Automated Stereospecific ^{1}H NMR Assignments and Their Impact on the Precision of Protein Structure Determinations in Solution

Peter Güntert, Werner Braun, Martin Billeter, and Kurt Wüthrich*

Contribution from the Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule—Hönggerberg, CH-8093 Zürich, Switzerland. *Received June 30, 1988*

Abstract: Two sets of constraints on proton–proton distances and dihedral angles, which mimic data that can be obtained from nuclear magnetic resonance experiments in solution, were derived from the crystal structure of the protein basic pancreatic trypsin inhibitor (BPTI). In one of these data sets, all prochiral groups of protons were replaced by pseudoatoms. In the second set, stereospecific assignments were used for all β-methylene groups, all protons of glycine and proline, the methyl groups of valine and leucine, and the ring protons of phenylalanine and tyrosine. Comparison of the BPTI structures calculated from these data with the distance geometry program DISMAN showed that, with otherwise identical distance constraints, the use of stereospecific assignments results in significantly improved precision of the structure determination for the polypeptide backbone as well as the amino acid side chains. The paper further describes the program HABAS, which determines stereospecific assignments by a systematic analysis of the proton–proton scalar couplings and the intraresidual and sequential proton–proton nuclear Overhauser effects. To investigate to what extent stereospecific assignments could be obtained for a predetermined completeness and precision of the input data set, HABAS was used for test calculations with a standard dipeptide unit and a database derived from a group of high-resolution protein crystal structures. From these data we estimate that with the precision presently achieved for NMR measurements with proteins, stereospecific assignments can be obtained for approximately half of the β-methylene protons. Quite generally, this ratio can be expected to be higher for β-proteins than for those that contain predominantly α-helical secondary structure.

I. Introduction

It has by now been quite widely accepted that as a second method besides X-ray diffraction in single crystals, nuclear magnetic resonance (NMR)[1] in solution can be used for the determination of the complete three-dimensional structure of proteins.[2-6] Presently, considerable effort is directed at improvements of the efficiency as well as the precision of such structure determinations, for example, with the use of ever more sophisticated NMR techniques.[6,7] and mathematical methods for the structural interpretation of the NMR data.[6,8-13] It is a fundamental advantage of the NMR method for protein structure determination that it can depend on qualitative experimental constraints on the conformation,[2,10,12] which makes it both robust and efficient in practical applications. Besides experimental limitations, quantitative distance measurements would be intrinsically difficult because the observed NOEs depend not only on the proton–proton distances but also on the effective rotational correlation times,[14-16] which may be variable for different locations in a protein molecule.[6,16] Stereospecific assignments for prochiral groups of protons can yield more precise structures without requirements for more quantitative distance measurements. It had already been demonstrated that the precision of protein solution conformations determined from qualitative NMR constraints can be comparable to that of a refined high-resolution crystal structure,[17] provided that a sufficiently large number of constraints is available, and that as far as possible stereospecific assignments were determined for the prochiral groups of protons.[18] The present paper describes investigations of the improvements in the precision of protein structure determinations that can be anticipated from stereospecific assignments, using input data sets derived from the crystal structure of the protein BPTI. It further introduces the program HABAS, which performs an automated analysis of the experimentally accessible, local NMR parameters to obtain stereospecific ^{1}H NMR assignments before the start of the distance geometry calculations.

The generally used sequential resonance assignment procedure for proteins[2,6] does not yield stereospecific assignments for the individual protons in prochiral groups. To deal with this situation a set of pseudoatoms replacing the prochiral groups was introduced.[19] This is inevitably a compromise, since the use of these pseudoatoms reduces the precision of the experimental conformational constraints.[6] In special situations some stereospecific assignments resulted in the course of the three-dimensional structure determination,[18,20,21] and recently, a procedure for obtaining stereospecific assignments during metric matrix distance geometry calculations was proposed.[22] Conversely, empirical procedures for obtaining stereospecific assignments before the structure determination have also been described, which use

(1) Abbreviations used: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; BPTI, basic pancreatic trypsin inhibitor; RMSD, root-mean-square distance; DISMAN, distance geometry program for proteins; HABAS, program for obtaining stereospecific resonance assignments for α- and β-protons in proteins.

(2) Wüthrich, K.; Wider, G.; Wagner, G.; Braun, W. *J. Mol. Biol.* **1982**, *155*, 311–319.
(3) Braun, W.; Wider, G.; Lee, K. H.; Wüthrich, K. *J. Mol. Biol.* **1983**, *169*, 921–948.
(4) Williamson, M. P.; Havel, T. F.; Wüthrich, K. *J. Mol. Biol.* **1985**, *182*, 295–315.
(5) Kline, A. D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1986**, *189*, 377–382.
(6) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(7) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press: Oxford, U.K., 1987.
(8) Havel, T. F.; Wüthrich, K. *Bull Math. Biol.* **1984**, *46*, 673–698.
(9) Kaptein, R.; Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F. *J. Mol.* Biol. **1985**, *182*, 179–182.
(10) Braun, W.; and Gō, N. *J. Mol. Biol.* **1985**, *186*, 611–626.
(11) Clore, G. M.; Gronenborn, A. M.; Brünger, A. T.; Karplus, M. *J. Mol. Biol.* **1985**, *186*, 435–455.
(12) Havel, T. F.; Wüthrich, K. *J. Mol. Biol.* **1985**, *182*, 281–294.
(13) Braun, W. *Q. Rev. Biophys.* **1987**, *19*, 115–157.
(14) (a) Solomon, I. *Phys. Rev.* **1955**, *99*, 559–565. (b) Noggle, J. H.; Schirmer, R. E. *The Nuclear Overhauser Effect*; Academic Press: New York, 1971.
(15) Wagner, G.; Wüthrich, K. *J. Magn. Reson.* **1979**, *33*, 675–680.
(16) Olejniczak, E. T.; Dobson, C. M.; Karplus, M.; Levy, R. M. *J. Am. Chem. Soc.* **1984**, *106*, 1923–1930.
(17) Billeter, M.; Kline, A. D.; Braun, W.; Huber, R.; Wüthrich, K. *J. Mol. Biol.*, in press.
(18) (a) Kline, A. D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1988**, *204*, 675–724. (b) Billeter, M.; Schaumann, Th.; Braun, W.; Wüthrich, K. *Biopolymers*, in press.
(19) Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1983**, *169*, 949–961.
(20) Senn, H.; Billeter, M.; Wüthrich, K. *Eur. Biophys. J.* **1984**, *11*, 3–5.
(21) Zuiderweg, E. R. P.; Boelens, R.; Kaptein, R. *Biopolymers* **1985**, *24*, 601–611.
(22) Weber, P. L.; Morrison, R.; Hare, D. *J. Mol. Biol.* **1988**, *204*, 483–487.

0002-7863/89/1511-3997$01.50/0

manual screening of the spin–spin couplings and the intraresidual and sequential NOEs.[18,23,24] The program HABAS replaces these empirical approaches by an unbiased screening of the local constraints, whereby for the torsion angle χ^1 either a continuous population of all values from 0 to +180°, or of limited ranges near the staggered rotamers, for example, 60 ± 20, 180 ± 20, and −60 ± 20°, can be assumed.[25] In addition to the determination of stereospecific assignments, HABAS analyses the local NMR constraints in terms of allowed regions in local conformation space, rather than individual, discrete points thereof. The starting structures for distance geometry calculations, e.g., using the program DISMAN,[10,13] can then be chosen randomly from this locally constrained conformation space.

