 hg1 3.3 1.5 0.2
46 ha 46 hg2 3.3 1.5 0.2
46 hb* 46 hg* 7.0 5.2 0.0
46 hg* 46 he* 7.0 5.2 0.0
46ha 47 hn 5.0 3.2 0.0
46 hn 47 hn 2.4 0.6 0.3
46hb* 47 hn 6.0 4.2 0.0
46 hb* 47 hd* 8.0 6.2 0.0
46 hb* 47 he* 8.0 6.2 0.0

! residue 47Y
47 hn 47 hd* 7.0 5.2 0.0
47 ha 47 hb* 4.3 2.5 0.2
47 hb1 47 hd* 5.3 3.5 0.2
47 hb2 47 hd* 5.3 3.5 0.2
47ha 48 hn 5.0 3.2 0.0
47 hn 48 hn 2.4 0.6 0.3
47 ha 48 he* 5.3 3.5 0.2
47hb* 48 hn 6.0 4.2 0.0
47 hb* 48 hd* 8.0 6.2 0.0
47hd* 47 hn 5.3 3.5 0.2
47hd* 47ha 5.3 3.5 0.2
! residue 48F
48 hn 47 hd* 7.0 5.2 0.0
48 hn 48 hd* 7.0 5.2 0.0
48 ha 48 hb* 4.3 2.5 0.2
48 ha 48 hd* 5.3 3.5 0.2
48 hb1 48 hd* 5.3 3.5 0.2
48 hb2 48 hd* 5.3 3.5 0.2
48ha 49 hn 3.3 1.5 0.2
48hb* 49 hn 6.0 4.2 0.0
48ha 51hb* 6.0 4.2 0.0
48 hb1 51 hb* 4.3 2.5 0.2
48 hb2 51 hb* 4.3 2.5 0.2
48 hb* 51 ha 4.3 2.5 0.2

! residue 49K
49 ha 49 hb1 3.3 1.5 0.2
49 ha 49 hb2 3.3 1.5 0.2
49 ha 49 hd* 6.0 4.2 0.0
49 hg* 49 ha 4.3 2.5 0.2
49 hg* 49 hb* 5.3 3.5 0.2
49 hg* 49 hd* 4.4 2.6 0.3
49ha 50 hn 2.4 0.6 0.3
49 hn 50 hn 5.0 3.2 0.0
49hb* 50 hn 4.3 2.5 0.0
49ha 51 hn 5.0 3.2 0.0

! residue 50H
50 hn 51 hn 3.3 1.5 0.2
! residue 51S
51 ha 51 hb* 3.4 1.6 0.3

! residue 52N
52 ha 52 hb* 4.3 2.5 0.2
52ha 53 hn 2.4 0.6 0.3
52hb* 53 hn 4.3 2.5 0.2
52ha 55 hn 3.3 1.5 0.2

! residue 53F
53 hn 53 hd* 5.3 3.5 0.2
53 ha 53 hd* 5.3 3.5 0.2
53 hb* 53 hd* 6.3 4.5 0.2
53ha 54 hn 3.3 1.5 0.2
53 hn 54 hn 3.3 1.5 0.2
53hb* 54 hn 4.3 2.5 0.2

! residue 54A
54 hn 53 hd* 7.0 5.2 0.0
54 hn 55 hn 3.3 1.5 0.2
54ha 57hb 3.3 1.5 0.2
54 ha 57 hg1* 6.5 4.7 0.0
54 ha 57 hg2* 4.8 3.0 0.2

! residue 55S
55 hn 56 hn 3.3 1.5 0.2
55hb* 56 hn 4.3 2.5 0.2
55ha 58hb* 4.3 2.5 0.2
55 ha 55 hb* 6.0 4.2 0.0

! residue 56F
56 hn 56 hd* 5.3 3.5 0.2
56 ha 56 hd* 5.3 3.5 0.2
56 hb1 56 hd* 5.3 3.5 0.2
56 hb2 56 hd* 5.3 3.5 0.2
56ha 57 hn 5.0 3.2 0.0
56 hn 57 hn 3.3 1.5 0.2
56hb* 57 hn 4.3 2.5 0.2
56 he* 57 hg2* 8.5 6.7 0.0
56ha 59hb* 4.3 2.5 0.2
56 ha 60 hn 3.3 1.5 0.2
56he* 60 hd* 9.9 8.1 0.0
56 hz 60 hd* 7.9 6.1 0.0

! residue 57V
57 ha 57 hg1* 3.9 2.1 0.3
57 ha 57 hg2* 3.9 2.1 0.3
57 hb 57 hg1* 3.9 2.1 0.3
57 hb 57 hg2* 3.9 2.1 0.3
57 hn 56 hd* 7.0 5.2 0.0
57ha 58 hn 5.0 3.2 0.0
57 hn 58 hn 3.3 1.5 0.2
57 hg1* 58 hn 6.5 4.7 0.0
57ha 60 hn 3.3 1.5 0.2
57 ha 60 hd* 7.9 6.1 0.0
57 hb* 60 hn 4.3 2.5 0.2
57 hg1* 60 hd* 8.9 7.1 0.0
57 hg2* 60 hd* 7.2 5.4 0.2
57 hg2* 58 hn 6.5 4.7 0.0
57 hg1* 84 hz3 4.8 3.0 0.2
57 hg2* 84 hz3 6.5 4.7 0.0

! residue 58R
58 ha 58 hb* 3.4 1.6 0.3
58ha 59 hn 5.0 3.2 0.0
58 hn 59 hn 3.3 1.5 0.2
58ha 61 hn 3.3 1.5 0.2
58ha 61hb* 3.4 1.6 0.3
58 ha 62 hn 3.3 1.5 0.2

! residue 59Q
59 hg1 60 ha 3.3 1.5 0.2
59 hg2 60 ha 3.3 1.5 0.2
59 hn 60 hn 3.3 1.5 0.2

59hb* 60 hn 4.3 2.5 0.2
59 hg* 60 hn 4.3 2.5 0.2
59ha 62hb1 5.0 3.2 0.0
59 ha 62 hb2 5.0 3.2 0.0
59 hb* 63 he* 6.3 4.5 0.2

! residue 60L
60 ha 60 hd* 7.9 6.1 0.0
60 ha 60 hb* 4.3 2.5 0.2
60 hb1 60 hg 5.0 3.2 0.0
60 hb2 60 hg 5.0 3.2 0.0
60 hd* 60 hb1 6.2 4.4 0.2
60 hd* 60 hb2 6.2 4.4 0.2
60 hg 60 hd* 7.9 6.1 0.0
60 ha 61 hn 5.0 3.2 0.0
60 hn 61 hn 3.3 1.5 0.2
60 hb* 61 hn 4.3 2.5 0.2
60 ha 63hb* 4.3 2.5 0.2
60 ha 63 hd* 5.3 3.5 0.2
60 hd* 63 hd* 9.9 8.1 0.0
60 hb* 65 hb* 5.3 3.5 0.2
60 hd* 65 hb1 8.9 7.1 0.0
60 hd* 65 hb2 8.9 7.1 0.0

! residue 61N
61ha 62 hn 5.0 3.2 0.0
61 hn 62 hn 3.3 1.5 0.2
61hb* 62 hn 4.3 2.5 0.2
61ha 64 hn 5.0 3.2 0.0
! residue 62M
62 ha 62 hb* 4.3 2.5 0.2
62 ha 62 hg* 4.3 2.5 0.2
62 hn 63 hn 2.4 0.6 0.3
62ha 64 hn 5.0 3.2 0.0

! residue 63Y
63 ha 63 hb* 6.0 4.2 0.0
63 hb1 63 hd* 5.3 3.5 0.2
63 hb2 63 hd* 5.3 3.5 0.2
63ha 64 hn 5.0 3.2 0.0
63 hn 64 hn 3.3 1.5 0.2
63hd* 63ha 5.3 3.5 0.2

! residue 64G
64ha* 65 hn 5.0 3.2 0.0
64 hn 65 hn 3.3 1.5 0.2

! residue 65W
65 hn 65 hd1 5.0 3.2 0.0
65ha 66 hn 2.4 0.6 0.3
65hb* 66 hn 6.0 4.2 0.0
65hd1 65 hn 5.0 3.2 0.0
65hd1 65hb* 6.0 4.2 0.0
65ha 88ha 3.3 1.5 0.2
! residue66H
66 ha 66 hb1 3.3 1.5 0.2
66 ha 66 hb2 3.3 1.5 0.2
66ha 67 hn 2.4 0.6 0.3
66 hn 67 hn 2.4 0.6 0.3
66hb* 67 hn 4.3 2.5 0.2
66 hn 87 hn 5.0 3.2 0.0
66 hn 88 ha 5.0 3.2 0.0

! residue 67K
67 ha 67 hb* 4.3 2.5 0.2
67ha 68 hn 3.3 1.5 0.2
67ha 87 hn 5.0 3.2 0.0
67ha 86ha 3.3 1.5 0.2

! residue 68V
68ha 69 hn 2.4 0.6 0.3
68 hn 69 hn 5.0 3.2 0.0
68 hg* 69 hn 5.3 3.5 0.3
68 hg* 87 hn 6.2 4.4 0.2

! residue 69Q
69 ha 69 hb1 3.3 1.5 0.2
69 ha 69 hb2 3.3 1.5 0.2
69 hb* 69 hg* 4.4 2.6 0.3

! residue 70D
70 ha 71 hn 3.3 1.5 0.2
70 hn 71 hn 5.0 3.2 0.0

! residue 71V
71 ha 71 hb 2.4 0.6 0.3
71 hb 71 hg1* 4.8 3.0 0.2
71ha 72 hn 3.3 1.5 0.2
71hb 72 hn 5.0 3.2 0.0

! residue 72K
72 ha 72 hb* 3.4 1.6 0.3
72 hg* 72 hd* 4.4 2.6 0.3
72ha 73 hn 3.3 1.5 0.2
72hb* 73 hn 6.0 4.2 0.0

! residue 73S
73ha 74 hn 3.3 1.5 0.2
73 hn 74 hn 3.3 1.5 0.2

! residue 74G
74 hn 75 hn 3.3 1.5 0.2

! residue75S
75 hn 76 hn 3.3 1.5 0.2
75hb* 76 hn 6.0 4.2 0.0

! residue 76M
76 ha 76 hb1 3.3 1.6 0.2
76 ha 76 hb2 3.3 1.6 0.2
76ha 77 hn 3.3 1.5 0.2
76 hn 77 hn 3.3 1.5 0.2
! residue77L
77 ha 77 hb* 4.3 2.5 0.2
77 ha 77 hg 3.3 1.5 0.2
77 ha 77 hd* 6.2 4.4 0.2
77 hb* 77 hd* 7.2 5.4 0.2
77 hb* 77 hd* 6.3 4.5 0.3
77 hg* 77 hd* 5.3 3.5 0.3
77ha 78 hn 3.3 1.5 0.2
77hb* 78 hn 4.3 2.5 0.2
77 hd* 78 hn 6.2 4.4 0.2

! residue 78S
78ha 79 hn 3.3 1.5 0.2
78 hn 79 hn 3.3 1.5 0.2
78hb* 79 hn 6.0 4.2 0.0

! residue 79N

! residue 80N
80 ha 81 hn 3.3 1.5 0.2

! residue 81D

! residue 82S
82 ha 82 hb* 3.4 1.6 0.3
82ha 83 hn 3.3 1.5 0.2
82 hn 83 hn 2.4 0.6 0.3
82hb* 83 hn 4.3 2.5 0.2

! residue 83R
83 ha 83 hb* 4.3 2.5 0.2
83ha 84 hn 3.3 1.5 0.2
83 hb* 83 hd* 5.3 3.5 0.2
83hb* 84 hn 4.3 2.5 0.2

! residue 84W
84 hn 84 hd1 5.0 3.2 0.0
84 hb* 84 he3 4.3 2.5 0.2
84ha 85 hn 2.4 0.6 0.3
84hb* 85 hn 4.3 2.5 0.2
84hd1 84 hn 3.3 1.5 0.2
84hd1 84ha 3.3 1.5 0.2
84hd1 84hb* 4.3 2.5 0.2

```
! residue 85E
85 ha 85 hb* 4.3 2.5 0.2
85 ha 85 hg* 4.3 2.5 0.2
85 hb1 85 hg* 4.3 2.5 0.2
85 hb2 85 hg* 4.3 2.5 0.2
85ha 86 hn 2.4 0.6 0.3
85hb* 86 hn 6.0 4.2 0.0

! residue 86F
86  hn 86 hd* 5.3 3.5 0.2
86 ha 86 hb* 6.0 4.2 0.0
86 ha 86 hd* 5.3 3.5 0.2
86 hb* 86 hd* 6.3 4.5 0.2
86ha 87 hn 2.4 0.6 0.3
86hb* 87 hn 6.0 4.2 0.0

! residue 87E
87 ha 87 hb* 4.3 2.5 0.2
87ha 88 hn 2.4 0.6 0.3
87hb* 88 hn 6.0 4.2 0.0

! residue 88N
88 ha 89 hb1 5.0 3.2 0.0
88 ha 89 hb2 5.0 3.2 0.0

! residue 89E
89 hb* 89 hg* 4.4 2.6 0.3
89ha 90 hn 3.3 1.5 0.2
89hn 90 hn 3.3 1.5 0.2
89hb* 90 hn 6.0 4.2 0.0

! residue 90R
90 ha 91 hn 3.3 1.5 0.2
90 hb* 91 hn 4.3 2.5 0.2

! residue 91H
91 ha 91 hb* 4.3 2.5 0.2
91ha 92 hn 3.3 1.5 0.2

! residue 92A
92 ha 92 hb* 4.8 3.0 0.2

!Hydrogen bond
constraints
!Residue 1 Residue 2
Constraints
!distance (+distance) (-
distance)
7 hn 3 o 2.1 0.3 0.2
7 n 3 o 3.1 0.3 0.2
8 hn 4 o 2.1 0.3 0.2
8 n 4 o 3.1 0.3 0.2
9 hn 5 o 2.1 0.3 0.2
9 n 5 o 3.1 0.3 0.2
13 hn 9 o 2.1 0.3 0.2
13 n 9 o 3.1 0.3 0.2
31 hn 86 o 2.1 0.3 0.2
31 n 86 o 3.1 0.3 0.2
32 hn 23 o 2.1 0.3 0.2
32 n 23 o 3.1 0.3 0.2
33 hn 84 o 2.1 0.3 0.2
33 n 84 o 3.1 0.3 0.2
56 hn 52 o 2.1 0.3 0.2
56 n 52 o 3.1 0.3 0.2
57 hn 53 o 2.1 0.3 0.2
57 n 53 o 3.1 0.3 0.2
58 hn 54 o 2.1 0.3 0.2
58 n 54 o 3.1 0.3 0.2
59 hn 55 o 2.1 0.3 0.2
59 n 55 o 3.1 0.3 0.2
60 hn 56 o 2.1 0.3 0.2
60 n 56 o 3.1 0.3 0.2
61 hn 57 o 2.1 0.3 0.2
61 n 57 o 3.1 0.3 0.2
62 hn 58 o 2.1 0.3 0.2
62 n 58 o 3.1 0.3 0.2
66 hn 87 o 2.1 0.3 0.2
66 n 87 o 3.1 0.3 0.2
```