In the first part of this paper the study of the influence of stereospecific assignments on the precision of protein structure determination by NMR uses the assumption that stereospecific assignments are available for distinct classes of prochiral groups of protons. While this provides a useful, general guideline, it is further of interest to assess the extent to which stereospecific assignments can be derived, depending on the completeness and precision of the available NMR data. To this end the program HABAS is applied for test calculations using simulations of NMR input data derived from a group of high-resolution crystal structures of small proteins.

II. Test Calculations on the Impact of Stereospecific Assignments on the Precision of a Protein Structure Determination

To investigate the influence of stereospecific resonance assignments on the precision of protein structure determinations by ^{1}H NMR, test calculations with the small globular protein BPTI (58 residues) were carried out. Two sets of conformational constraints were derived from the regularized crystal structure of BPTI, which differed only in stereospecific assignments. In the data set NOST no stereospecific assignments were used and all prochiral groups of protons were represented by pseudoatoms.[19] When preparing the data set WIST it was assumed that stereospecific assignments were available for the following groups: β-methylene protons, α-methylene protons of glycine, γ- and δ-methylene protons of proline, methyl groups of valine and leucine, and ring protons of tyrosine and phenylalanine. The remaining prochiral groups of protons were again represented by pseudoatoms. The data set WIST thus contained stereospecific assignments for 73 of the total of 101 prochiral groups of protons in BPTI. With each of the two constraint sets, four structures were calculated by using the distance geometry program DISMAN.[10] These two groups of structures were then compared with each other and with the regularized crystal structure from which the input data had been obtained.

Simulated Input Data Sets. Because the program DISMAN works with fixed bond lengths and bond angles, the simulated constraint sets for the test calculations with DISMAN were extracted from a structure with standard geometry of the amino acid residues. For this the BPTI crystal structure[26] (code of the Protein Data Bank: 4PTI)[27] was regularized with the program DISMAN, by using 2990 exact distances from the unregularized crystal structure as constraints. The resulting regularized crystal structure, XRAY, with the desired standard geometry and all hydrogen atoms attached, coincided closely with the unregularized structure, with RMSD values[28] of 0.27 Å for the backbone atoms and 0.35 Å for all heavy atoms. (This difference is small compared to the difference between the two crystal forms I and II of BPTI, where the RMSD for the backbone atoms is 0.4 Å.[29]) The regularized structure contained seven violations of steric constraints greater than 0.2 Å; the maximal violation was 0.32 Å. The advantage of using a regularized structure as the source of the distance constraints is that this ensures a clearcut distinction between possible effects arising either from the limited accuracy of the simulated data or from distortions of the standard geometry.[10,13]

Distance constraints were derived from the regularized crystal structure following the strategy previously used for test calculations without stereospecific assignments,[8,12] whereby for the steric constraints the standard parameters employed with DISMAN were used.[10] All interresidual proton–proton distances shorter that 4.0 Å were considered, as well as the intraresidual distances shorter than 4.0 Å from backbone amide or α-protons to the side-chain protons attached to C^γ or beyond. These precise distances were substituted by corresponding upper limits on the distances in order to mimic a typical NMR input for a structure calculation.[6] For the long-range constraints, the upper limit was 4.0 Å throughout. For the intraresidual constraints and for constraints between protons in sequentially adjacent residues, upper limits of 2.5, 3.0, 3.5, and 4.0 Å were used, where the limit <2.5 Å applies to all distances shorter than 2.5 Å, the limit <3.0 Å to all distances in the range from 2.5 to 3.0 Å, etc. Whenever a prochiral group of protons was represented by a centrally located pseudoatom, the appropriate correction was added to these upper bounds.[6,19] A survey of the distance constraints used is afforded by Table I. [In addition, the constraints on the intraresidual distances $d_{N\beta2}(i,i)$

Table I. Survey of the Distance Constraints Used in the Input for the Distance Geometry Calculations with BPTI and Analysis of the Residual Constraint Violations

type of constraint[a]	NOST no. of constraints	NOST violations[b] >0.2 Å	NOST violations[b] >0.5 Å	WIST no. of constraints	WIST violations[b] >0.2 Å
intraresidual	78	0.5	0.3 (0.76)	88	0.3 (0.22)
neighbor residue	172	1.3	0.5 (0.93)	211	1.5 (0.31)
long range	379	4.3	1.0 (0.80)	528	2.0 (0.37)
steric		6.3	0.3 (0.58)		2.5 (0.43)

[a] *Neighbor residue* distance constraints are between atoms in sequentially neighboring residues. All other interresidual constraints are *long range*. The steric lower limit distance constraints are those imposed by the van der Waals volumes of the atoms as described in ref 10. [b] Four structures were calculated for each of the two input data sets, NOST and WIST. Among the four structures the average number of residual violations exceeding the indicated limit is given, and the values in parentheses are the largest individual residual violations.

Table II. Survey of the Dihedral Angle Constraints Used in the Input for the Distance Geometry Calculations with BPTI[a] and Analysis of the Residual Violations

type of constraint[b]	NOST no. of constraints	NOST violations[c] >5°	WIST no. of constraints	WIST violations[c] >5°
$0° < \Delta\phi \le 90°$	32	0.5 (10.8)	32	
$90° < \Delta\phi \le 300°$	9		9	
$0° < \Delta\psi \le 90°$	35	0.3 (16.4)	35	
$90° < \Delta\psi \le 300°$	6		6	
$0° < \Delta\chi^1 \le 90°$	21	0.5 (6.8)	38	
$90° < \Delta\chi^1 \le 300°$	20	0.3 (5.5)	3	

[a] These dihedral angle constraints resulted from a combined analysis of the spin–spin coupling constants $^3J_{HN\alpha}$, $^3J_{\alpha\beta2}$, and $^3J_{\alpha\beta3}$, the intraresidual distance constraints $d_{N\beta2}(i,i)$ and $d_{N\beta3}(i,i)$, and the sequential distance constraints $d_{\alpha N}$, d_{NN}, $d_{\beta2N}$, and $d_{\beta3N}$ using the program HABAS. [b] $\Delta\phi$, $\Delta\psi$, and $\Delta\chi^1$ indicate the full size of the allowed dihedral angel intervals. [c] Four structures were calculated for each of the two input data sets, NOST and WIST. Among the four structures the average number of residual violations exceeding 5° is given, and in parentheses, the largest of these violations is indicated.

(23) Hyberts, S.; Märki, W.; Wagner G., *Eur. J. Biochem.* **1987**, *164*, 625–635.

(24) (a) Arseniev, A.; Schultze, P.; Wörgötter, E.; Braun, W.; Wagner, G.; Vašák, M.; Kägi, J. H. R.; Wüthrich, K. *J. Mol. Biol.* **1988**, *201*, 637–657. (b) Wagner, G.; Braun, W.; Havel, T. F.; Schaumann, T.; Gō, N.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 611–639.

(25) (a) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 755–791. (b) McGregor, M. J.; Islam, S. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1987**, *198*, 295–310.

(26) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr. B* **1983**, *39*, 480–490.

(27) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.

(28) (a) Nyburg, S. C. *Acta Crystallogr. B* **1974**, *30*, 251–253. (b) McLachlan, A. D. *J. Mol. Biol.* **1979**, *128*, 49–79.

(29) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J. Mol. Biol.* **1984**, *180*, 301–329.

Table III. Average and Standard Deivations of the Pairwise RMSD Values among the BPTI Structures Used in This Study

structures compared[a]	RMSD[b] backbone (N, C^{α}, C'), Å			RMSD[b] all heavy atoms, Å		
	3–55	β-sheet	α-helix	3–55	β-sheet	α-helix
XRAY/NOST	0.9 ± 0.2	0.2 ± 0.1	0.7 ± 0.3	1.6 ± 0.3	1.4 ± 0.3	1.3 ± 0.3
XRAY/WIST	0.4 ± 0.1	0.2 ± 0.1	0.2 ± 0.1	1.1 ± 0.1	1.1 ± 0.1	0.7 ± 0.1
NOST/WIST	1.0 ± 0.3	0.3 ± 0.1	0.7 ± 0.3	1.7 ± 0.2	1.3 ± 0.3	1.4 ± 0.2
NOST/WIST	1.1 ± 0.2	0.3 ± 0.1	0.9 ± 0.3	1.8 ± 0.3	1.4 ± 0.2	1.4 ± 0.3
WIST/WIST	0.5 ± 0.1	0.2 ± 0.1	0.2 ± 0.1	1.2 ± 0.1	0.9 ± 0.2	0.8 ± 0.1

[a] XRAY is the structure 4PTI from the Protein Data Bank[26,27] after regularization. NOST and WIST are the two groups of four structures calculated from the corresponding input data sets (see text). For example, the comparisons XRAY/NOST and NOST/WIST yield 4 and 16 pairwise RMSDs, respectively. [b] RMSDs were calculated for the residues 3–55, rather than for the complete structure (residues 1–58), to exclude chain termination effects. The β-sheet and the α-helix comprise the residues 18–35 and 48–55, respectively.

and $d_{N\beta3}(i, i)$ are included implicitly in the constraints on the dihedral angles ϕ, ψ, and χ^1 listed in Table II].

Allowed intervals for the dihedral angles ϕ, ψ, and χ^1 were determined by the program HABAS (See section III below. Note that in this application HABAS has not been used to obtain stereospecific assignments. The determinations of allowed dihedral angle intervals is a part of the program that is independent from the stereospecific assignment part. Restrictions on the allowed dihedral angle ranges may result even if no unambiguous stereospecific assignments can be derived from the available data). For this, spin–spin coupling constants were calculated from the regularized crystal structure by using Karplus-type relations calibrated for use with proteins.[30,31] To mimic the precision of a typical NMR experiment, these J values were taken to define the center of an interval of half-width 2.0 Hz. These intervals were then combined with the sequential distance constraints $d_{\alpha N}$, d_{NN}, $d_{\beta 2N}$, and $d_{\beta 3N}$ and the intraresidual constraints $d_{N\beta2}(i, i)$ and $d_{N\beta3}(i, i)$ (see ref 6 for the notation used) to define allowed ranges for the dihedral angles. For the residues for which stereospecific assignments had been assumed in WIST, these allowed ranges were further confined so as to include only the values that were compatible with the correct assignments. A survey of all dihedral angle constraints thus obtained is afforded by Table II. In addition to the data in Tables I and II, the input for the DISMAN calculations contained three constraints for each of the disulfide bonds, using the parameters described by Williamson et al.[4]

Results. The DISMAN program has several options for generating starting structures. The option used here for the calculations with both NOST and WIST was to choose the variable dihedral angles within those limits that are allowed by the $d_{\alpha N}$, d_{NN}, and ϕ distance and dihedral angle constraints. Different starting structures were generated for the calculations with NOST and WIST, respectively. Convergent structures for the constraint sets NOST and WIST were selected according to their residual constraint violations. In Tables I and II it is shown that the four final structures of each group satisfy nearly all distance and dihedral angle constraints perfectly. The number of violations of distance constraints by more than 0.2 Å or angle constraints by more than 5° is always small relative to the total number of constraints, whereby the structures obtained from the data set WIST converged slightly better than the structures obtained without stereospecific assignments.

In the following, two criteria are used to evaluate the calculated structures. One is the average RMSD relative to the regularized crystal structure from which the input data were taken, which indicates how faithfully this structure was reproduced. The second criterion is the average RMSD among the four structures in each group, which indicates how precisely the atom positions are determined and further provides information on the sampling by the DISMAN program.

RMSD values were calculated separately for the backbone, and for the complete structure including the amino acid side chains (Table III). Table III shows that the improvement of the structures with stereospecific assignments is particularly pronounced for the polypeptide backbone. The backbone atoms of the WIST structures are significantly nearer to the corresponding atoms in the regularized crystal structure than those of the NOST structures, with an RMSD value of 0.4 Å as compared to 0.9 Å. The RMSDs among the different structures of the group NOST are bigger than among the different structures WIST. The RMSD values for all heavy atoms show a similar reduction, 0.6 Å, between the structures NOST and WIST. As the absolute RMSD values of the backbone atoms are smaller than those for all heavy atoms, the relative improvement with the use of stereospecific assignments is particularly striking for the backbone. The differences between the average RMSD values for the WIST and NOST structures are in all comparisons larger than the standard deviations for the pairwise RMSDs among the individual structures within each group, which emphasizes that the improvements achieved with the stereospecific assignments are indeed significant. To study the influence of stereospecific assignments on the regular secondary structures, separate RMSD values were calculated for the β-sheet region 18–35 and for the α-helical region 48–55. The improvement in the α-helical region is especially pronounced. The RMSD values dropped from 0.9 Å among all NOST structures to 0.2 Å among all WIST structures (Table III). The polypeptide backbone fold of the NOST structures in the β-sheet region is already well-defined, and here the improvement with stereospecific assignments is only marginal. When the RMSDs for all heavy atoms are considered, one finds similar improvements with stereospecific assignments for the β-sheet, the α-helix, and the complete molecule (Table III).

Visual impressions of the results in Table III are afforded by Figures 1–3, which show molecular models produced with the molecular graphics program CONFOR.[32] Figure 1A shows the regularized crystal structure (thick line) superimposed with the four NOST structures. Figure 1B shows the spread of the four WIST structures (thin lines) around the regularized crystal structure. The distribution of the structures calculated with the NOST and WIST input data around the regularized crystal structure, from which the input data were taken, is consistent with unbiased sampling (Figure 1 and Table III). The regions that contribute most to the improvements of the backbone conformations in the WIST structures with respect to the NOST structures are the segments 5–10, the β-turn region near 25, the loop at 36–40, and the α-helical region 48–55. In Figure 2 the side chains are also shown. While the improvement in the backbone of the β-sheet is only marginal, some side chains that were not well-defined by the NOST data set were significantly more tightly constrained, e.g., Ile-18, Ile-19, Leu-29, Phe-33, and Tyr-35. The aromatic side chains of Tyr-21, Phe-22, and Tyr-23 are already quite well confined in the NOST structures, which is probably largely due to the internal packing restrictions.[6] For these side chains the improvement by the stereospecific assignments is less pronounced.

Figure 3 presents two direct comparisons of the four structures NOST with the four structures WIST. Figure 3A illustrates the improved precision of the backbone structure determination by

(30) Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741–751.

(31) DeMarco, A.; Llinás, M.; Wüthrich, K. *Biopolymers* **1978**, *17*, 617–636.

(32) Billeter, M.; Engeli, M.; Wüthrich, K. *Mol. Graphics* **1985**, *3*, 79–83, 97–98.