Table E.3 Dihedral restraints for Chapter 2

```
restraints
! dihe_tbl
!Dihedral Restraints
! used in calculation of structures described in Chapter 2

! Torsion psi constraints

!Residue 1 Residue 2 Residue 3 Residue 4 Constraint
!        Constraint = Force     Angle  (+/- Angle)     Exponent
!                     Constant        Value            Range                    Potential
!!    residue 4  A
  assign (resid 3 and name c) (resid 4 and name n)
       (resid 4 and name ca) (resid 4 and name c) 1.0 -60.0 30.0 2
!!    residue 12  M
  assign (resid 11 and name c) (resid 12 and name n)
       (resid 12 and name ca) (resid 12 and name c) 1.0 -60.0 30.0 2
!!    residue 13  V
  assign (resid 12 and name c) (resid 13 and name n)
       (resid 13 and name ca) (resid 13 and name c) 1.0 -60.0 30.0 2
!!    residue 14  N
  assign (resid 13 and name c) (resid 14 and name n)
       (resid 14 and name ca) (resid 14 and name c) 1.0 -60.0 30.0 2
!!    residue 16  K
  assign (resid 15 and name c) (resid 16 and name n)
       (resid 16 and name ca) (resid 16 and name c) 1.0 -60.0 30.0 2
!!    residue 19  E
  assign (resid 18 and name c) (resid 19 and name n)
       (resid 19 and name ca) (resid 19 and name c) 1.0 -60.0 30.0 2
!!    residue 21  F
  assign (resid 20 and name c) (resid 21 and name n)
       (resid 21 and name ca) (resid 21 and name c) 1.0 -120 40.0 2
!!    residue 24  W
  assign (resid 23 and name c) (resid 24 and name n)
       (resid 24 and name ca) (resid 24 and name c) 1.0 -60.0 30.0 2
!!    residue 25  A
  assign (resid 24 and name c) (resid 25 and name n)
       (resid 25 and name ca) (resid 25 and name c) 1.0 -60.0 30.0 2
!!    residue 31  I
  assign (resid 30 and name c) (resid 31 and name n)
       (resid 31 and name ca) (resid 31 and name c) 1.0 -120.0 40.0 2
!!    residue 32  V
  assign (resid 31 and name c) (resid 32 and name n)
       (resid 32 and name ca) (resid 32 and name c) 1.0 -120.0 40.0 2
!!    residue 33  V
  assign (resid 32 and name c) (resid 33 and name n)
       (resid 33 and name ca) (resid 33 and name c) 1.0 -120.0 40.0 2
!!    residue 35  N
  assign (resid 34 and name c) (resid 35 and name n)
       (resid 35 and name ca) (resid 35 and name c) 1.0 -120.0 40.0 2
!!    residue 36  R
```

```
  assign (resid 35 and name c) (resid 36 and name n)
       (resid 36 and name ca) (resid 36 and name c) 1.0 -60.0 30.0 2
!!    residue 37  E
  assign (resid 36 and name c) (resid 37 and name n)
       (resid 37 and name ca) (resid 37 and name c) 1.0 -60.0 30.0 2
!!    residue 39  F
  assign (resid 38 and name c) (resid 39 and name n)
       (resid 39 and name ca) (resid 39 and name c) 1.0 -60.0 30.0 2
!!    residue 41  Q
  assign (resid 40 and name c) (resid 41 and name n)
       (resid 41 and name ca) (resid 41 and name c) 1.0 -60.0 30.0 2
!!    residue 43  V
  assign (resid 42 and name c) (resid 43 and name n)
       (resid 43 and name ca) (resid 43 and name c) 1.0 -120.0 40.0 2
!!    residue 47  Y
  assign (resid 46 and name c) (resid 47 and name n)
       (resid 47 and name ca) (resid 47 and name c) 1.0 -120.0 40.0 2
!!    residue 48  F
  assign (resid 47 and name c) (resid 48 and name n)
       (resid 48 and name ca) (resid 48 and name c) 1.0 -120.0 40.0 2
!!    residue 51  S
  assign (resid 50 and name c) (resid 51 and name n)
       (resid 51 and name ca) (resid 51 and name c) 1.0 -120.0 40.0 2
!!    residue 53  F
  assign (resid 52 and name c) (resid 53 and name n)
       (resid 53 and name ca) (resid 53 and name c) 1.0 -60.0 30.0 2
!!    residue A  54
  assign (resid 53 and name c) (resid 54 and name n)
       (resid 54 and name ca) (resid 54 and name c) 1.0 -60.0 30.0 2
!!    residue F  56
  assign (resid 55 and name c) (resid 56 and name n)
       (resid 56 and name ca) (resid 56 and name c) 1.0 -60.0 30.0 2
!!    residue 57  V
  assign (resid 56 and name c) (resid 57 and name n)
       (resid 57 and name ca) (resid 57 and name c) 1.0 -60.0 30.0 2
!!    residue 58  R
  assign (resid 57 and name c) (resid 58 and name n)
       (resid 58 and name ca) (resid 58 and name c) 1.0 -60.0 30.0 2
!!    residue 59  Q
  assign (resid 58 and name c) (resid 59 and name n)
       (resid 59 and name ca) (resid 59 and name c) 1.0 -60.0 30.0 2
!!    residue 60  L
  assign (resid 59 and name c) (resid 60 and name n)
       (resid 60 and name ca) (resid 60 and name c) 1.0 -60.0 30.0 2
!!    residue 62  M
  assign (resid 61 and name c) (resid 62 and name n)
       (resid 62 and name ca) (resid 62 and name c) 1.0 -60.0 30.0 2
!!    residue 63  Y
  assign (resid 62 and name c) (resid 63 and name n)
       (resid 63 and name ca) (resid 63 and name c) 1.0 -120.0 40.0 2
!!    residue 66  H
  assign (resid 65 and name c) (resid 66 and name n)
```

```
        (resid 66 and name ca) (resid 66 and name c) 1.0 -120.0 40.0 2
!!      residue 73  S
  assign (resid 72 and name c) (resid 73 and name n)
        (resid 73 and name ca) (resid 73 and name c) 1.0 -60.0 30.0 2
!!      residue 77  L
  assign (resid 76 and name c) (resid 77 and name n)
        (resid 77 and name ca) (resid 77 and name c) 1.0 -60.0 30.0 2
!!      residue 84  W
  assign (resid 83 and name c) (resid 84 and name n)
        (resid 84 and name ca) (resid 84 and name c) 1.0 -120.0 40.0 2
!!      residue 91  H
  assign (resid 90 and name c) (resid 91 and name n)
        (resid 91 and name ca) (resid 91 and name c) 1.0 -120.0 40.0 2


!! Chi 1 restraints

!       Residue 1       Residue 2`      Residue 3       Residue 4       Constraint
!       Constraint = Force      Angle   (+/- Angle)     Exponent
!                       Constant        Value           Range                   Potential

!!      residue 6  V
    assign (resid 6 and name n)
        (resid 6 and name ca)
        (resid 6 and name cb)
        (resid 6 and name cg1) 1.0 180.0 20.0 2
!!      residue 32  V
    assign (resid 32 and name n)
        (resid 32 and name ca)
        (resid 32 and name cb)
        (resid 32 and name cg1) 1.0 180.0 20.0 2
!!      residue 33  V
    assign (resid 33 and name n)
        (resid 33 and name ca)
        (resid 33 and name cb)
        (resid 33 and name cg1) 1.0 180.0 20.0 2
!!      residue 40  V
    assign (resid 40 and name n)
        (resid 40 and name ca)
        (resid 40 and name cb)
        (resid 40 and name cg1) 1.0 180.0 20.0 2
!!      residue 43  V
    assign (resid 43 and name n)
        (resid 43 and name ca)
        (resid 43 and name cb)
        (resid 43 and name cg1) 1.0 180.0 20.0 2
!!      residue 57  V
    assign (resid 57 and name n)
        (resid 57 and name ca)
        (resid 57 and name cb)
        (resid 57 and name cg1) 1.0 180.0 20.0 2
!!      residue 22  I
```

```
    assign (resid 22 and name n)
        (resid 22 and name ca)
        (resid 22 and name cb)
        (resid 22 and name cg1) 1.0 180.0 20.0 2
!!      residue 31  I
    assign (resid 31 and name n)
        (resid 31 and name ca)
        (resid 31 and name cb)
        (resid 31 and name cg1) 1.0 -60.0 20.0 2
```

Table E.4 NOE and H-bond restraints for Chapter 3

restraints

```
! noe_tbl.1_031695
! constraints list for
! X-PLOR 3.1
! used to calculate
! structures described
! in Chapter 3

! proper X-PLOR format
! for these constraints
! is:
! assign (resid N and name
! hX) (resid M and name
hY)
!3.0 1.2 0.3
! these have been
! abbreviated so that
! only the residue numbers
! and atom names and
! constraint ranges are
! given
! namely format:
! N hX M hY 3.0 1.2 0.3
!

! residue 1A

1 hb* 2 hn 6.5 4.7 0.0

! residue 2R

2 ha 3 hd* 4.3 2.5 0.2
2 hn 3 hd* 4.3 2.5 0.2

! residue 3P

3 ha 3 hb1 2.4 0.6 0.3
3 ha 3 hb2 3.3 1.5 0.2
3 hg* 3 hd* 5.3 3.5 0.2
3 hb2 3 hg* 4.3 2.5 0.2

3 ha 4 hn 2.4 0.6 0.3

3 ha 5 hn 5.0 3.2 0.0
3 hb* 5 hn 6.0 4.2 0.0
3 hb2 6 hn 3.3 1.5 0.0
3 hd* 6 hb 4.3 2.5 0.2
3 hg* 6 hg2* 5.8 4.0 0.2
3 hd* 6 hg2* 5.8 4.0 0.2
3 hb1 63 he2 5.0 3.2 0.0
3 hg* 63 he2 4.3 2.5 0.2
3 hg* 63 hd2 6.0 4.2 0.0

! residue 4A

4 ha 4 hn 3.3 1.5 0.2
4 hb* 4 hn 3.9 2.1 0.3

4 ha 5 hn 3.3 1.5 0.2
4 hn 5 hn 3.3 1.5 0.2
4 hb* 5 hn 3.9 2.1 0.3

4 ha 7 hb* 4.3 2.5 0.2
4 ha 7 hn 5.0 3.2 0.0

! residue 5F

5 hn 5 ha 3.3 1.5 0.2
5 hn 5 hb1 3.3 1.5 0.2
5 hn 5 hb2 3.3 1.5 0.2
5 hn 5 hd* 5.3 3.5 0.2
5 ha 5 hb1 3.3 1.5 0.2
5 ha 5 hb2 5.0 3.2 0.0
5 ha 5 hd2 2.4 0.6 0.3
5 hb1 5 hd1 2.4 0.6 0.3
5 hb2 5 hd2 3.3 1.5 0.2

5 ha 6 hn 5.0 3.2 0.0
5 hn 6 hn 3.3 1.5 0.2
5 hb1 6 hn 3.3 1.5 0.2
5 hb2 6 hn 5.0 3.2 0.0
5 hd1 6 hn 5.0 3.2 0.0

5 ha 8 hn 5.0 3.2 0.0
5 ha 8 hb* 4.3 2.5 0.2

5 he1 9 hd2* 6.5 4.7 0.0
5 ha 48 he1 5.0 3.2 0.0
5 he1 60 ha 3.3 1.5 0.2
5 he1 60 hd2* 4.8 3.0 0.2
5 he1 60 hg 3.3 1.5 0.2
5 hd1 63 hd1 3.3 1.5 0.2

! residue 6V

6 hn 6 ha 5.0 3.2 0.0
6 hn 6 hb 2.4 0.6 0.3
6 hn 6 hg2* 3.9 2.1 0.3
6 ha 6 hb 5.0 3.2 0.0
6 ha 6 hg1* 4.8 3.0 0.2
6 ha 6 hg2* 4.8 3.0 0.2
6 hb 6 hg2* 3.9 2.1 0.3
6 hb 6 hg1* 4.8 3.0 0.2

6 ha 7 hn 5.0 3.2 0.0
6 hn 7 hn 3.3 1.5 0.2
6 hb 7 hn 3.3 1.5 0.2
6 hg* 7 hn 6.5 4.7 0.0

6 ha 9 hn 3.3 1.5 0.2
6 ha 9 hb1 3.3 1.5 0.2
6 ha 9 hb2 3.3 1.5 0.2
6 ha 9 hd2* 4.8 3.0 0.2
6 ha 10 hn 5.0 3.2 0.0
6 hg1* 10 he3 6.5 4.7 0.0
6 hg1* 10 he1 4.8 3.0 0.2

6 hg2* 63 hd1 4.8 3.0 0.2
6 hg1* 65 he1 4.8 3.0 0.2

! residue 7N

7 hn 7 ha 3.3 1.5 0.2
7 hn 7 hb* 3.9 2.1 0.2
7 ha 7 hb1 3.3 1.5 0.2
7 ha 7 hb2 3.3 1.5 0.2
7 ha 8 hn 5.0 3.2 0.0
7 hn 8 hn 3.3 1.5 0.2
7 hb* 8 hn 4.3 2.5 0.2
7 ha 10 hn 5.0 3.2 0.0
7 ha 10 hb* 4.3 2.5 0.2
7 ha 11 hn 5.0 3.2 0.0

! residue 8K

8 ha 8 hn 3.3 1.5 0.2
8 ha 8 hb* 4.3 2.5 0.2
8 ha 8 hg* 4.3 2.5 0.2
8 hb* 8 hg* 5.3 3.5 0.2
8 ha 9 hn 5.0 3.2 0.0
8 hn 9 hn 3.3 1.5 0.2
8 hb* 9 hn 6.0 4.2 0.0
8 hb* 48 he1 6.0 4.2 0.0
8 hd* 59 hg* 7.0 5.2 0.0

! residue 9L

9 hn 9 ha 3.3 1.5 0.2
9 hn 9 hb1 3.3 1.5 0.2
```