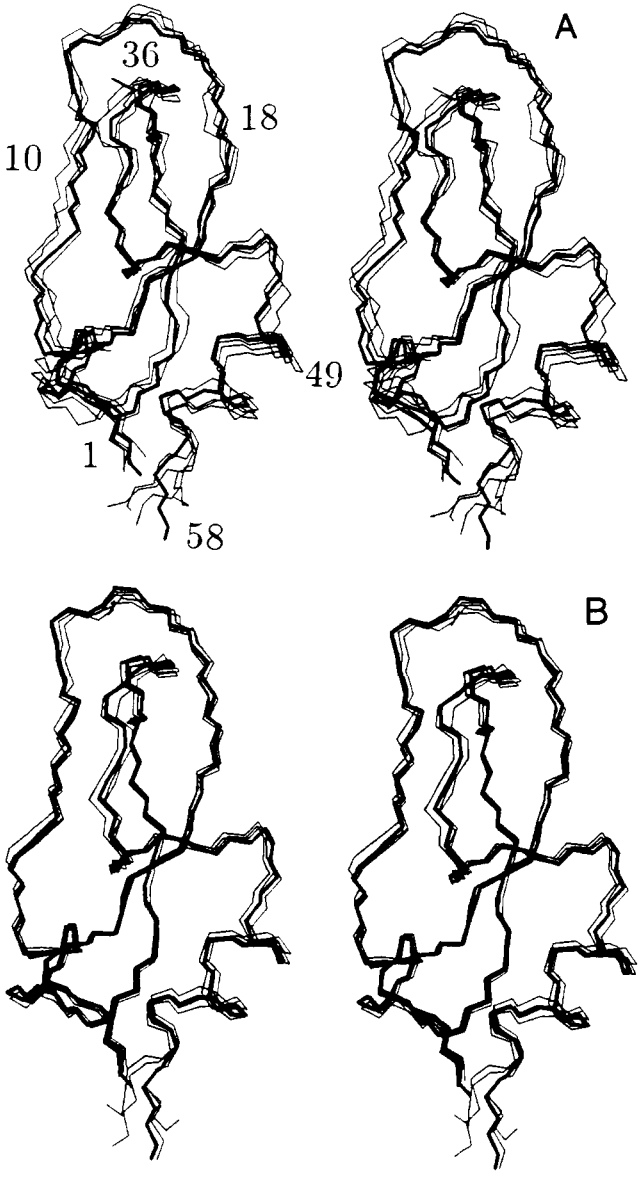

Figure 1. Stereo views affording comparisons of the regularized crystal structure of BPTI (thick line) with (A) four structures calculated from the data set NOST (thin lines) and (B) four structures calculated from the data set WIST (thin lines). The bonds connecting the backbone atoms N, C^{α} and C′ are shown, and the structures were superimposed for minimal pairwise RMSD of these atoms between the individual calculated structures and the regularized crystal structure. In (A) the locations of selected residues are indicated.

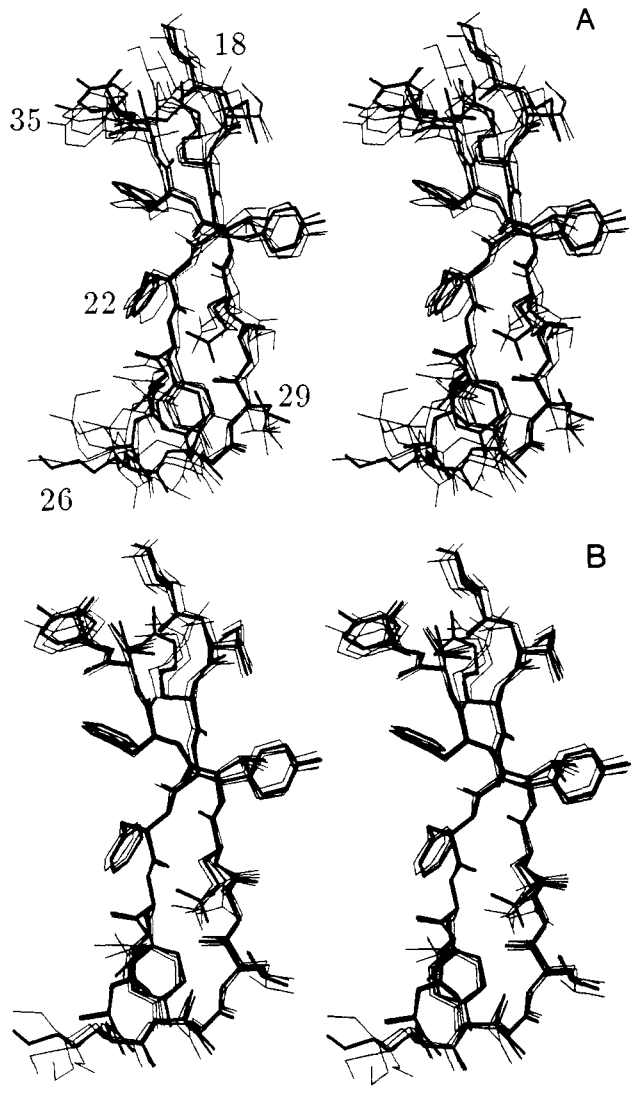

Figure 2. Same as Figure 1, except that all heavy atoms of the β-sheet formed by the residues 18–35 in BPTI are shown. In (A) the regularized crystal structure (thick line) is shown superimposed with four structures calculated from NOST and in (B) with four structures calculated from WIST. In (A) the locations of some residues are identified.

the four structures WIST. It further reveals a tendency of the BPTI structures calculated from the data with stereospecific assignments to be somewhat contracted relative to NOST, i.e., to have slightly reduced global dimensions. Figure 3B identifies in a clear fashion the sequence regions with least well determined spatial backbone structure. Thereby it is quite striking that the regions near residues 27 and 37 have not only the largest dispersion among the four NOST structures, but show also the most pronounced improvement when stereospecific assignments are used.

III. The Program HABAS for Automated Determination of Stereospecific ^{1}H NMR Assignments

The program HABAS systematically scans experimentally determined sets of structural constraints in proteins in order to obtain stereospecific assignments for β-methylene protons in amino acid side chains, and for the γ-methyl groups of valine. It is applied before the start of the structure calculations and uses experimental data corresponding to local conformational constraints that are available after determination of the sequence-specific ^{1}H NMR assignments.[6] No medium-range or long-range NOEs are considered. For a particular residue i, HABAS analyzes the constraints that depend only on the three dihedral angles ϕ_i, ψ_i, and χ_i^1 (Figure 4). This includes steric constraints, allowed ranges for proton–proton distances, relations between pairs of such distances, and spin–spin coupling constants. The following describes how these input data are handled in the preparation of the input. (HABAS is available upon request addressed to the authors.)

To describe steric constraints, a repulsive core radius is assigned to each atom in the polypeptide chain. A pair of atoms violates a steric constraint if the distance between the two atoms is smaller than the sum of their repulsive core radii. The same core radii were used as in the program DISMAN.[10] Upper bounds on ^{1}H–^{1}H distances are obtained from the corresponding ^{1}H–^{1}H NOEs, where HABAS makes use of the constraints on the intraresidual distances $d_{N\beta 2}(i, i)$ and $d_{N\beta 3}(i, i)$ and the sequential distances $d_{\alpha N}$, d_{NN}, $d_{\beta 2N}$, and $d_{\beta 3N}$ (see ref 6 for the notation used). In addition, for valine the intraresidual and sequential distances between amide protons and the γ-methyl groups are also considered.