9 hn 9 hb2 3.3 1.5 0.2
9 hn 9 hg 5.0 3.2 0.0
9 hn 9 hd2* 6.5 4.7 0.0
9 ha 9 hb2 2.4 0.6 0.3
9 ha 9 hg 3.3 1.5 0.2
9 hb1 9 hg 3.3 1.5 0.2
9 ha 9 hd2* 4.8 3.0 0.2
9 ha 9 hd1* 3.9 2.1 0.3
9 hb1 9 hd1* 4.8 3.0 0.2
9 hb1 9 hd2* 3.9 2.1 0.3
9 hb2 9 hd1* 4.8 3.0 0.2
9 hb2 9 hd2* 4.8 3.0 0.2

9 ha 10 hn 5.0 3.2 0.0
9 hn 10 hn 3.3 1.5 0.2
9 hb1 10 hn 3.3 1.5 0.2
9 hb2 10 hn 5.0 3.2 0.0
9 hd2* 10 hn 6.5 4.7 0.0
9 hg 10 hn 5.0 3.2 0.0
9 hd2* 10 hz3 6.5 4.7 0.0

9 ha 12 hb2 5.0 3.2 0.0
9 ha 12 hn 5.0 3.2 0.0
9 ha 13 hn 5.0 3.2 0.0
9 hd1* 12 he* 4.9 3.1 0.3

9 hd1* 31 hd* 4.9 3.1 0.3
9 hd2* 31 hd* 5.8 4.0 0.2
9 hd1* 56 he2 3.0 1.5 0.2
9 hd1* 56 hz 4.8 3.0 0.2
9 hd1* 60 hd2* 5.8 4.0 0.2
9 hd2* 60 hd2* 4.9 3.1 0.3
9 hd2* 65 he1 4.8 3.0 0.2

! residue 10W

10 hn 10 ha 3.3 1.5 0.2
10 hn 10 hb* 3.9 2.1 0.2
10 hn 10 he3 5.0 3.2 0.0
10 ha 10 hb* 4.3 2.5 0.2
10 hn 10 hd1 5.0 3.2 0.0
10 hn 11 hn 3.3 1.5 0.2
10 hb* 11 hn 4.3 2.5 0.2
10 ha 13 hn 5.0 3.2 0.0
10 ha 13 hb 3.3 1.5 0.2
10 ha 13 hg1* 6.5 4.7 0.0
10 ha 13 hg2* 4.8 3.0 0.2
10 ha 14 hb* 4.3 2.5 0.2
10 hd1 10 hn 5.0 3.2 0.0
10 hd1 10 ha 5.0 3.2 0.0
10 hd1 10 hb* 4.3 2.5 0.2
10 ha 24 hz2 5.0 3.2 0.0

! residue 11S

11 ha 12 hn 3.3 1.5 0.2
11 hn 12 hg* 6.0 4.2 0.0
11 hb* 12 hn 4.3 2.5 0.2

11 ha 14 hn 5.0 3.2 0.0

! residue 12M

12 hn 12 ha 5.0 3.2 0.0
12 hn 12 hb* 4.3 2.5 0.2
12 hn 12 hg1 5.0 3.2 0.0
12 hn 12 hg2 5.0 3.2 0.0

12 ha 13 hn 5.0 3.2 0.0
12 ha 12 hg* 4.3 2.5 0.2
12 hn 13 hn 5.0 3.2 0.0
12 hb* 13 hn 3.4 1.6 0.3
12 hg* 13 hn 6.0 4.2 0.0
12 he* 13 hg1* 7.5 5.7 0.0
12 ha 15 hn 5.0 3.2 0.0
12 ha 18 hd22 3.3 1.5 0.2
12 ha 18 hd21 3.3 1.5 0.2
12 ha 47 he* 4.8 3.0 0.2
12 he* 43 hg1* 5.8 4.0 0.2
12 he* 44 ha 4.8 3.0 0.2
12 he* 44 hd1* 5.8 4.0 0.2
12 hb* 47 hd* 8.0 6.2 0.0
12 hb* 47 he1 6.0 4.2 0.0
12 hg* 47 he1 6.0 4.2 0.0

! residue 13V

13 hn 13 ha 3.3 1.5 0.2
13 hn 13 hb 3.3 1.5 0.2
13 hn 13 hg1* 6.5 4.7 0.0
13 hn 13 hg2* 4.8 3.0 0.2
13 ha 13 hg1* 3.5 1.7 0.3
13 ha 13 hg2* 3.5 1.7 0.3

13 ha 14 hn 5.0 3.2 0.0
13 hn 14 hn 3.3 1.5 0.2
13 hb 14 hn 3.3 1.5 0.2
13 hg1* 14 hn 6.5 4.7 0.0
13 hg2* 14 hn 6.5 4.7 0.0
13 ha 15 hn 5.0 3.2 0.0
13 ha 22 hg2* 4.8 3.0 0.2
13 hg1* 22 hg2* 5.8 4.0 0.2
13 hg1* 22 hd1* 7.5 5.7 0.0
13 hg1* 22 hg12 4.2 3.0 0.2
13 hg2* 22 hg2* 5.8 4.0 0.2
13 hg2* 22 hd* 5.8 4.0 0.2
13 hg1* 23 ha 4.8 3.0 0.2
13 hg1* 24 hn 4.8 3.0 0.2
13 hb24 he1 5.0 3.2 0.0
13 hg1* 24 he1 6.5 4.7 0.0
13 hg2* 24 he1 4.8 3.0 0.2
13 hg1* 31 hg2* 7.5 5.7 0.0
13 hg1* 31 hd* 5.8 4.0 0.2
13 hg1* 32 hn 6.5 4.7 0.0

! residue 14N

14 hn 14 ha 3.3 1.5 0.2
14 ha 14 hb1 3.3 1.5 0.2
14 ha 14 hb2 3.3 1.5 0.2
14 hn 15 hn 3.3 1.5 0.2
14 hb* 15 hn 4.3 2.5 0.2

14 ha 24 he1 3.3 1.5 0.2

! residue 15D

15 hn 15 ha 3.3 1.5 0.2
15 hn 15 hb* 4.3 2.5 0.2
15 ha 16 hn 3.3 1.5 0.2
15 hb* 22 hg2* 5.8 4.0 0.2

! residue 16K

16 hn 16 ha 3.3 1.5 0.2
16 hn 16 hb* 5.6 3.8 0.0
16 ha 16 hb* 3.4 1.6 0.3
16 ha 16 hg* 4.3 2.5 0.2
16 ha 16 hd* 6.0 4.2 0.0
16 hg* 16 hd* 4.4 2.6 0.3
16 hg* 16 he* 5.3 3.5 0.2
16 hb* 16 hg* 4.4 2.6 0.3
16 hd* 16 he* 4.4 2.6 0.3
16 ha 17 hn 3.3 1.5 0.2
16 hn 17 hn 3.3 1.5 0.2
16 hb* 17 hn 4.3 2.5 0.2
16 ha 19 hn 5.0 3.2 0.0
16 ha 19 hb1 5.0 3.2 0.0
16 ha 19 hb2 5.0 3.2 0.0

! residue 17S

17 hn 17 ha 3.3 1.5 0.2
17 hn 17 hb1 3.3 1.5 0.2
17 hn 17 hb2 3.3 1.5 0.2
17 ha 17 hb1 3.3 1.5 0.2
17 ha 17 hb2 3.3 1.5 0.2
17 ha 19 hn 5.0 3.2 0.0

! residue 18N

18 hn 18 hb* 3.9 2.1 0.2
18 ha 18 hb* 4.3 2.5 0.2
18 hd22 18 hb* 6.0 4.2 0.0
18 hd21 18 hb* 3.4 1.6 0.3
18 ha 19 hn 3.3 1.5 0.2
18 ha 20 hn 5.0 3.2 0.0
18 ha 21 hn 5.0 3.2 0.0
18 ha 21 hd2 3.3 1.5 0.2
18 hb* 22 hg2* 5.8 4.0 0.2
18 hb* 39 hd1 4.3 2.5 0.2
18 hd21 43 hg2* 6.5 4.7 0.0
18 hd22 43 hg2* 6.5 4.7 0.0
18 ha 43 hg2* 4.8 3.0 0.2
18 hb1 43 hg* 6.2 4.4 0.0
18 hb2 43 hg* 6.2 4.4 0.0
18 hd2* 47 he2 3.3 1.5 0.2

! residue 19E

19 hn 19 ha 3.3 1.5 0.2
19 hn 19 hb1 3.3 1.5 0.2
19 hn 19 hb2 3.3 1.5 0.2
19 hn 19 hg* 4.3 2.5 0.2

19 ha 20 hn 3.3 1.5 0.2
19 hn 20 hn 5.0 3.2 0.0
19 hb* 20 hn 4.3 2.5 0.2
19 hg* 20 hn 6.0 4.2 0.0

! residue 20K

20 hn 20 ha 3.3 1.5 0.2
20 hn 20 hb* 3.9 2.1 0.2
20 ha 20 hb* 3.4 1.6 0.3
20 ha 20 hd* 6.0 4.2 0.0
20 hd* 20 he* 4.4 2.6 0.3
20 ha 21 hn 5.0 3.2 0.0
20 hn 21 hn 3.3 1.5 0.2
20 hb* 21 hn 6.0 4.2 0.0

! residue 21F

21 hn 21 ha 5.0 3.2 0.0
21 hn 21 hb1 5.0 3.2 0.0
21 hn 21 hb2 3.3 1.5 0.2
21 hn 21 hd* 7.0 5.2 0.0
21 ha 21 hd1 2.4 0.6 0.3
21 ha 21 hb1 3.3 1.5 0.2
21 hb2 21 hd* 7.0 5.2 0.0
21 hb1 21 hd1 3.3 1.5 0.2
21 ha 22 hn 5.0 3.2 0.0
21 hb1 22 hn 5.0 3.2 0.0
21 hb2 22 hn 3.3 1.5 0.2
22 hn 21 hd* 7.0 5.2 0.0
21 ha 34 hd* 6.0 4.2 0.0
21 hd* 39 ha 5.3 3.5 0.2
21 hb2 43 hg* 4.8 3.0 0.2
21 hd2 43 hg2* 6.5 4.7 0.0

! residue 22I

22 hn 22 ha 3.3 1.5 0.2
22 hn 22 hg2* 4.8 3.0 0.2
22 ha 22 hb 3.3 1.5 0.2
22 ha 22 hg11 3.3 1.5 0.2
22 ha 22 hg12 5.0 3.2 0.0
22 ha 22 hg2* 4.4 2.6 0.2
22 ha 22 hd* 4.8 3.0 0.2
22 hb 22 hd* 4.8 3.0 0.2
22 hb 22 hg2* 3.5 1.7 0.3
22 hg11 22 hd* 4.8 3.0 0.2
22 hg12 22 hd* 4.8 3.0 0.2
22 hg2* 22 hd* 4.9 3.1 0.3
22 ha 23 hn 2.4 0.6 0.3
22 hb 23 hn 5.0 3.2 0.0
22 hd* 23 hn 6.5 4.7 0.0
22 ha 31 hg2* 4.8 3.0 0.2
22 hg11 31 hg2* 4.8 3.0 0.2
22 hd* 31 hd* 5.8 4.0 0.2
22 ha 33 ha 3.3 1.5 0.2
22 ha 33 hg2* 6.5 4.7 0.0
22 hb 33 hg2* 6.5 4.7 0.0
22 hg11 33 hg2* 4.8 3.0 0.2
22 hd* 33 hg2* 4.9 3.1 0.3
22 ha 34 hd* 4.3 2.5 0.2
22 hg2* 39 hd1 4.8 3.0 0.2
22 hd* 39 hd2 6.5 4.7 0.0

! residue 23H

23 hn 23 ha 5.0 3.2 0.0
23 ha 24 hn 2.4 0.6 0.3
23 hb* 24 hn 4.3 2.5 0.2
23 hn 32 hn 3.3 1.5 0.2

! residue 24W

24 hn 24 ha 3.3 1.5 0.2
24 hn 24 hb1 5.0 3.2 0.0
24 hn 24 hb2 5.0 3.2 0.0
24 hn 24 hd1 3.3 1.5 0.2
24 ha 24 hb1 2.4 0.6 0.3
24 ha 24 hb2 3.3 1.5 0.2
24 hb1 24 hd1 5.3 3.5 0.2
24 hb2 24 hd1 5.3 3.5 0.2
24 ha 25 hn 2.4 0.6 0.3
24 hb1 25 hn 3.3 1.5 0.2
24 hb2 25 hn 5.0 3.2 0.0
24 ha 32 hn 5.0 3.2 0.0
24 ha 31 ha 3.3 1.5 0.2

! residue 25S

25 hn 25 ha 5.0 3.2 0.0
25 hn 25 hb* 5.6 3.8 0.0
25 ha 25 hb* 4.3 2.5 0.2

25 ha 26 hn 3.3 1.5 0.2
25 hn 26 hn 5.0 3.2 0.0
25 hb* 26 hn 6.0 4.2 0.0
25 hb* 27 hn 3.4 1.6 0.3
25 hn 31 ha 5.0 3.2 0.0
25 hn 32 hg2* 6.5 4.7 0.0
25 ha 32 hg2* 4.8 3.0 0.2
25 hb1 32 hg2* 4.8 3.0 0.2
25 hb2 32 hg2* 4.8 3.0 0.2

! residue 26T

26 hn 26 ha 3.3 1.5 0.2
26 hn 26 hb 3.3 1.5 0.2
26 hn 26 hg* 6.5 4.7 0.0
26 ha 26 hg* 3.9 2.1 0.3
26 ha 27 hn 3.3 1.5 0.2
26 hg* 27 hn 6.5 4.7 0.0

! residue 27S

! residue 28G

28 ha* 29 hn 5.0 3.2 0.0

! residue 29E

29 hn 29 ha 3.3 1.5 0.2
29 hn 29 hb* 5.6 3.8 0.0
29 hn 29 hb* 5.6 3.8 0.0
29 hn 29 hg* 6.0 43.2 0.0
29 hb* 29 hg* 4.4 2.6 0.3
29hn 30 hn 3.3 1.5 0.2
29hb* 30 hn 6.0 4.2 0.0

! residue 30S

30 ha 30 hb* 3.4 1.6 0.3
30 ha 30 hn 5.0 3.2 0.0
30 ha 31 hn 2.4 0.6 0.3
30 hb1 31 hn 5.0 3.2 0.0
30 hb2 31 hn 5.0 3.2 0.0
30 ha 65 hz3 3.3 1.5 0.2
30 ha 87 ha 3.3 1.5 0.2
30 hb* 87 ha 4.3 2.5 0.2
30 hb* 87 hg* 7.0 5.2 0.0

! residue 31I

31 hn 31 ha 5.0 3.2 0.0
31 hn 31 hb 3.3 1.5 0.2
31 hn 31 hg1* 6.5 4.7 0.0
31 ha 31 hb 3.3 1.5 0.2
31 ha 31 hg1* 5.0 3.2 0.0
31 ha 31 hg2* 4.8 3.0 0.2
31 hb 31 hd* 4.8 3.0 0.2
31 hb 31 hg2* 4.8 3.0 0.2
31 hg1* 31 hd* 5.8 4.0 0.2
31 hg2* 31 hd* 4.9 3.1 0.3
31 ha 32 hn 2.4 0.6 0.3
31 hg2* 32 hn 4.8 3.0 0.2

31 hd* 60 hd1* 7.5 5.7 0.0
31 hd* 60 hd2* 5.8 4.0 0.2
31 hd* 65 he1 6.5 4.7 0.0
31 hn 86 hn 3.3 1.5 0.2
31 hb 86 hd2 5.0 3.2 0.0
31 hg2* 86 hd* 6.8 5.0 0.2
31 hd* 86 hd* 6.8 5.0 0.2
31 hn 87 ha 5.0 3.2 0.0

! residue 32V

32 hn 32 ha 5.6 3.2 0.0
32 hn 32 hb 3.3 1.5 0.2
32 hn 32 hg1* 6.5 4.7 0.0
32 hn 32 hg2* 4.8 3.0 0.2
32 ha 32 hb 3.3 1.5 0.2
32 ha 32 hg1* 3.5 1.7 0.3
32 ha 32 hg2* 3.5 1.7 0.3
32 ha 33 hn 2.4 0.6 0.3
32 hg1* 33 hn 6.5 4.7 0.0
32 hg1* 34 ha 4.8 3.0 0.2
32 hg1* 83 hd1* 5.8 4.0
0.2
32 ha 85 ha 3.3 1.5 0.2
32 hg1* 34 ha 6.5 4.7 0.0
32 hg1* 34 hd* 7.5 5.7 0.0
32 hg1* 34 hg* 5.8 4.0 0.2
32 hg1* 33 hn 6.5 4.7 0.0
32 hg1* 34 ha 4.8 3.0 0.2
32 hg2* 85 ha 4.8 3.0 0.2
32 hg2* 85 ha 6.5 4.7 0.0

! residue 33V

33 ha 33 hg1* 3.5 1.7 0.3
33 ha 33 hg2* 3.5 1.7 0.3
33 ha 34 hd1 3.3 1.5 0.2
33 ha 34 hd2 3.3 1.5 0.2
33 ha 34 hg* 4.3 2.5 0.2
33 hg1* 35 hn 4.8 3.0 0.2
33 hg1* 36 hn 6.5 4.7 0.0
33 hg1* 36 ha 4.8 3.0 0.2
33 hb 84 hn 3.3 1.5 0.2
33 hb 84 hb* 4.3 2.5 0.2
33 hn 85 ha 3.3 1.5 0.2
33 hg2* 86 hd* 8.5 6.7 0.0
33 hg2* 86 he2 4.8 3.0 0.2
33 hg1* 37 hn 4.8 3.0 0.2
33 hg1* 84 he1 6.5 4.7 0.0
33 hg2* 84 hd1 6.5 4.7 0.0
33 hg1* 39 hn 6.5 4.7 0.0
33 hg1* 39 hb* 5.8 4.0 0.2
33 hg1* 39 hd2 6.5 4.7 0.0

! residue 34P

34 ha 34 hb1 2.4 0.6 0.3
34 ha 34 hb2 3.3 1.5 0.2
34 hb* 34 hg* 5.3 3.5 0.2
34 hg* 34 ha 6.0 4.2 0.0
34 hg* 34 hd1 4.3 2.5 0.2
34 hg* 34 hd2 4.3 2.5 0.2
34 ha 35 hn 5.0 3.2 0.0
34 hg* 35 hn 6.0 4.2 0.0
34 hd* 35 hn 6.0 4.2 0.0
34 hb2 35 hn 3.3 1.5 0.2
34 ha 83 ha 5.0 3.2 0.0

! residue 35N

35 hn 35 ha 3.3 1.5 0.2
35 hn 35 hb* 3.9 2.1 0.2
35 ha 35 hb* 4.3 2.5 0.2

35 ha 36 hn 2.4 0.6 0.3
35 hb* 36 hn 6.0 4.2 0.0

! residue 36R

36 hn 36 ha 3.3 1.5 0.2
36 hn 36 hb1 3.3 1.5 0.2
36 hn 36 hb2 3.3 1.5 0.2
36 hg1 36 ha 5.0 3.2 0.0
36 hg2 36 ha 5.0 3.2 0.0
36 hn 37 hn 5.0 3.2 0.0
36 hb* 37 hn 3.4 1.6 0.3
36 ha 39 hn 5.0 3.2 0.0
36 hn 82 ha 5.0 3.2 0.0
36 hn 82 hb* 6.0 4.2 0.0
36 hb* 82 ha 6.0 4.2 0.0
36 hd* 84 he1 4.3 2.5 0.2
36 hb* 84 he1 3.4 1.6 0.3
36 hg* 40 hn 6.0 4.2 0.0

! residue 37E

37 hn 37 ha 3.3 1.5 0.2
37 hn 37 hb* 3.9 2.1 0.2
37 hn 37 hg* 6.0 4.2 0.0
37 ha 37 hb* 4.3 2.5 0.2
37 hb1 37 hg* 3.4 1.6 0.3
37 hb2 37 hg* 3.4 1.6 0.3
37 hn 38 hn 3.3 1.5 0.2
37 hb* 38 hn 4.3 2.5 0.2
37 hg* 38 hn 6.0 4.2 0.0
37 hb* 38 ha 6.0 4.2 0.0
37 hg* 38 ha 4.3 2.5 0.2
37 hn 39 hn 5.0 3.2 0.0
37 ha 40 hn 5.0 3.2 0.0
37 ha 40 hb 5.0 3.2 0.0
37 ha 40 hg2* 4.8 3.0 0.2
37 hn 82 hb* 6.0 4.2 0.0

! residue 38R

38 hn 38 ha 3.3 1.5 0.2
38 hn 38 hb* 3.9 2.1 0.2
38 ha 38 hb2 3.3 1.5 0.2
38 hn 38 hd* 6.0 4.2 0.0
38 ha 39 hn 5.0 3.2 0.0
38 hn 39 hn 3.3 1.5 0.2
38 hb1 39 hn 3.3 1.5 0.2
38 ha 41 hb* 4.3 2.5 0.2
38 ha 42 hn 5.0 3.2 0.0

! residue 39F

39 hn 39 ha 3.3 1.5 0.2
39 hn 39 hb* 3.9 2.1 0.2
39 hn 39 hd* 5.3 3.5 0.2
39 ha 39 hd1 3.3 1.5 0.2
39 hb1 39 hd2 3.3 1.5 0.2
39 hb2 39 hd1 3.3 1.5 0.2
39 ha 40 hn 5.0 3.2 0.0
39 hn 40 hn 3.3 1.5 0.2
39 hb* 40 hn 4.3 2.5 0.2
39 hb* 40 hg1* 7.5 5.7 0.0
39 hd2 40 hg1* 6.5 4.7 0.0
39 hd2 40 hg2* 4.8 3.0 0.2
39 hd2 40 hn 5.0 3.2 0.0
39 ha 43 hg2* 5.0 3.2 0.0
39 hd1 43 hb 3.3 1.5 0.2
39 hd2 44 hd1* 6.5 4.7 0.0
39 hd2 44 hd2* 6.5 4.7 0.0

! residue 40V

40 hn 40 ha 3.3 1.5 0.2
40 hn 40 hb 3.3 1.5 0.2
40 hn 40 hg1* 6.5 4.7 0.0
40 hn 40 hg2* 4.8 3.0 0.2
40 ha 40 hb 5.0 3.2 0.0
40 ha 40 hg1* 3.5 1.7 0.3
40 ha 40 hg2* 3.5 1.7 0.3

40 ha 41 hn 5.0 3.2 0.0
40 hn 41 hn 3.3 1.5 0.2
40 hb 41 hn 4.3 2.5 0.2
40 hg2* 41 hn 6.5 4.7 0.0
40 hg1* 41 hn 6.5 4.7 0.0
40 hg1* 41 ha 4.8 3.0 0.2

40 ha 43 hb 5.0 3.2 0.0
40 ha 44 hn 3.3 1.5 0.2
40 ha 44 hb1 3.3 1.5 0.2
40 ha 44 hb2 3.3 1.5 0.2
40 ha 44 hd2* 4.8 3.0 0.2
40 hg1* 44 hd1* 7.5 5.7
0.0
40 hg1* 44 hd2* 5.8 4.0
0.2

40 hg2* 53 ha 6.5 4.7 0.0
40 hg2* 56 hb* 7.5 5.7 0.0
40 hg2* 53 hn 4.8 3.0 0.2

! residue 41Q

41 hn 41 ha 3.3 1.5 0.2
41 hn 41 hb* 3.9 2.1 0.2
41 hn 41 hg* 4.3 2.5 0.2
41 ha 41 hg1 3.3 1.5 0.2
41 ha 41 hg2 3.3 1.5 0.2
41 hb* 41 hg* 4.4 2.6 0.3
41 he2* 41 hg* 7.0 5.2 0.0
41 hn 39 hd2 5.0 3.2 0.0
41 ha 42 hn 5.0 3.2 0.0
41 hb* 42 hn 4.3 2.5 0.2

! residue 42E

42 hn 42 ha 3.3 1.5 0.2
42 hn 42 hb2 3.3 1.5 0.2
42 hn 42 hb1 5.0 3.2 0.0
42 hn 42 hg* 4.3 2.5 0.2
42 ha 42 hb1 3.3 1.5 0.2
42 ha 42 hg* 3.4 1.6 0.3
42 hb2 42 hg* 4.3 2.5 0.2
42 ha 43 hn 5.0 3.2 0.0
42 hn 43 hn 2.4 0.6 0.3
42 hb2 43 hn 3.3 1.5 0.2
42 hb1 43 hn 3.3 1.5 0.2
42 hg* 43 hn 6.0 4.2 0.0
42 hb1 43 hg2* 4.8 3.0 0.2

! residue 43V

43 hn 43 ha 3.3 1.5 0.2
43 hn 43 ha 3.3 1.5 0.2
43 hn 43 hg2* 4.8 3.0 0.2
43 ha 43 hb 5.0 3.2 0.0
43 ha 43 hg* 6.2 4.4 0.2
43 hb 43 hg1* 3.9 2.1 0.3

43 ha 44 hn 5.0 3.2 0.0
43 hn 44 hn 2.4 0.6 0.3
43 hb 44 hn 3.3 1.5 0.2
43 hg* 44 hn 7.9 6.1 0.0

43 ha 45 hd* 6.0 4.2 0.0
43 ha 46 hn 5.0 3.2 0.0
43 ha 46 hb1 3.3 1.5 0.2
43 ha 46 hb2 3.3 1.5 0.2
43 hg1* 47 hd2 3.9 2.1 0.3
43 hn 45 hd* 6.0 4.2 0.0

! residue 44L

44 hn 44 ha 3.3 1.5 0.2
44 hn 44 hb1 3.9 2.1 0.2
44 hn 44 hb2 2.4 0.6 0.3
44 hn 44 hg 5.0 3.2 0.0
44 hn 44 hd1* 4.8 3.0 0.2
44 ha 44 hb1 3.9 2.1 0.2
44 ha 44 hb2 3.9 2.1 0.2
44 ha 44 hg 5.0 3.2 0.0
44 ha 44 hd1* 4.8 3.0 0.2
44 ha 44 hd2* 6.5 4.7 0.0
44 hb1 44 hd1* 6.5 4.7 0.0
44 hb1 44 hd2* 4.8 3.0 0.2
44 hn 45 hd* 3.4 1.6 0.3
44 hb1 45 hd1 3.3 1.5 0.2
44 hb1 45 hd2 3.3 1.5 0.2
44 hb2 45 hd* 6.0 4.2 0.0
44 ha 47 hn 5.0 3.2 0.0
44 ha 48 hn 5.0 3.2 0.0
44 ha 47 hd2 4.4 2.6 0.3
44 hd1* 51 hb* 7.5 5.7 0.0
44 hd2* 51 hb* 5.8 4.0 0.2
44 hd1* 52 ha 6.5 4.7 0.0
44 hd2* 52 ha 4.8 3.0 0.2
44 hd2* 52 hb* 7.5 5.7 0.0
44 hd2* 53 ha 6.8 5.0 0.2
44 hd2* 53 hb1 6.5 4.7 0.0
44 hd2* 53 hn 6.5 4.7 0.0
44 hd* 55 hb* 8.9 7.1 0.0
44 hd1* 56 hb* 7.5 5.7 0.0
44 hd1* 56 hd1 6.5 4.7 0.0
44 hd1* 48 hb1 6.5 4.7 0.0
44 hd1* 48 hb2 6.5 4.7 0.0
44 ha 48 hd2 5.0 3.2 0.0
44 hd1* 48 hd2 3.9 2.1 0.3
44 hd2* 48 hd2 6.5 4.7 0.0

! residue 45P

45 ha 45 hb1 3.3 1.5 0.2
45 ha 45 hb2 3.3 1.5 0.2

45 hb* 45 hd* 7.0 5.2 0.0
45 ha46 hn 5.0 3.2 0.0
45 hd* 46 hn 3.4 1.6 0.3
45 ha 48 hn 3.3 1.5 0.2
45 ha 49 ha 5.0 3.2 0.0

! residue 46K

46 hn 46 ha 3.3 1.5 0.2
46 hn 46 hb* 3.9 2.1 0.2
46 hn 46 hg* 4.3 2.5 0.2
46 hn 46 hd* 4.3 2.5 0.2
46 ha 46 hb2 3.3 1.5 0.2
46 ha 46 hg1 3.3 1.5 0.2
46 ha 46 hg2 3.3 1.5 0.2
46 hb* 46 hg* 7.0 5.2 0.0
46 hg* 46 he* 7.0 5.2 0.0