Besides the constraints on individual distances, HABAS also accepts relations between the two distances from a proton A to the two protons B and B′ of a methylene group. The relational constraint is fulfilled if $d(A, B) > d(A, B') + \Delta d$, where Δd is an arbitrary parameter (usually $\Delta d = 0$). This option of HABAS takes into account that relative values for two NOEs are often

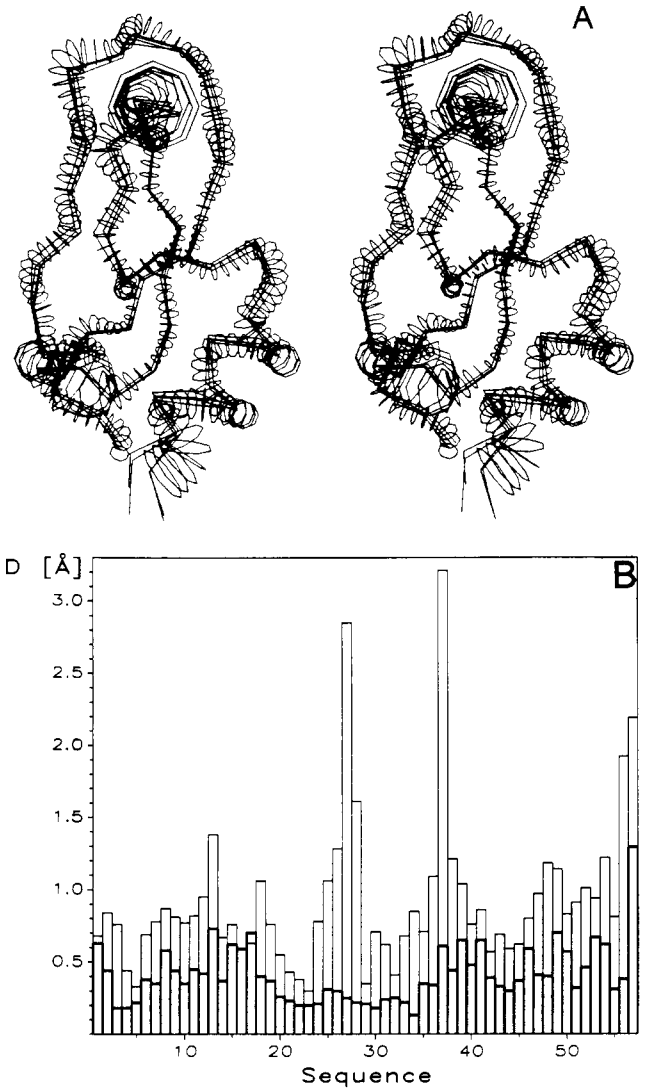

Figure 3. (A) Stereo view affording a comparison of the two groups of four BPTI structures calculated from the data sets NOST and WIST, respectively. All structures were superimposed so as to minimize the RMSD of the C^α atoms with respect to those in the regularized crystal structure. The four structures WIST are represented by straight lines connecting the C^α positions. The structures NOST are represented by a single set of circles, with five circles per residue. The centers of these circles were obtained by calculating the average of the C^α positions in the four structures NOST and fitting a spline function through these average positions. The planes of the circles are perpendicular to this spline function. At the position of a given C^α atom the radius of the circle is equal to the largest of the four distances between this C^α atom in the individual structures and in the average, andd the radii of the four circles between two neighboring C^α positions are smoothly interpolated. (B) Plot versus the amino acid sequence of the largest of the four displacements, D, between a given C^α atom in the individual structures and the corresponding average for the structures calculated from the data set NOST (thin line) and those from WIST (thick line). The values for the displacements plotted here for NOST correspond to the radii of the circles around the C^α positions in (A). For the C-terminal residue Ala-58 the displacements among the four structures NOST are very large, and they were therefore not included in the drawings.

more easily accessible than estimates of absolute distance values.

To determine allowed intervals for individual dihedral angles, HABAS makes use of the following Karplus-type relations, which were calibrated with experimental data in peptides and proteins.[30,31]

$$^3J_{HN\alpha}(\phi) = 6.4\cos^2(\phi - 60°) - 1.4\cos(\phi - 60°) + 1.9 \quad (1)$$

$$^3J_{\alpha\beta_2}(\chi^1) = 9.5\cos^2(\chi^1 - 120°) - 1.6\cos(\chi^1 - 120°) + 1.8 \quad (2)$$

$$^3J_{\alpha\beta_3}(\chi^1) = 9.5\cos^2\chi^1 - 1.6\cos\chi^1 + 1.8 \quad (3)$$

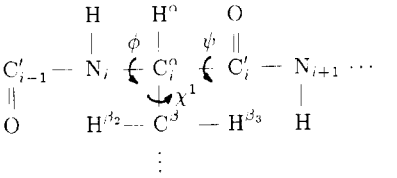

Figure 4. Dipeptide segment examined by the program HABAS in each step of the calculation. A residue with a β-methylene group is shown.

The experimental data, $^3J^{exp}$, are supplemented with an arbitrary parameter, Δ^3J, defining the precision of the experiment (Δ^3J is usually chosen in the range 1.0–2.0 Hz). The corresponding dihedral angle is then constrained within the range that corresponds to the interval of spin–spin coupling constants from ($^3J^{exp} - \Delta^3J$) to ($^3J^{exp} + \Delta^3J$). Note that eq 3 applies for the β-methine proton in Val and eq 2 for the β-methine proton in Ile or Thr.

To determine stereospecific assignments for a pair of β-methylene protons, HABAS goes through a process that corresponds to two subsequent, independent grid searches of the three-dimensional space defined by ϕ, ψ, and χ^1 for conformations that fulfill all experimental constraints. The two grid searches are for the two possible stereospecific assignments, and for each assignment, the number of conformations that are consistent with all constraints is computed. If all those conformations fulfil the constraints for only one of the two possible stereospecific assignments, then this stereospecific assignment is considered to be unambiguously identified by the input data used. In the present form, the program applies a grid search with steps of $\Delta\phi = \Delta\psi = \Delta\chi^1 = 10°$. The values for proton–proton distances and spin–spin couplings, which are needed for the grid search, are obtained from the peptide segments Ala-Ser-Ala, representing all non-proline residues, and Ala-Pro-Ala in the ECEPP standard geometry.[33,34] Of the $36^3 = 46\,656$ conformations generated in the course of this grid search, many are not allowed due to steric hindrance (see ref 10 for the core radii used). Therefore, only 13 050 conformations need to be checked against the experimental data.

Valine is treated as a special case of the non-proline residues. In the place of C^γ and $H^{\beta 2}$ two pseudo atoms $Q^{\gamma 1}$ and $Q^{\gamma 2}$ representing the γ-methyl groups[19] are attached to C^β. The scalar coupling between the α-proton and the β-methine proton is analyzed with eq 3, and the resulting information on χ^1 is combined with the distance constraints for the intraresidual and sequential distances $d_{NQ^{\gamma 1}}(i, i)$, $d_{NQ^{\gamma 2}}(i, i)$, $d_{N\beta}(i, i)$, $d_{Q^{\gamma 1}N}(i, i+1)$, $d_{Q^{\gamma 2}N}(i, i+1)$, $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ to determine the stereospecific assignments of the γ-methyl groups.

(33) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361–2381.

(34) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883–1887.

(35) Teeter, M. M. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6014–6018.