46 ha 47 hn 5.0 3.2 0.0
46 hn 47 hn 2.4 0.6 0.3
46 hn 47 hd2 3.3 1.5 0.2
46 hb2 47 hn 5.0 3.2 0.0
46 hb1 47 hn 5.0 3.2 0.0
46 hb2 47 hd2 3.3 1.5 0.2
46 hb1 47 hd2 3.3 1.5 0.2
46 hb* 47 he2 6.0 4.2 0.0

! residue 47Y

47 hn 47 ha 5.0 3.2 0.0
47 hn 47 hb2 3.3 1.5 0.2
47 hn 47 hb1 5.0 3.2 0.0
47 hn 47 hd2 3.3 1.5 0.2
47 ha 47 hb2 5.0 3.2 0.0
47 ha 47 hb1 3.3 1.5 0.2
47 hb2 47 hd* 5.3 3.5 0.2
47 hb1 47 hd1 3.3 1.5 0.2
47 ha 48 hn 5.0 3.2 0.0
47 hn 48 hn 2.4 0.6 0.3
47 hb2 48 hn 3.3 1.5 0.2
47 hb1 48 hn 5.0 3.2 0.0
47 hb2 48 hd2 5.0 3.2 0.0
47 hb1 48 hd2 5.0 3.2 0.0
47 hd* 47 ha 4.3 2.5 0.2
47 hd2 47 hn 3.3 1.5 0.2

! residue 48F

48 hn 48 ha 5.0 3.2 0.0
48 hn 48 hb2 3.3 1.5 0.2
48 hn 48 hb1 5.0 3.2 0.0
48 hn 47 hd2 5.0 3.2 0.0
48 hn 48 hd2 5.0 3.2 0.0
48 ha 48 hb2 5.0 3.2 0.0
48 ha 48 hb1 3.3 1.5 0.2
48 ha 48 hd1 3.3 1.5 0.2
48 hb1 48 hd1 3.3 1.5 0.2
48 hb2 48 hd2 3.3 1.5 0.2
48 ha49hn 3.3 1.5 0.2
48 hb1 49 hn 5.0 3.2 0.0
48 ha 51 hb* 6.0 4.2 0.0
48 hb1 51 hb* 4.3 2.5 0.2
48 hb2 51 hb* 4.3 2.5 0.2

! residue 49K

49 hn 49 ha 3.3 1.5 0.2
49 hn 49 hb* 3.9 2.1 0.2
49 hn 49 hg* 4.3 2.5 0.2
49 ha 49 hb1 2.4 0.6 0.3
49 ha 49 hb2 5.0 3.2 0.0
49 ha 49 hd* 6.0 4.2 0.0
49 hg* 49 ha 4.3 2.5 0.2
49 hg* 49 hb1 3.4 1.6 0.3
49 hg* 49 hb2 4.3 2.5 0.2
49 hg* 49 hd* 4.4 2.6 0.3

49 ha 50 hn 3.3 1.5 0.2
49 hn 50 hn 5.0 3.2 0.0
49 hb1 50 hn 3.3 1.5 0.2
49 hb2 50 hn 5.0 3.2 0.0
49 hg* 50 hn 6.0 4.2 0.0

49 ha 51 hn 5.0 3.2 0.0

! residue 50H

50 hn 50 ha 3.3 1.5 0.0
50 hn 50 hb2 3.3 1.5 0.2
50 hn 50 hb1 5.0 3.2 0.0
50 ha 50 hb1 2.4 0.6 0.3
50 hn 51 hn 3.3 1.5 0.2

! residue 51S

51 ha 51 hb* 3.4 1.6 0.3
51 hn 51 hb* 3.0 1.2 0.3

51 ha 52 hn 3.3 1.5 0.2
51 hb* 52 hn 4.3 2.5 0.2

! residue 52N

52 ha 52 hb* 4.3 2.5 0.2
52 hd21 52 hb* 4.3 1.6 0.3
52 hd22 52 hb* 4.3 2.5 0.2

52 ha 53 hn 2.4 0.6 0.3
52 hb* 53 hn 4.3 2.5 0.2
52 ha 55 hn 5.0 3.2 0.0

! residue 53F

53 ha 53 hn 3.3 1.5 0.2
53 hb* 53 hn 3.9 2.1 0.2
53 hd* 53 hn 5.3 3.5 0.2
53 ha 53 hb1 5.0 3.2 0.2
53 ha 53 hd* 5.3 3.5 0.2
53 hb1 53 hd* 5.3 3.5 0.2
53 hb2 53 hd* 5.3 3.5 0.2

53 ha 54 hn 5.0 3.2 0.2
53 hn 54 hn 3.3 1.5 0.2
53 hb2 54 hn 3.3 1.5 0.2
53 hb1 54 hn 5.0 3.2 0.0
53 hd2 54 hn 5.0 3.2 0.0

53 ha 57 hn 5.0 3.2 0.0

! residue 54A

54 ha 54 hn 3.3 1.5 0.2
54 hb* 54 hn 4.8 3.0 0.2

54 ha 55 hn 5.0 3.2 0.0
54 hn 55 hn 3.3 1.5 0.2
54 hb* 55 hn 4.8 3.0 0.2

54 ha 57 hn5.0 3.2 0.0
54 ha 57 hb 3.3 1.5 0.2
54 ha 57 hg1* 6.5 4.7 0.0
54 ha 57 hg2* 4.8 3.0 0.2
54 ha 58 hn 5.0 3.2 0.0

! residue 55S

55 hn 56 hn 3.3 1.5 0.2
55 hb* 56 hn 4.3 2.5 0.2

55 ha 58 hn 3.3 1.5 0.2
55 ha 58 hb* 4.3 2.5 0.2
55 ha 55 hb* 6.0 4.2 0.0

! residue 56F

56 ha 56 hn 3.3 1.5 0.2
56 hb* 56 hn 3.9 2.1 0.2
56 hd* 56 hn 5.3 3.5 0.2
56 ha 56 hd2 3.3 1.5 0.2
56 hb1 56 hd* 5.3 3.5 0.2
56 hb2 56 hd* 5.3 3.5 0.2

56 ha 57 hn 5.0 3.2 0.0
56 hn 57 hn 3.3 1.5 0.2
56 hb* 57 hn 4.3 2.5 0.2
56 hd2 57 hn 5.0 3.2 0.0

56 ha 59 hn 5.0 3.2 0.0
56 ha 59 hb1 5.0 3.2 0.0
56 ha 60 hn 5.0 3.2 0.0
56 he2 60 hd1* 4.8 3.0 0.2
56 he2 60 hd2* 6.5 4.7 0.0
56 hz 60 hd1* 6.5 4.7 0.0
56 hz 60 hd2* 4.8 3.0 0.2

! residue 57V

57 ha 57 hn 3.3 1.5 0.2
57 hb 57 hn 2.4 0.6 0.3
57 hg1* 57 hn 4.8 3.0 0.2
57 hg2* 57 hn 3.9 2.1 0.3
57 ha 57 hg1* 3.5 1.7 0.3
57 ha 57 hg2* 3.5 1.7 0.3

57 ha 58 hn 5.0 3.2 0.0
57 hn 58 hn 3.3 1.5 0.2
57 hg1* 58 hn 4.8 3.0 0.2
57 hg2* 58 hn 6.5 4.7 0.0

57 ha 60 hn 3.3 1.5 0.2
57 ha 60 hd1* 3.9 2.1 0.3
57 hg1* 60 hd1* 7.5 5.7
0.0
57 hg2* 60 hd1* 5.8 4.0
0.2
57 hg1* 61 hd22 6.5 4.7
0.0
57 hg1* 61 hd21 6.5 4.7
0.0

57 hg1* 84 hz3 4.8 3.0 0.2
57 hg2* 84 hz3 6.5 4.7 0.0
57 hg1* 86 he* 6.8 5.0 0.2
57 hg2* 86 he* 6.8 5.0 0.2
57 hg2* 86 hd* 8.5 6.7 0.0

! residue 58R

58 hn 58 hb* 3.9 2.1 0.2
58 ha 58 hb* 3.4 1.6 0.3

58 ha 59 hn 5.0 3.2 0.0
58 hn 59 hn 3.3 1.5 0.2
58 hb* 59 hn 4.3 2.5 0.2

58 ha 61 hn 3.3 1.5 0.2
58 ha 61 hb* 3.4 1.6 0.3
58 ha 62 hn 5.0 3.2 0.0

! residue 59Q

59 ha 59 hn 3.3 1.5 0.2
59 hb* 59 hn 3.9 2.1 0.2
59 hg* 59 hn 4.3 2.5 0.2
59 ha 59 hb1 3.3 1.5 0.2
59 ha 59 hg* 3.9 2.1 0.2
59 he2* 59 hg* 6.0 4.2 0.0
59 hn 60 hn 3.3 1.5 0.2
59 hb2 60 hn 3.3 1.5 0.2
59 hb1 60 hn 3.3 1.5 0.2
59 hg* 60 hn 6.0 4.2 0.0
59 ha 62 hn 5.0 3.2 0.0
59 ha 62 hb1 5.0 3.2 0.0
59 ha 62 hb2 5.0 3.2 0.0
59 hb2 63 he1 3.3 1.5 0.2
59 hb1 63 he1 3.3 1.5 0.2

! residue 60L

60 ha 60 hn 3.3 1.5 0.2
60 hb2 60 hn 2.4 0.6 0.3
60 hb1 60 hn 3.3 1.5 0.2

60 hd1* 60 hn 6.5 4.7 0.0
60 hd2* 60 hn 6.5 4.7 0.0
60 ha 60 hd1* 6.5 4.7 0.0
60 ha 60 hd2* 3.9 2.1 0.3
60 ha 60 hb1 3.3 1.5 0.2
60 ha 60 hb2 3.3 1.5 0.2
60 hb1 60 hg 5.0 3.2 0.0
60 hb2 60 hg 5.0 3.2 0.0
60 hd1* 60 hb1 4.8 3.0 0.2
60 hd2* 60 hb1 4.8 3.0 0.2
60 hd1* 60 hb2 4.8 3.0 0.2
60 hd2* 60 hb2 6.5 4.7 0.0
60 ha 61 hn 5.0 3.2 0.0
60 hn 61 hn 3.3 1.5 0.2
60 hb1 61 hn 3.3 1.5 0.2
60 hb2 61 hn 3.3 1.5 0.2

60 ha 63 hn 5.0 3.2 0.0
60 ha 63 hd1 3.3 1.5 0.2
60 hb1 65 hb2 5.0 3.2 0.0
60 hb2 65 hb2 3.3 1.5 0.2
60 hb2 65 hb1 5.0 3.2 0.0
60 hd2* 65 hb2 6.5 4.7 0.0
60 hd2* 65 hb1 6.5 4.7 0.0
60 hd1* 86 hd* 6.8 5.0 0.2
60 hd2* 86 hd* 8.5 6.7 0.0

! residue 61N

61 ha 61 hn 3.3 1.5 0.2
61 hb* 61 hn 3.9 2.1 0.2
61 ha 61 hb* 4.3 2.5 0.2
61 hd22 61 ha 5.0 3.2 0.0
61 hd21 61 ha 3.3 1.5 0.2
61 hd21 61 hb* 5.3 3.5 0.2
61 ha 62 hn 5.0 3.2 0.0
61 hn 62 hn 3.3 1.5 0.2
61 hb* 62 hn 4.3 2.5 0.2
61 ha 64 hn 5.0 3.2 0.0

! residue 62M

62 ha 62 hn 3.3 1.5 0.2
62 hb* 62 hn 3.9 2.1 0.2
62 hg* 62 hn 4.3 2.5 0.2
62 ha 62 hb* 4.3 2.5 0.2
62 ha 62 hg* 4.3 2.5 0.2
62 ha 63 hn 5.0 3.2 0.0
62 hn 63 hn 2.4 0.6 0.3
62 hb* 63 hn 4.3 2.5 0.2
62ha 64 hn 5.0 3.2 0.0

! residue 63Y

63 ha 63 hn 5.0 3.2 0.0
63 hb2 63 hn 3.3 1.5 0.2
63 ha 63 hn 5.0 3.2 0.0
63 hd* 63 hb2 5.3 3.5 0.2
63 ha 63 hb1 3.3 1.5 0.2
63 hb2 63 hd1 3.3 1.5 0.2
63 hn 63 hd1 3.3 1.5 0.2
63 ha 64 hn 5.0 3.2 0.0
63 hn 64 hn 3.3 1.5 0.2
63 hd2 63 ha 3.3 1.5 0.2

! residue 64G

64 ha* 65 hn 5.0 3.2 0.0
64 hn 65 hn 3.3 1.5 0.2

! residue 65W

65 ha 65 hn 5.0 3.2 0.0
65 hb1 65 hn 5.0 3.2 0.0
65 hb2 65 hn 3.3 1.5 0.2
65 hn 65 hd1 5.0 3.2 0.0

65 ha66 hn 2.4 0.6 0.3
65 hb1 66 hn 5.0 3.2 0.0
65 hd1 65 hn 5.0 3.2 0.0
65 hb1 65 hd1 5.0 3.2 0.0
65 hb2 65 hd1 5.0 3.2 0.0
65 ha 88 ha 3.3 1.5 0.2
65 ha 89 hn 5.0 3.2 0.0

! residue 66H

66 ha 66 hn 3.3 1.5 0.2
66 hb1 66 hn 5.0 3.2 0.0
66 hb2 66 hn 3.3 1.5 0.2
66 ha 66 hb2 5.0 3.2 0.0
66 ha 66 hb1 2.4 0.6 0.3
66 ha 67 hn 2.4 0.6 0.3
66 hb1 67 hn 3.3 1.5 0.2
66 hb2 67 hn 3.3 1.5 0.2

66 hn 87 hn 5.0 3.2 0.0
66 hn 88 ha 5.0 3.2 0.0

! residue 67K

67 ha 67 hn 3.3 1.5 0.2
67 hb* 67 hn 3.9 2.1 0.2
67 ha 67 hb* 4.3 2.5 0.2

67 ha 68 hn 3.3 1.5 0.2

67 ha 87 hn 5.0 3.2 0.0
67 ha 86 ha 3.3 1.5 0.2
67 ha 86 hd1 5.3 3.5 0.2
67 hn 86 hd* 7.0 5.2 0.0

! residue 68V

68 ha 68 hn 3.3 1.5 0.2
68 hb 68 hn 5.0 3.2 0.0
68 hg* 68 hn 5.3 3.5 0.3

68 ha 69 hn 2.4 0.6 0.3
68 hn 69 hn 5.0 3.2 0.0
68 hg* 69 ha 6.2 4.4 0.2
68 hg* 69 hn 5.3 3.5 0.3

68 hn 86 ha 5.0 3.2 0.0
68 hg* 87 hn 6.2 4.4 0.2

! residue 69Q

69 ha 69 hn 3.3 1.5 0.2
69 hn 69 hb* 3.0 1.2 0.3
69 hn 69 hg* 4.3 2.5 0.2
69 ha 69 hb1 3.3 1.5 0.2
69 ha 69 hb2 3.3 1.5 0.2
69 hb* 69 hg* 4.4 2.6 0.3
69 he2* 69 hg1* 5.3 3.5
0.2
69 ha 70 hn 3.3 1.5 0.2
69 hn 70 hn 3.3 1.5 0.2

! residue 70D

70 hn 70 hb* 3.9 2.1 0.2
70 ha 71 hn 3.3 1.5 0.2
70 hn 71 hn 5.0 3.2 0.0

! residue 71V

71 ha 71 hb 3.3 1.5 0.2
71 ha 71 hn 3.3 1.5 0.2
71 hb 71 hn 5.0 3.2 0.0
71 hn71 hg1* 4.8 3.0 0.2
71 ha 72 hn 3.3 1.5 0.2
71 hb 72 hn 5.0 3.2 0.0

! residue 72K

72 hn 72 ha 3.3 1.5 0.2
72 hn 72 hb* 3.9 2.1 0.2
72 ha 72 hb* 3.4 1.6 0.3
72 hg* 72 hd* 4.4 2.6 0.3
72 ha 73 hn 3.3 1.5 0.2
72 hb* 73 hn 6.0 4.2 0.0

! residue 73S

73 ha 73 hb* 3.4 1.6 0.3
73 ha 74 hn 3.3 1.5 0.2
73 hn 74 hn 3.3 1.5 0.2

! residue 74G

74 hn 75 hn 3.3 1.5 0.2

! residue 75S

75 ha 76 hn 3.3 1.5 0.2
75 hn 76 hn 3.3 1.5 0.2
75 hb* 76 hn 6.0 4.2 0.0

! residue 76M

76 ha 76 hb1 3.3 1.6 0.2
76 ha 76 hb2 3.3 1.6 0.2
76 ha 76 hg* 4.3 2.5 0.2
76 hn 76 hb* 3.0 1.2 0.3
76 ha 77 hn 3.3 1.5 0.2
76 hn 77 hn 3.3 1.5 0.2
76 ha 78 hn 3.3 1.5 0.2

! residue 77L

77 ha 77 hb* 4.3 2.5 0.2
77 ha 77 hg 3.3 1.5 0.2
77 ha 77 hd* 5.8 4.0 0.2
77 hb* 77 hd* 7.2 5.4 0.2
77 hb* 77 hd* 6.3 4.5 0.3
77 hg* 77 hd* 5.3 3.5 0.3
77 ha78 hn 3.3 1.5 0.2
77 hb* 78 hn 4.3 2.5 0.2
77 hd* 78 hn 6.2 4.4 0.2

! residue 78S

78 ha 79 hn 3.3 1.5 0.2
78 hn 79 hn 3.3 1.5 0.2
78 hb* 79 hn 6.0 4.2 0.0

! residue 79N

! residue 80N

80 ha 81 hn 3.3 1.5 0.2

! residue 81D

! residue 82S

82 ha 82 hb* 3.4 1.6 0.3
82 ha 83 hn 3.3 1.5 0.2
82 hn83 hn 2.4 0.6 0.3
82 hb* 83 hn 4.3 2.5 0.2

82 hb* 84 he1 6.0 4.2 0.0

! residue 83R

83 ha 83 hn 3.3 1.5 0.2
83 hb* 83 hn 3.9 2.1 0.2
83 hg* 83 hn 4.3 2.5 0.2
83 ha 83 hb* 4.3 2.5 0.2
83 ha 84 hn 3.3 1.5 0.2
83 hb* 83 hd* 5.3 3.5 0.2

83 hb* 84 hn 4.3 2.5 0.2

! residue 84W

84 ha 84 hn 3.3 1.5 0.2
84 hb* 84 hn 3.9 2.1 0.2
84 hd1 84 hn 5.0 3.2 0.0
84 hb* 84 he3 4.3 2.5 0.2
84 ha 85 hn 2.4 0.6 0.3
84 hb* 85 hn 4.3 2.5 0.2
84 he3 86 he* 7.0 5.2 0.0
84 hd1 84 hn 3.3 1.5 0.2
84 hd1 84 ha 3.3 1.5 0.2
84 hd1 84 hb* 4.3 2.5 0.2

! residue 85E

85 ha 85 hn 3.3 1.5 0.2
85 hb* 85 hn 3.9 2.1 0.2
85 ha 85 hb* 4.3 2.5 0.2
85 ha 85 hg* 4.3 2.5 0.2
85 hb1 85 hg* 4.3 2.5 0.2
85 hb2 85 hg* 4.3 2.5 0.2
85 ha 86 hn 2.4 0.6 0.3
85 hb* 86 hn 6.0 4.2 0.0

! residue 86F

86 ha 86 hn 3.3 1.5 0.2
86 hb2 86 hn 3.3 1.5 0.2
86 hb1 86 hn 5.0 3.2 0.0
86 hn 86 hd* 5.3 3.5 0.2
86 ha 86 hb1 3.3 1.5 0.2
86 ha 86 hd* 4.3 2.5 0.2
86 hb2 86 hd* 5.3 3.5 0.2
86 hb1 86 hd* 5.3 3.5 0.2

86 ha 87 hn 2.4 0.6 0.3
86 hb1 87 hn 5.0 3.2 0.0
86 hd* 87 hn 7.0 5.2 0.0

! residue 87E

87 ha 87 hb* 4.3 2.5 0.2
87 hn 87 hb* 3.9 2.1 0.2
87 hn 87 hg* 6.0 4.2 0.0

87 ha 88 hn 2.4 0.6 0.3
87 hb* 88 hn 6.0 4.2 0.0
87 hg* 88 hn 6.0 4.2 0.0

! residue 88N

88 ha 88 hn 5.0 3.2 0.0
88 hb* 88 hn 3.9 2.1 0.2
88 ha 89 hb* 6.0 4.2 0.0

! residue 89E

89 ha 89 hn 3.3 1.5 0.2
89 hb* 89 hn 5.6 3.8 0.0
89 hg* 89 hn 4.3 2.5 0.2
89 hb* 89 hg* 4.4 2.6 0.3

89 ha 90 hn 3.3 1.5 0.2
89 hn 90 hn 3.3 1.5 0.2
89 hb* 90 hn 6.0 4.2 0.0

! residue 90R

90 ha 91 hn 3.3 1.5 0.2
90 hb* 91 hn 3.9 2.1 0.2
90 hb* 91 hn 4.3 2.5 0.2

! residue 91H

91 hn 91 hb* 3.9 2.1 0.2
91 ha 91 hb* 4.3 2.5 0.2
91 ha 92 hn 3.3 1.5 0.2

! residue 92A

92 ha 92 hb* 4.8 3.0 0.2

! Hydrogen bond constraints

7 hn 3 o 2.1 0.3 0.2
7 n 3 o 3.1 0.3 0.2
8 hn 4 o 2.1 0.3 0.2
8 n 4 o 3.1 0.3 0.2
9 hn 5 o 2.1 0.3 0.2
9 n 5 o 3.1 0.3 0.2
13 hn 9 o 2.1 0.3 0.2
13 n 9 o 3.1 0.3 0.2
21 hn 18 o 2.1 0.3 0.2
21 n 18 o 3.1 0.3 0.2
25 hn 30 o 2.1 0.3 0.2
25 n 30 o 3.1 0.3 0.2
31 hn 86 o 2.1 0.3 0.2
31 n 86 o 3.1 0.3 0.2
32 hn 23 o 2.1 0.3 0.2
32 n 23 o 3.1 0.3 0.2
33 hn 84 o 2.1 0.3 0.2
33 n 84 o 3.1 0.3 0.2
42 hn 38 o 2.1 0.3 0.2
42 n 38 o 3.1 0.3 0.2
44 hn 40 o 2.1 0.3 0.2
44 n 40 o 3.1 0.3 0.2
46 hn 43 o 2.1 0.3 0.2
46 n 43 o 3.1 0.3 0.2
48 hn 44 o 2.1 0.3 0.2
48 n 44 o 3.1 0.3 0.2
57 hn 53 o 2.1 0.3 0.2
57 n 53 o 3.1 0.3 0.2
58 hn 54 o 2.1 0.3 0.2
58 n 54 o 3.1 0.3 0.2
59 hn 55 o 2.1 0.3 0.2
59 n 55 o 3.1 0.3 0.2
60 hn 56 o 2.1 0.3 0.2
60 n 56 o 3.1 0.3 0.2
61 hn 57 o 2.1 0.3 0.2
61 n 57 o 3.1 0.3 0.2
62 hn 58 o 2.1 0.3 0.2
62 n 58 o 3.1 0.3 0.2
66 hn 87 o 2.1 0.3 0.2
66 n 87 o 3.1 0.3 0.2
86 hn 31 o 2.1 0.3 0.2
86 n 31 o 3.1 0.3 0.2