(36) Leijonmarck, M.; Liljas, A. *J. Mol. Biol.* **1987**, *195*, 555–579.

(37) Carter, C. W., Jr.; Kraut, J.; Freer, S. T.; Xuong, N.-H.; Alden, R. A.; Bartsch, R. G. *J. Biol. Chem.* **1974**, *249*, 4212–4225.

(38) Blundell, T. L.; Pitts, J. E.; Tickle, I. J.; Wood, S. P.; Wu, C.-W. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4175–4179.

(39) Almassy, R. J.; Fontecilla-Camps, J. C.; Suddath, F. L.; Bugg, C. E. *J. Mol. Biol.* **1983**, *170*, 497–527.

(40) Mathews, F. S.; Argos, P.; Levine, M. *Cold Spring Harbor Symp. Quant. Biol.* **1972**, *36*, 387–395.

(41) Bourne, P. E.; Sato, A.; Corfield, P. W. R.; Rosen, L. S.; Birken, S.; Low, B. W. *Eur. J. Biochem.* **1985**, *153*, 521–527.

(42) Bode, W.; Epp, O.; Huber, R.; Laskowski, M., Jr.; Ardelt, W. *Eur. J. Biochem.* **1985**, *147*, 387–395.

(43) Matsuura, Y.; Takano, T.; Dickerson, R. E. *J. Mol. Biol.* **1982**, *156*, 389–395.

(44) Watenpaugh, K. D.; Sieker, L. C.; Jensen, L. H. *J. Mol. Biol.* **1980**, *138*, 615–633.

(45) Guss, J. M.; Harrowell, P. R.; Murata, M.; Norris, V. A.; Freeman, H. C. *J. Mol. Biol.* **1986**, *192*, 361–387.

With its grid search approach, HABAS in principle never misses possible stereospecific assignments, nor will it make erroneous assignments. In practice, however, if one uses "hard limits" and discards all conformations that cause any violation of at least one constraint, experimental errors and the internal mobility of proteins may distort the results. Initial experience showed that with these hard limits one may draw incorrect conclusions if the input data contain small inconsistencies, or if the covalent structure deviates somewhat from the standard ECEPP geometry. Therefore, the program contains a "soft limit" option, with which small violations of the constraints can be tolerated. The size of these allowed violations can be specified by the user and adjusted to the quality of the experimental data.

HABAS goes through a complete search with all values for the dihedral angles χ^1. In some of the empirical approaches to stereospecific assignments using local constraints,[18,24] it was assumed that χ^1 would be within narrow ranges about the staggered rotamers with $\chi^1 = -60$, 60, and 180°. To enable investigations on the impact of this assumption, HABAS includes the option to produce a separate output corresponding to a search of the χ^1 values over a range of ±20° about the three staggered rotamers. Comparison of the results obtained with and without this restriction showed that it does not significantly improve the outcome of the analysis, but that such restricted screening of the χ^1 values may even lead to inconsistencies and erroneous stereospecific assignments (see section IV and Table VI).

As a side product of the grid search for obtaining stereospecific resonance assignments, HABAS yields constraints on the torsion angles ϕ, ψ, and χ^1 which are compatible with all the local NOE distance constraints and the measured spin–spin coupling constants. If the experimental data are not sufficient for obtaining stereospecific assignments, the program still identifies the smallest intervals for the three dihedral angles, which include all conformations that are consistent with the structural constraints for either of the two possible stereospecific assignments. Obviously, these constraints are redundant with the experimental constraints from which they are computed. However, there are indications that the use of these supplementary constraints on the dihedral angles improves convergence for the first stages of structure calculations with the program DISMAN. For example, such dihedral angle constraints were used in section II for the test calculations with BPTI.

The program was applied with a data set derived from the regularized crystal structure of BPTI (see footnotes to Table IV). Table IV illustrates the output format of HABAS and presents the different types of results that one may obtain in applications of the program. For both possible stereospecific assignments of each prochiral group of protons the number of allowed conformations, the allowed values of the dihedral angles ϕ, ψ, and χ^1, and for each dihedral angle the smallest interval containing all these allowed values are given. For the first example listed in the table, Tyr-10, an unambiguous stereospecific assignment was obtained, since $N_r = 0$ (see following section). Pro-13 was stereospecifically assigned, as were all other prolines. Quite generally, since the dihedral angles ϕ and χ^1 are fixed in proline, stereospecific assignments for βCH_2 can be obtained from a minimum of experimental constraints, e.g., when the spin–spin coupling constants $^3J_{\alpha\beta 2}$ and $^3J_{\alpha\beta 3}$ are available, or from the relative intensities of the NOEs from H^α to the two β-protons, using that in Pro the distance H^α–$H^{\beta 2}$ is always longer than the distance H^α–$H^{\beta 3}$. Phe-22 and Phe-33 are examples of residues for which no stereospecific assignments were obtained, since $N_r \neq 0$. For Phe-22 the allowed values for the χ^1 angle constitute two nonoverlapping, well-separated intervals, each of which corresponds to one of the two stereospecific assignments for βCH_2. In this situation, if one can determine by independent, additional procedures that χ^1 falls into only one of the two separated intervals (in Phe-22 either near −120° or near +90°), the stereospecific assignments can be determined from these additional measurements (The additional data would usually be longer range NOE distance constraints). Val-34 demonstrates that stereospecific assignments can also be obtained for the γCH_3 groups of valine. Thr-54 was included to illustrate that for residues with a β-methine proton, the values of the dihedral angles ϕ, ψ, and χ^1 that are consistent with all constraints can be determined with the program HABAS, even though in this case no problem of stereospecific assignments must be solved.

Table IV. HABAS Results for Selected Residues of BPTI from a Test Calculation Using Input Data Derived from the Regularized Crystal Structure 4PTI[a,26,27]

residue[b]	N_i[c]	N_r[c]		allowed dihedral angle values[d] (-180 … -90 … 0 … 90)	HABAS constraints on ϕ, ψ, and χ^1[e]
Tyr 10	131	0	ϕ :	/////	-125°...-75°
			ψ :	///////////	55°...175°
			χ^1 :	// /	165°...-165°
Pro 13	9	0	ψ :	/////// //	135°...-35°
Phe 22	167	65	ϕ :	\XXXXXXXX X	55°...-75°
			ψ :	//XXXX	115°...175°
			χ^1 :	\\ ////	65°...-125°
Phe 33	262	88	ϕ :	XXXXX XXX// XXX//	55°...-75°
			ψ :	//XXXX	115°...175°
			χ^1 :	\\ \\\ ////	15°...-125°
Val 34	68	0	ϕ :	////	-95°...-55°
			ψ :	///////	55°...125°
			χ^1 :	// //	155°...-165°
Thr 54	50		ϕ :	/////	-135°...-85°
			ψ :	///	-75°...-45°
			χ^1 :	/////	-95°...-45°

[a] For the distances $d_{N\beta 2}(i, i)$, $d_{N\beta 3}(i, i)$, $d_{\alpha N}$, d_{NN}, $d_{\beta 2N}$, and $d_{\beta 3N}$, as well as for the pairs of distances $d_{N\alpha}(i, i)$ for glycine, and $d_{NCH_3}(i, i)$ and $d_{CH_3N}(i, i+1)$ for valine (see ref 6 for the notation used), upper limits were extracted from the regularized crystal structure whenever the distances were shorter than 4.0 Å. The values 2.5, 3.0, 3.5, or 4.0 Å were used as upper limits, where the limit <2.5 Å replaced all distances shorter than this value, <3.0 Å all distances in the range 2.50–2.99 Å, etc. Spin–spin coupling constants $^3J_{HN\alpha}$, $^3J_{\alpha\beta 2}$, and $^3J_{\alpha\beta 3}$ were calculated from the crystal structure by using eq 1–3, and Δ^3J was always set to 2.0 Hz. Violations of distance constraints and steric constraints up to 0.1 Å, and of coupling constant constraints up to 0.5 Hz, were tolerated (soft limit option; see text). [b] For underlined residues stereospecific assignments i for βCH_2, or for $\beta CH(CH_3)_2$ in valine, were obtained. [c] The two possible stereospecific assignments are i and r. N_i and N_r are the numbers of conformations found in the grid search that fulfil the constraints for i and r, respectively. [d] /// indicates values that are allowed for the stereospecific assignment i, \\\ is the same for the reversed stereospecific assignment r, and X indicates values that are allowed for both stereospecific assignments. [e] For each dihedral angle the two numbers are the bounds enclosing the smallest interval that contains all allowed values. When a stereospecific assignment was obtained by HABAS, this interval includes only the values that are compatible with this assignment. Otherwise, for example, for Phe-22 and Phe-33, this interval extends over all values that would be compatible with either of the two possible stereospecific assignments.

IV. Extent to Which Stereospecific Assignments in Proteins Can Be Determined

Any approach to the determination of stereospecific assignments for prochiral groups of protons will be limited on several different levels. The most obvious limitation in practice arises if the chemical shifts of the different prochiral protons are accidentally degenerate, so that they cannot be individually observed in the NMR experiments, or if there is high rotational mobility about the single bonds. These situations are not explicitely considered here, but it should be kept in mind that their occurrence will always lower the percentage of accessible stereospecific assignments relative to the results of test calculations, which assume that the prochiral protons can be individually observed and that one deals with immobilized prochiral groups.

With methods such as HABAS, which use exclusively local constraints, the stereospecific assignment can in principle always be obtained, provided that a sufficient amount of exact data is available and that the assumption of standard geometry is strictly valid. In the present practice of NMR studies with proteins, however, it is not easy, for example, to obtain reliable, nontrivial lower bounds on ^1H–^1H distances from NOESY experiments, to

Table V. Results of a Systematic Test of the Program HABAS with the Dipeptide Segment of Figure 4

constraints[a]	Δ^3J,[b] Hz	fraction (%) of the conformations with[c] $N_w = 0$	$N_c \geq 10N_w$	$N_c \geq 2N_w$	$N_w > N_c$
A	1.0	73	74	82	12
	2.0	52	52	70	12
	3.0	32	37	60	16
B	1.0	81	82	87	8
	2.0	57	62	80	8
	3.0	42	50	72	12

[a] In the data set A it was assumed that all ^1H–^1H distances used by HABAS (see section III) are constrained in the intervals between the sum of the two core radii and an upper limit of 2.5, 3.0, 3.5, and 4.0 Å, where for each distance the shortest constraint that includes its actual value is used. The spin–spin coupling constraints $^3J_{HN\alpha}$, $^3J_{\alpha\beta 2}$, and $^3J_{\alpha\beta 3}$ were derived from the molecular conformation by using eq 1–3. In set B the two constraints with the allowed ranges extending from the sum of the two core radii to 3.5 and 4.0 Å, respectively, were replaced by 2.5 Å $< d <$ 3.5 Å and 3.0 Å $< d <$ 4.0 Å, and a new lower bound, $d >$ 4.0 Å, was used for all distances longer than this limit. All other constraints were the same as in A. [b] Δ^3J defines the assumed accuracy of the ^3J measurements, with the allowed range extending from $(^3J - \Delta^3J)$ to $(^3J + \Delta^3J)$. [c] The result of the grid search, which was conducted in steps of 10° over the range from −180 to 180° for all three dihedral angles ϕ, ψ, and χ^1: N_c is the number of sterically allowed conformations for which the *correct* stereospecific assignment was obtained, and N_w is the corresponding number for the reverse, *wrong* stereospecific assignment.

measure upper distance bounds more accurately than within approximately 0.2 Å, or to measure ^3J coupling constants more accurately than to ±1.0 Hz. As a consequence, when using conformational constraints with the precision that is accessible in NMR measurements with proteins, the occurrence of certain local conformations may preclude unambiguous determination of stereospecific assignments. To get an estimate of the percentage of assignments that can be expected with real experimental data, we used two kinds of test calculations. In the first test, the polypeptide segment of Figure 4 was subjected to grid searches using different assumptions about the precision of the experimental data, and the percentage of the conformations enabling unambiguous stereospecific assignments was evaluated. In the second approach, HABAS was used with a database derived from a selection of high-resolution protein structures in single crystals and in solution.

Table V lists the results obtained by systematic screening of all sterically allowed conformations of the peptide segment in Figure 4 with two different sets of constraints. The constraint set A corresponds to data that can presently routinely be obtained from NMR experiments.[6] In the set B, more stringent constraints were introduced (for details see footnotes to Table V). Constraint sets of the types A and B were derived from each of the 13 050 sterically allowed conformations of the peptide in a grid search with 10° intervals for the three dihedral angles ϕ, ψ, and χ^1, and the number of allowed conformations for the two possible stereospecific assignments for βCH_2, N_c and N_w, respectively, were calculated by HABAS. Table V shows that the constraints chosen were not sufficient to establish stereospecific assignments for all of the 13 050 conformations. For example, using the constraint set A with Δ^3J = 2.0 Hz, one gets unambiguous stereospecific assignments for the βCH_2 group (with $N_w = 0$) for just over 50% of the conformations. Somewhat higher percentages of stereospecific assignments were obtained if Δ^3J in A was reduced to 1.0 Hz, or when more stringent NOE distance constraints were assumed with the data set B (Table V). For approximately 10% of the conformations (those with $N_w > N_c$) the program had a tendency to indicate the wrong stereospecific assignment, but there was not a single case where unambiguous evidence for an erroneous assignment (with $N_c = 0$) was obtained.

Table V provides an additional result of interest with respect to the criteria to be used in defining unambiguous stereospecific assignments by HABAS. It shows that similar percentages of conformations with $N_w = 0$ or $N_c \geq 10N_w$ are obtained. Therefore, the determination of stereospecific assignments can be based on the more stringent criterion that $N_w = 0$ without running the risk of substantial reduction in the extent of the assignments achieved.

The results of Table V can be applied to proteins only in an indirect fashion, since they do not account for the preferential population of certain local conformations in globular protein structures. Therefore a second test was performed using as input the constraints derived from a group of 13 protein crystal structures taken from the Protein Data Bank[27] and the solution structure of Tendamistat.[18] The first four columns in Table VI describe the origin and the resolution of these protein structures, and the fifth column lists the number of β-methylene groups contained in them. Hydrogen atoms were attached to the crystal structures, and two simulated NMR input data sets of similar precision to those used in Table V were generated. L is a low-precision input data set as it can be obtained routinely from present NMR experiments. H is a higher precision input data set with tighter constraints for ^1H–^1H distances and more accurate ^3J coupling constant values (L corresponds to data set A of Table V with Δ^3J = 2.0 Hz and H to data set B of Table V with Δ^3J = 1.0 Hz. See also footnotes to Table VI). In Table VI the columns LF and HF list the results obtained with a grid search over the entire χ^1 range and LS and HS those from a limited search near the three staggered rotamers, i.e., from 40 to 80, 160 to −160 and −80 to −40°.