Table E.5 Dihedral restraints for Chapter 3

```
restraints
! Dihedral restraints for structure
! calculations described in Chapter 3

restraints dihedral reset
! Torsion psi constraints
!     residue 4  A
  assign (resid 3 and name c) (resid 4 and name n)
       (resid 4 and name ca) (resid 4 and name c) 1.0 -60.0 30.0 2

!     residue 12  M
  assign (resid 11 and name c) (resid 12 and name n)
       (resid 12 and name ca) (resid 12 and name c) 1.0 -60.0 30.0 2

!     residue 13  V
  assign (resid 12 and name c) (resid 13 and name n)
       (resid 13 and name ca) (resid 13 and name c) 1.0 -60.0 30.0 2

!     residue 14  N
  assign (resid 13 and name c) (resid 14 and name n)
       (resid 14 and name ca) (resid 14 and name c) 1.0 -60.0 30.0 2

!     residue 16  K
  assign (resid 15 and name c) (resid 16 and name n)
       (resid 16 and name ca) (resid 16 and name c) 1.0 -60.0 30.0 2

!     residue 19  E
  assign (resid 18 and name c) (resid 19 and name n)
       (resid 19 and name ca) (resid 19 and name c) 1.0 -60.0 30.0 2

!     residue 21  F
  assign (resid 20 and name c) (resid 21 and name n)
       (resid 21 and name ca) (resid 21 and name c) 1.0 -120 40.0 2

!     residue 24  W
  assign (resid 23 and name c) (resid 24 and name n)
       (resid 24 and name ca) (resid 24 and name c) 1.0 -60.0 30.0 2

!     residue 25  A
  assign (resid 24 and name c) (resid 25 and name n)
       (resid 25 and name ca) (resid 25 and name c) 1.0 -60.0 30.0 2

!     residue 31  I
  assign (resid 30 and name c) (resid 31 and name n)
       (resid 31 and name ca) (resid 31 and name c) 1.0 -120.0 40.0 2

!     residue 32  V
  assign (resid 31 and name c) (resid 32 and name n)
       (resid 32 and name ca) (resid 32 and name c) 1.0 -120.0 40.0 2
```

```
!     residue 33  V
  assign (resid 32 and name c) (resid 33 and name n)
       (resid 33 and name ca) (resid 33 and name c) 1.0 -120.0 40.0 2

!     residue 35  N
  assign (resid 34 and name c) (resid 35 and name n)
       (resid 35 and name ca) (resid 35 and name c) 1.0 -120.0 40.0 2

!     residue 36  R
  assign (resid 35 and name c) (resid 36 and name n)
       (resid 36 and name ca) (resid 36 and name c) 1.0 -60.0 30.0 2

!     residue 37  E
  assign (resid 36 and name c) (resid 37 and name n)
       (resid 37 and name ca) (resid 37 and name c) 1.0 -60.0 30.0 2

!     residue 39  F
  assign (resid 38 and name c) (resid 39 and name n)
       (resid 39 and name ca) (resid 39 and name c) 1.0 -60.0 30.0 2

!     residue 41  Q
  assign (resid 40 and name c) (resid 41 and name n)
       (resid 41 and name ca) (resid 41 and name c) 1.0 -60.0 30.0 2

!     residue 43  V
  assign (resid 42 and name c) (resid 43 and name n)
       (resid 43 and name ca) (resid 43 and name c) 1.0 -120.0 40.0 2

!     residue 47  Y
  assign (resid 46 and name c) (resid 47 and name n)
       (resid 47 and name ca) (resid 47 and name c) 1.0 -120.0 40.0 2

!     residue 48  F
  assign (resid 47 and name c) (resid 48 and name n)
       (resid 48 and name ca) (resid 48 and name c) 1.0 -120.0 40.0 2

!     residue 51  S
  assign (resid 50 and name c) (resid 51 and name n)
       (resid 51 and name ca) (resid 51 and name c) 1.0 -120.0 40.0 2

!     residue 53  F
  assign (resid 52 and name c) (resid 53 and name n)
       (resid 53 and name ca) (resid 53 and name c) 1.0 -60.0 30.0 2

!     residue A  54
  assign (resid 53 and name c) (resid 54 and name n)
       (resid 54 and name ca) (resid 54 and name c) 1.0 -60.0 30.0 2

!     residue F  56
  assign (resid 55 and name c) (resid 56 and name n)
       (resid 56 and name ca) (resid 56 and name c) 1.0 -60.0 30.0 2
```

```
!     residue 57  V
  assign (resid 56 and name c) (resid 57 and name n)
       (resid 57 and name ca) (resid 57 and name c) 1.0 -60.0 30.0 2

!     residue 58  R
  assign (resid 57 and name c) (resid 58 and name n)
       (resid 58 and name ca) (resid 58 and name c) 1.0 -60.0 30.0 2

!     residue 59  Q
  assign (resid 58 and name c) (resid 59 and name n)
       (resid 59 and name ca) (resid 59 and name c) 1.0 -60.0 30.0 2

!     residue 60  L
  assign (resid 59 and name c) (resid 60 and name n)
       (resid 60 and name ca) (resid 60 and name c) 1.0 -60.0 30.0 2

!     residue 62  M
  assign (resid 61 and name c) (resid 62 and name n)
       (resid 62 and name ca) (resid 62 and name c) 1.0 -60.0 30.0 2

!     residue 63  Y
  assign (resid 62 and name c) (resid 63 and name n)
       (resid 63 and name ca) (resid 63 and name c) 1.0 -120.0 40.0 2

!     residue 66  H
  assign (resid 65 and name c) (resid 66 and name n)
       (resid 66 and name ca) (resid 66 and name c) 1.0 -120.0 40.0 2

!     residue 68  V

!     residue 73  S
  assign (resid 72 and name c) (resid 73 and name n)
       (resid 73 and name ca) (resid 73 and name c) 1.0 -60.0 30.0 2

!     residue 77  L
  assign (resid 76 and name c) (resid 77 and name n)
       (resid 77 and name ca) (resid 77 and name c) 1.0 -60.0 30.0 2

!     residue 84  W
  assign (resid 83 and name c) (resid 84 and name n)
       (resid 84 and name ca) (resid 84 and name c) 1.0 -120.0 40.0 2

!     residue 91  H
  assign (resid 90 and name c) (resid 91 and name n)
       (resid 91 and name ca) (resid 91 and name c) 1.0 -120.0 40.0 2
!
            Chi 1 and Chi 2 restraints

!     residue 5  F
   assign (resid 5 and name n) (resid 5 and name ca)
        (resid 5 and name cb) (resid 5 and name cg) 1.0 180.0 20.0 2
!     residue 6  V
```

```
    assign (resid 6 and name n) (resid 6 and name ca)
        (resid 6 and name cb) (resid 6 and name cg1) 1.0 180.0 20.0 2
!    residue 7  N
    assign (resid 7 and name n) (resid 7 and name ca)
        (resid 7 and name cb) (resid 7 and name cg) 1.0 -60.0 20.0 2
!    residue 9  L
    assign (resid 9 and name n) (resid 9 and name ca)
        (resid 9 and name cb) (resid 9 and name cg) 1.0 180.0 20.0 2
!    residue 9  L chi-2
    assign (resid 9 and name ca) (resid 9 and name cb)
        (resid 9 and name cg) (resid 9 and name cd1) 1.0 60.0 20.0 2


!    residue 13  V
!    assign (resid 13 and name n) (resid 13 and name ca)
!        (resid 13 and name cb) (resid 13 and name cg1) 1.0 180.0 20.0 2 rotamer avging

!    residue 21  F
    assign  (resid 21 and name n) (resid 21 and name ca)
        (resid 21 and name cb) (resid 21 and name cg) 1.0 -60.0 20.0 2

!    residue 22  I
    assign (resid 22 and name n) (resid 22 and name ca)
        (resid 22 and name cb) (resid 22 and name cg1) 1.0 180.0 20.0 2
!    residue 22  I chi_2
    assign (resid 22 and name ca) (resid 22 and name cb)
        (resid 22 and name cg1) (resid 22 and name cd1) 1.0 180.0 20.0 2

!    residue 24  W
    assign (resid 24 and name n) (resid 24 and name ca)
        (resid 24 and name cb) (resid 24 and name cg) 1.0 -60.0 20.0 2

!    residue 26  T
    assign (resid 26 and name n) resid 26 and name ca)
        (resid 26 and name cb) (resid 26 and name og1) 1.0 -60.0 20.0 2

!    residue 30  S

!    residue 31
    assign (resid 31 and name n) (resid 31 and name ca)
        (resid 31 and name cb) (resid 31 and name cg1) 1.0 -60.0 20.0 2
    assign (resid 31 and name ca) (resid 31 and name cb)
        (resid 31 and name cg1) (resid 31 and name cd1) 1.0 180.0 20.0 2

!    residue 32  V
    assign (resid 32 and name n) (resid 32 and name ca)
        (resid 32 and name cb) (resid 32 and name cg1) 1.0 180.0 20.0 2

!    residue 33  V
    assign (resid 33 and name n) (resid 33 and name ca)
        (resid 33 and name cb) (resid 33 and name cg1) 1.0 180.0 20.0 2
```

```
!    residue 38  R

!    residue 40  V
     assign (resid 40 and name n) (resid 40 and name ca)
          (resid 40 and name cb) (resid 40 and name cg1) 1.0 180.0 20.0 2

!    residue 42  E
     assign (resid 42 and name n) (resid 42 and name ca)
          (resid 42 and name cb) (resid 42 and name cg) 1.0 -60.0 20.0 2

!    residue 43  V
     assign (resid 43 and name ca) (resid 43 and name cb)
          (resid 43 and name cg1) 1.0 180.0 20.0 2

!    residue 44  L
     assign (resid 44 and name n) (resid 44 and name ca)
          (resid 44 and name cb) (resid 44 and name cg) 1.0 180.0 20.0 2

!    residue 44 L        (chi-2)
     assign (resid 44 and name ca) (resid 44 and name cb)
          (resid 44 and name cg) (resid 44 and name cd1) 1.0 60.0 20.0 2

!    residue 47  Y
     assign (resid 47 and name n) (resid 47 and name ca)
          (resid 47 and name cb) (resid 47 and name cg) 1.0 -60.0 20.0 2

!    residue 48  F
     assign (resid 48 and name n) (resid 48 and name ca)
          (resid 48 and name cb) (resid 48 and name cg) 1.0 -60.0 20.0 2

!    residue 49  K

!    residue 50  H
     assign (resid 50 and name n) (resid 50 and name ca)
          (resid 50 and name cb) (resid 50 and name cg) 1.0 -60.0 20.0 2


!    residue 57  V
     assign (resid 57 and name n) (resid 57 and name ca)
          (resid 57 and name cb) (resid 57 and name cg1) 1.0 180.0 20.0 2

!    residue 59  Q
     assign (resid 59 and name n) (resid 59 and name ca)
          (resid 59 and name cb) resid 59 and name cg) 1.0 -60.0 20.0 2
!    residue 60  L
     assign (resid 60 and name n) (resid 60 and name ca)
          (resid 60 and name cb) (resid 60 and name cg) 1.0 -60.0 20.0 2

!    residue 60  L    (chi 2)
     assign (resid 60 and name ca) (resid 60 and name cb)
          (resid 60 and name cg) (resid 60 and name cd1) 1.0 180.0 20.0 2
```

```
!    residue 63  Y
    assign (resid 63 and name n) (resid 63 and name ca)
        (resid 63 and name cb) (resid 63 and name cg) 1.0 -60.0 20.0 2

!    residue 65  W
    assign (resid 65 and name n) (resid 65 and name ca)
        (resid 65 and name cb) (resid 65 and name cg) 1.0 -60.0 20.0 2

!    residue 66  H
    assign (resid 66 and name n) (resid 66 and name ca)
        (resid 66 and name cb) (resid 66 and name cg) 1.0 -60.0 20.0 2

!    residue 86  F
    assign (resid 86 and name n) (resid 86 and name ca)
        (resid 86 and name cb) (resid 86 and name cg) 1.0 -60.0 20.0 2


    scale=1.0
end
```

Appendix F NOEs,<rmsd>s, angular order parameters

The following programs (Appendix D) were used to generate the data in the respective columns of this table: B (ptnoe.c), C (noeclass_writeall.c -> ptnoe.c), D (ptbkrmsd.c), E (ptnonhrmsd.c), F (phicalc_s.c), G (chi1_s.c), and H (chi2_s.c). Plots of these data are given in Fig. 3.6.

Table F NOEs,<rmsd>s, angular order parameters for high resolution structures

Table of NOEs, <rmsd>s, and angular order parameters for high resolution structures described in Chapter 3.

	number of NOEs		<rmsd>		angular order parameters		
residue	total	long range	backbone	heavy atm	S(ϕ)	S(χ1)	S(χ2)
A	B	C	D	E	F	G	H
A1	1	0	3.1875	3.3940	0.497	0.000	0.000
R2	3	0	1.5225	3.2091	0.213	0.401	0.554
P3	20	3	0.7250	0.8371	0.932	1.000	1.000
A4	10	0	0.6550	0.6980	0.994	0.000	0.000
F5	36	5	0.4925	0.6109	0.988	0.987	0.994
V6	38	2	0.4875	0.5386	0.992	0.991	0.000
N7	20	0	0.5750	0.8200	0.995	0.987	0.284
K8	18	2	0.6150	1.2078	0.980	0.437	0.832
L9	54	7	0.5050	0.5450	0.990	0.979	0.991
W10	36.	1	0.5525	1.2007	0.981	0.979	0.388
S11	7	0	0.5975	0.7683	0.956	0.398	0.000
M12	31	9	0.5075	0.7850	0.981	0.496	0.720
V13	42	14	0.4850	0.4700	0.974	0.994	0.000
N14	16	1	0.7350	1.2413	0.969	0.596	0.545
D15	10	1	1.0975	2.2450	0.497	0.345	0.106
K16	25	0	1.0600	1.9544	0.964	0.501	0.542
S17	14	0	1.1550	1.3050	0.000	0.424	0.000
N18	22	9	0.6250	0.8625	0.760	0.819	0.558
E19	17	0	0.9150	1.4311	0.993	0.690	0.540
K20	18	0	0.8350	1.4578	0.788	0.438	0.555
F21	29	4	0.8825	0.8118	0.957	0.977	0.980
I22	50	18	0.6325	0.6250	0.440	0.995	0.994
H23	9	2	0.5500	1.3900	0.959	0.284	0.173
W24	29	8	0.5250	1.0664	0.983	0.978	0.792
S25	18	5	0.7750	0.9350	0.974	0.456	0.000
T26	13	0	1.4825	1.6267	0.637	0.261	0.000
S27	3	0	2.1375	2.3733	0.799	0.427	0.000
G28	1	0	2.1775	2.2100	0.499	0.000	0.000
E29	13	0	1.6625	2.7378	0.429	0.666	0.538
S30	13	4	0.7275	0.8500	0.641	0.250	0.000
I31	42	17	0.4875	0.5500	0.972	0.991	0.991
V32	35	11	0.5125	0.5800	0.981	0.990	0.000
V33	29	15	0.5125	0.5943	0.984	0.997	0.000
P34	27	3	0.5700	0.6314	0.965	1.000	1.000
N35	13	0	0.6075	1.0538	0.979	0.673	0.350
R36	23	5	0.6775	1.7555	0.990	0.804	0.731

E37	25	1	0.7850	1.4367	0.973	0.564	0.752
R38	18	0	0.5825	1.4282	0.896	0.817	0.508
F39	36	7	0.4675	0.5536	0.997	0.998	0.997
V40	41	3	0.4350	0.5271	0.990	0.994	0.000
Q41	24	0	0.5250	1.1222	0.989	0.813	0.583
E42	23	0	0.5475	0.9544	0.986	0.986	0.561
V43	39	8	0.4450	0.5171	0.964	0.993	0.000
L44	61	13	0.4825	0.5913	0.999	0.996	0.987
P45	16	0	0.5100	0.5386	0.993	1.000	1.000
K46	31	0	0.5825	1.0800	0.994	0.926	0.483
Y47	42	5	0.7200	0.7408	0.975	0.993	0.993
F48	38	2	0.7975	0.9727	0.969	0.979	0.951
K49	29	0	1.4075	1.8567	0.420	0.823	0.559
H50	14	0	1.5550	2.2833	0.195	0.983	0.136
S51	13	2	1.2800	1.4200	0.938	0.483	0.000
N52	14	3	1.1125	2.1050	0.614	0.634	0.171
F53	27	5	0.7275	1.7618	0.963	0.536	0.431
A54	17	0	0.5900	0.6240	0.980	0.000	0.000
S55	11	1	0.5675	0.7433	0.990	0.363	0.000
F56	30	5	0.4925	0.7164	0.985	0.976	0.858
V57	36	5	0.5100	0.5871	0.993	0.993	0.000
R58	17	0	0.4925	1.5436	0.987	0.655	0.483
Q59	27	1	0.4325	0.9267	0.994	0.988	0.424
L60	63	9	0.4725	0.5775	0.998	0.995	0.994
N61	24	0	0.5900	0.9075	0.995	0.897	0.449
M62	21	0	0.6700	1.1287	0.990	0.673	0.466
Y63	30	5	0.7325	0.7250	0.975	0.986	0.993
G64	6	0	0.8975	0.9325	0.918	0.000	0.000
W65	29	6	0.7975	0.7429	0.625	0.995	0.986
H66	17	2	0.9525	1.4067	0.926	0.982	0.116
K67	14	4	1.6100	2.5033	0.860	0.291	0.572
V68	13	2	2.5450	2.7986	0.520	0.421	0.000
Q69	20	0	4.4675	5.6489	0.579	0.442	0.534
D70	6	0	4.6675	5.4850	0.263	0.283	0.103
V71	12	0	5.2800	5.6857	0.450	0.173	0.000
K72	12	0	5.9025	6.6600	0.340	0.457	0.452
S73	6	0	6.8000	7.0117	0.976	0.353	0.000
G74	3	0	7.2300	7.1900	0.205	0.000	0.000
S75	4	0	6.7675	7.0567	0.467	0.331	0.000
M76	14	0	6.2300	6.9437	0.058	0.547	0.536
L77	17	0	6.4475	7.0712	0.969	0.545	0.772
S78	7	0	5.9450	6.2250	0.134	0.132	0.000
N79	3	0	5.0400	5.6038	0.479	0.532	0.323
N80	1	0	3.9875	4.8713	0.370	0.517	0.207
D81	1	0	2.7475	3.4975	0.406	0.466	0.135
S82	10	4	1.2625	1.3900	0.379	0.074	0.000
R83	17	2	1.0750	1.8291	0.837	0.530	0.667
W84	28	8	0.7975	1.1214	0.932	0.716	0.696
E85	20	4	0.6950	1.2567	0.891	0.314	0.710
F86	37	15	0.6350	0.8282	0.943	0.984	0.936
E87	19	7	0.8625	1.6033	0.864	0.431	0.597
N88	10	2	1.6425	1.9787	0.878	0.108	0.259
E89	13	1	2.6575	3.6289	0.276	0.322	0.669
R90	6	0	4.1250	5.4882	0.330	0.417	0.480
H91	8	0	5.9550	6.8067	0.903	0.472	0.157
A92	0	0	7.8625	7.9500	0.450	0.000	0.000

Appendix G Table of Heteronuclear NOE data

These data represent the results of analysis of a reference experiment and a heteronuclear NOE experiment described in Chapter 3. Data were automatically peak-picked using the program CAPP followed by manual editing and additional peak-picking of low threshhold peaks using PIPP. Peak-pick lists generated by PIPP were sorted by chemical shift for the reference and NOE experiment. Comparison of peak-picks in these two lists to the ^{1}H-^{15}N assignments allowed the identification of 79 peaks which could be used to calculate heteronuclear NOEs. The intensities of these peaks are given below (columns B, C, F and G). Two intensities are given in cases were the peaks were split in the ^{1}H dimension due to Hα-HN coupling (upfield peaks in columns C and G). These intensities were averaged to obtain columns E and I. Cells containing zeros (with no decimal place) correspond to data which could not be measured. In a small number of cases one half of a doublet was overlapped with other peaks or streaks so the better half of the doublet was taken (column J). The NOE is computed by dividing entries in column J by entries in column E to give column K. Some NOEs showed unphysical values and were ignored to obtain column L which is plotted in Fig. 3.6F.

Table G Heteronuclear NOE data

Table of Heteronuclear NOE data

	reference intensities				NOE intensities					NOEs	
A	B	C	D	E	F	G	H	I	J	K	L
res	I(1)	I(2)	I(2)'	<I>	I(1)	I(2)	I(2)'	<I>	I(best)	NOE	NOE[b]
A1											0
R2	4.73		4.73	4.73	1.08		1.08	1.08	1.08	0.23	0.23
P3											0
A4	5.74		5.74	5.74	4.69		4.69	4.69	4.69	0.82	0.82
F5	4.77		4.77	4.77	4.10		4.10	4.10	4.10	0.86	0.86
V6	2.63		2.63	2.63	2.61		2.61	2.61	2.61	0.99	0.99
N7	4.56		4.56	4.56	4.08		4.08	4.08	4.08	0.89	0.89
K8	4.37		4.37	4.37	4.00		4.00	4.00	4.00	0.92	0.92
L9	4.41		4.41	4.41	4.01		4.01	4.01	4.01	0.91	0.91
W10	3.54		3.54	3.54	3.22		3.22	3.22	3.22	0.91	0.91
S11	2.88	2.72	2.72	2.80	2.71	2.41	2.41	2.56	2.56	0.91	0.91
M12	4.77		4.77	4.77	4.48		4.48	4.48	4.48	0.94	0.94
V13	3.34		3.34	3.34	3.15		3.15	3.15	3.15	0.94	0.94
N14	0		0	0	3.14		3.14	3.14	3.14	0	0
D15	4.28		4.28	4.28	4.25		4.25	4.25	4.25	0.99	0.99
K16	4.63		4.63	4.63	4.19		4.19	4.19	4.19	0.90	0.90
S17	4.30		4.30	4.30	0		0	0	0	0	0
N18	3.44	2.97	2.97	3.205	3.33	2.62	2.62	2.975	2.975	0.93	0.93
E19	5.54		5.54	5.54	5.18		5.18	5.18	5.18	0.94	0.94
K20	4.46		4.46	4.46	3.98		3.98	3.98	3.98	0.89	0.89
F21	2.62	2.37	2.37	2.495	2.46	2.23	2.23	2.345	2.345	0.94	0.94

I22	3.58		3.58	3.58	3.64		3.64	3.64	3.64	1.02	1.02
H23	0		0	0	0		0	0	0	0	0
W24	3.35		3.35	3.35	3.14		3.14	3.14	3.14	0.94	0.94
S25	2.50		2.50	2.50	2.40	2.16	2.16	2.28	2.28	0.91	0.91
T26	1.72		1.72	1.72	1.35		1.35	1.35	1.35	0.78	0.78
S27	0		0	0	2.80		2.80	2.80	2.80	2.80	0
G28	0		0	0	0		0	0	0	0	0
E29	2.72	2.48	2.48	2.60	2.31	2.08	2.08	2.195	2.195	0.84	0.84
S30	1.98	1.95	1.95	1.965	1.73		1.73	1.73	1.73	0.88	0.88
I31	2.07	2.04	2.04	2.055	2.01	1.75	1.75	1.88	1.88	0.91	0.91
V32	2.64	2.37	2.37	2.505	2.57	2.13	2.13	2.35	2.35	0.94	0.94
V33	2.24	2.16	2.16	2.20	2.23	1.96	1.96	2.095	2.095	0.95	0.95
P34											
N35	2.36	2.06	2.06	2.21	2.12	1.84	1.84	1.98	1.98	0.90	0.90
R36	3.33		3.33	3.33	3.19		3.19	3.19	3.19	0.96	0.96
E37	3.89		3.89	3.89	3.54		3.54	3.54	3.54	0.91	0.91
R38	3.23		3.23	3.23	3.15		3.15	3.15	3.15	0.98	0.98
F39	3.14		3.14	3.14	2.97		2.97	2.97	2.97	0.95	0.95
V40	0		0	0	0		0	0	0	0	0
Q41	3.53		3.53	3.53	3.00		3.00	3.00	3.00	0.85	0.85
E42	0		0	0	0		0	0	0	0	0
V43	2.49	2.26	2.26	2.375	2.57	2.16	2.16	2.365	2.365	1.00	1.00
L44	3.99		3.99	3.99	3.87		3.87	3.87	3.87	0.97	0.97
P45											
K46	3.48		3.48	3.48	3.19		3.19	3.19	3.19	0.92	0.92
Y47	2.69	2.40	2.40	2.545	2.24	2.04	2.04	2.14	2.14	0.84	0.84
F48	3.05	2.74	2.74	2.895	2.82	2.49	2.49	2.655	2.655	0.92	0.92
K49	3.72		3.72	3.72	0		0	0	0	0	0
H50	3.7	3.51	3.51	3.605	2.58	2.58	2.58	2.58	2.58	0.72	0.72
S51	4.41	3.58	3.58	3.995	3.08	2.53	2.53	2.805	2.805	0.70	0.70
N52	0		0	0	0		0	0	0	0	0
F53	4.15		4.15	4.15	3.90		3.90	3.90	3.90	0.94	0.94
A54	4.88		4.88	4.88	4.53		4.53	4.53	4.53	0.93	0.93
S55	4.3		4.30	4.30	4.12		4.12	4.12	4.12	0.96	0.96
F56	3.88		3.88	3.88	4.23		4.23	4.23	4.23	1.09	1.09
V57	3.89		3.89	3.89	3.73		3.73	3.73	3.73	0.96	0.96
R58	4.54		4.54	4.54	4.00		4.00	4.00	4.00	0.88	0.88
Q59	3.52		3.52	3.52	3.33		3.33	3.33	3.33	0.95	0.95
L60	4.35		4.35	4.35	4.19		4.19	4.19	4.19	0.96	0.96
N61	4.14		4.14	4.14	3.90		3.90	3.90	3.90	0.94	0.94
M62	4.77		4.77	4.77	4.16		4.16	4.16	4.16	0.87	0.87
Y63	2.94	2.49	2.49	2.715	2.88	2.35	2.35	2.615	2.615	0.96	0.96
G64	3.05		3.05	3.05	2.65		2.65	2.65	2.65	0.87	0.87
W65	2.67	2.39	2.39	2.53	2.64	2.04	2.04	2.34	2.34	0.92	0.92
H66	2.56	2.32	2.32	2.44	2.28	2.02	2.02	2.15	2.15	0.88	0.88
K67	3.76		3.76	3.76	3.42		3.42	3.42	3.42	0.91	0.91
V68	2.65	2.47	2.47	2.56	2.30	2.21	2.21	2.255	2.255	0.88	0.88
Q69	3.91	3.44	3.44	3.675	3.04	2.63	2.63	2.835	2.835	0.77	0.77
D70	0		0	0	0		0	0	0	0	0
V71	5.6	4.97	4.97	5.285	1.57		1.57	1.57	1.57	0.30	0.30
K72	5.24		5.24	5.24	1.41	1.28	1.28	1.345	1.345	0.26	0.26
S73	3.27	3.19	3.19	3.23	2.63	2.10	2.10	2.365	2.365	0.73	0
G74	5.47		5.47	5.47	0.78		0.78	0.78	0.78	0.14	0.14
S75	6.35	5.99	5.99	6.17	1.10	1.11	1.11	1.105	1.105	0.18	0.18
M76	5.74	5.26	5.26	5.50	1.37	1.21	1.21	1.29	1.29	0.23	0.23
L77	6.03	5.53	5.53	5.78	8.45	7.82	7.82	8.135	8.135	1.41	0
S78	5.91	5.90	5.90	5.905	1.41	1.52	1.52	1.465	1.465	0.25	0.25

N79	4.96	4.49	4.49	4.725	1.58	1.28	1.28	1.43	1.43	0.30	0.30
N80	4.04	3.84	3.84	3.94	1.98	1.91	1.91	1.945	1.945	0.49	0.49
D81	4.38	3.59	3.59	3.985	2.55	1.97	1.97	2.26	2.26	0.57	0.57
S82	5.47	5.50	5.50	5.485	9.74	8.41	8.41	9.075	9.075	1.65	0
R83	2.98	2.83	2.83	2.905	2.76		2.76	2.76	2.76	0.95	0.95
W84	2.32	2.17	2.17	2.245	2.07	2.03	2.03	2.05	2.05	0.91	0.91
E85	3.51		3.51	3.51	2.96		2.96	2.96	2.96	0.84	0.84
F86	2.48	3.24	3.24	2.86	2.29	2.83	2.83	2.56	2.29	0.80	0.80
E87	3.24	2.12	2.12	2.68	2.83	2.08	2.08	2.455	2.08	0.78	0.78
N88	2.28		2.28	2.28	2.00		2.00	2.00	2.00	0.88	0.88
E89	3.96		3.96	3.96	2.94		2.94	2.94	2.94	0.74	0.74
R90	5.34	5.16	5.16	5.25	2.60	2.65	2.65	2.625	2.625	0.49	0.49
H91	6.10	5.87	5.87	5.985	-0.46	-0.48	-0.48	-0.47	-0.47	-0.08	-0.08
A92	7.01	6.45	6.45	6.73	-6.23	-5.23	-5.23	-5.73	-5.73	-0.85	-0.85

Appendix H Table of Amide Exchange Data

^{1}H-^{15}N HMQC exchange spectra were obtained from a sample of uniformly ^{15}N-labeled HSF lyophilized from D_2O and dissolved in H_2O before acquisition. Data are listed in column format. An x indicates the presence an intense peak (o indicates attenuated peak) in spectra recorded 13.1 min (Column B), 23.1 min (Column C), 37.7 min (Column D), 77.7 min (Column E), 156 min (Column F), 180 min (Column G), 376 min (Column H) and 2 weeks after dissolution in D_2O (Column I). Columns J and K list the residue whose backbone carbonyl accepts an H-bond from the amide in the constraints list for structures computed in Chapters 2 and 3 respectively and column L lists residues accepting H-bonds in the X-ray structure (using the program hbondrev2.com in Appendix D). Differences in H-bonding partners are indicated in bold.

Table H Amide Exchange Data

Table of Amide Exchange Data

	Exchange Data								H-bonds		
A	B	C	D	E	F	G	H	I	J	K	L
res	13.1	23.1	37.7	77.7	156	180	376	2 wks	Acc 2	Acc 3	Acc X
A1											
R2											
P3	NA										
A4	x	x									
F5	x	x									
V6											
N7	x	x	x	x	x	x			3	3	3
K8	x	x	x	x	x	x			4	4	4
L9	x	x	x	x	x	x			5	5	5
W10											**6**
S11											**7**
M12											**8**
V13	x	x	x	x	x	x	x	x	9	9	9
N14	x	x	x	x	x	x	x				**10**
D15	x	x	x	x	x	x	x				
K16	x	o									
S17	x	x	x	x	x	x	x				
N18											**15**
E19	x	x	x	x	x	x	x				**16**
K20	x	x									
F21	x	x	x	x	x	x	x			18	18
I22											
H23											

A res	B 13.1	C 23.1	D 37.7	E 77.7	F 156	G 180	H 376	I 2 wks	J Acc 2	K Acc 3	L Acc X
W24	x	x	x	x	x	x					
S25	x	x	x	x	x	x	x			30	30
T26											
S27											
G28											
E29	x	x	x	x	x	x	x				
S30	x	x	x	x	x	x	o				
I31	x	x	x	x	x	x	x	x	86	86	86
V32	x	x	x	x	x	x	x	x	23	23	23
V33	x	x	x	x	x	x	x	x	84	84	84
P34	NA										
N35	x	x	x	x	x	x	x				
R36											
E37	x	x	x	x	x						
R38	x	x	x	x	x	x	x				
F39	x	x	x	x	x	x	x				**35**
V40											**36**
Q41											**37**
E42	x	x	x	x	x	x	x			38	38
V43	x	x	x	x	x	x	x	o			**39**
L44	x	x	x	x	x	x	x			40	**39**
P45	NA										
K46	x	x	x	x	x	x	x			43	43
Y47	x	x	x	x	x	x	x	o			**44**
F48	x	x	x	x	x	x	x			44	44
K49											
H50											
S51											
N52											
F53											
A54	x	x	x	x	x	x	o				
S55	x	x	x	x	x	x	x				
F56	x	x	x	x	x	x	x		52		52
V57	x	x	x	x	x	x	x	x	53	53	53
R58	x	x	x	x	x	x	x	x	54	54	54
Q59	x	x	x	x	x	x	o		55	55	55
L60	x	x	x	x	x	x	x	x	56	56	56
N61	x	x	x	x	x	x	x		57	57	57
M62	x	x	x	x	x	x	x		58	58	**59**
Y63											**60**
G64	x	x	x	x	x	x	o				**61**
W65	x	x	x	x	x	x	x				**60**
H66	x	x	x	x	x	x	x		87	87	87
K67	x	x									
V68	x	x	x	x	x	x	x				
Q69	x	x									
D70											
V71	x	x									
K72	x	x									
S73											
G74											
S75											
M76											

A res	B 13.1	C 23.1	D 37.7	E 77.7	F 156	G 180	H 376	I 2 wks	J Acc 2	K Acc 3	L Acc X
L77	x	x	x	x							
S78											
N79											
N80											
D81											
S82											
R83											
W84	x	x	x	x	x	x	o				**33**
E85	x	x	x								
F86	x	x	x	x	x	x	o			31	31
E87	x	x	x	x	x	x					**66**
N88											
E89											
R90	x	x									
H91											
A92											

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
1 CYCLOTRON ROAD
BERKELEY, CALIFORNIA 94720