The third row from the bottom in Table VI summarizes the result of these test calculations: Among all 496 non-proline, nonterminal β-methylene groups of the database, HABAS yielded 42% stereospecific assignments with the input LF and 77% with HF. With the assumption that all χ^1 values are near the staggered rotamers, the corresponding results are 49% for LS and 67% for HS. As the first and most important result we thus see that the extent of stereospecific assignments is limited and depends critically on the precision of the input data. The results obtained further imply that it is preferable to use HABAS with an unrestrained grid search of the χ^1 dihedral angle space, since the limitation to values near the staggered rotamers produced only slightly better results for the input L, and for input H, a smaller number of stereospecific assignments were actually obtained with the restricted χ^1 angle range. This reduction of the level of unambiguous stereospecific assignments stems from those residues in the proteins for which the χ^1 values are outside of the ranges of ±20° about the staggered rotamers.[25] Furthermore, for 10% and 15% of the β-methylene groups, respectively, either no sterically allowed conformations or only conformations consistent with the wrong stereospecific assignment were found with the restricted grid search. In contrast, the number of inconsistencies encountered with the complete grid search amounted to less than 0.5%.

The lowest two rows in Table VI show that with the input L there is a significantly higher percentage of stereospecific assignments for residues located in β-strands than for those in helices. With the higher precision input data H this difference disappears. The dependence on the secondary structure presents an explanation for at least part of the sizable variations in the extent of stereospecific assignments obtained for the individual proteins in Table VI. It also adds to earlier observations that β-proteins are generally more readily amenable to structural studies by NMR than α-proteins.[6]

V. Conclusions

There have been indications previously from practical experience with protein structures calculated from experimental NMR data that the use of stereospecific assignments for prochiral groups of protons contributes to improved precision of the structure determination.[18,24] However, since in these projects the input for the structure calculations was at the same time changed in other ways, the influence of the stereospecific assignments could not be properly assessed. The test calculations in section II of this paper now show that the inclusion of stereospecific assignments yields substantial improvements when combined with the usual qualitative conformational constraints corresponding to those that can be obtained from NMR experiments.[6] Thereby the improved

Table VI. Extent of Stereospecific Assignments for β-Methylene Groups Achieved with the Program HABAS by Using Simulated Input Data Derived from 13 Known Protein Structures

protein[a]	code[b]	resolution, Å	R factor, %	βCH_2 groups[c]	stereospecific assignments for βCH_2,[d] %			
					LF	LS	HF	HS
crambin[35]	1CRN	1.5	11	18	28	44 (–)	89	89 (–)
L7/L12 50S ribosomal protein[36]	1CTF	1.7	17	33	45	45 (6)	88	82 (6)
HIPIP[37]	1HIP	2.0	24	43	33	35 (21)	70	44 (33)
avian pancreatic polypeptide[38]	1PPT	1.4		22	41	45 (5)	73	73 (5)
scorpion neurotoxin[39]	1SN3	1.8	16	43	44	56 (5)	74	72 (9)
cytochrome b_5 (oxidized)[40]	2B5C	2.0		58	38	40 (16)	81	55 (28)
erabutoxin b[41]	2EBX	1.4	22	40	50	65 (5)	83	75 (13)
ovomucoid third domain[42]	2OVO	1.5	20	35	43	46 (3)	80	74 (11)
cytochrome c_{551} (red.)[43]	451C	1.6	19	42	38	43 (7)	76	71 (10)
BPTI[26]	4PTI	1.5	16	35	37	54 (–)	69	71 (–)
rubredoxin (oxidized)[44]	5RXN	1.2	12	31	52	52 (10)	74	74 (10)
plastocyanin[45]	6PCY	1.9	15	59	41	58 (15)	70	58 (24)
tendamistat[18]				37	43	51 (19)	86	73 (22)
residues in these proteins	870			496	42	49 (10)	77	67 (15)
residues in helices[e]	225			133	38	41 (10)	89	74 (17)
residues in β-sheets[e]	230			132	57	69 (8)	83	77 (11)

[a] For this analysis all but one of the crystal structures from the Protein Data Bank[27] were used that contain between 30 and 99 amino acid residues and for which data were collected to 2.0 Å or higher resolution (ferredoxin was not included because the refinement method used makes this structure unsuitable for the present study). In addition the solution structure of tendamistat (structure I of ref 18) was included. The root-mean-square deviation with respect to the ECEPP/2 standard geometry[33,34] of the lengths of all covalent bonds between N, C^α, C^β, C′, and O is less than 0.038 Å, and the root-mean-square deviation of all bond angles involving these atoms is less than 4° in all these structures. [b] File indentification code of the Protein Data Bank. [c] All β-methylene groups are considered, except those of all prolines and the chain-terminal residues. In HIPIP Ser-26 and Gln-50 were not used because the atoms O^γ and C^γ, respectively, are not listed in the coordinate file of the Protein Data Bank. [d] Stereospecific assignments were obtained by using two sets of simulated input constraints corresponding to different precision of the NMR measurements, L and H, as described in detail below. Each of the two data sets was used with the assumption that either all values for χ^1 are accessible (indicated by F), or that χ^1 could adopt only values within a limited range about the three staggered rotamers, i.e., 60 ± 20, 180 ± 20, and −60 ± 20° (S). For the low-precision input data (L) it was assumed that all ^1H–^1H distances used by HABAS (see section III) are constrained in the intervals between the sum of the two core radii and an upper limit of 2.5, 3.0, 3.5, and 4.0 Å, where for each distance the shortest constraint that includes its actual value is used. The spin–spin coupling constants $^3J_{HN\alpha}$, $^3J_{\alpha\beta 2}$, and $^3J_{\alpha\beta 3}$ were derived from the molecular conformations by using eq 1–3 with allowed deviations Δ^3J = 2.0 Hz. The higher precision data (H) were the same with the following exceptions: The two constraints with the allowed ranges extending from the sum of the two core radii to 3.5 and 4.0 Å, respectively, were replaced by 2.5 Å < d < 3.5 Å and 3.0 Å < d < 4.0 Å, and a new lower bound, d > 4.0 Å, was used for all distances longer than this limit. The allowed deviation of spin–spin coupling constants Δ^3J was 1.0 Hz. In parentheses the percentages of inconsistencies and erroneous assignments occurring when the investigation is done with the χ^1 angles restricted near the three rotamers (S) is given. [e] The secondary structure identification in the Protein Data Bank were used.[27]

precision of the polypeptide backbone conformation is an outstanding, and not necessarily expected result (Table III). These test calculations thus provided new motivation to work on improved techniques for obtaining stereospecific assignments and to investigate the extent to which stereospecific assignments could a priori be obtained.

The program HABAS screens all available local constraints in a grid search of the dipeptide conformation space and thus performs an unbiased search of stereospecific assignments. It is an improved alternative to the previously applied manual screening of local constraints prior to the three-dimensional structure calculations.[23,24] As such it might in the future be further improved by substituting part or all of the grid search by an analytical analysis. HABAS does not take account of long-range constraints in the three-dimensional structure, which can also lead to stereospecific assignments.[18,20–22] For practical purposes we recommend that HABAS, or in the future perhaps some improved version of this program, is applied to obtain the maximum possible number of stereospecific assignments prior to the start of the structure calculations. These results can then be supplemented by a suitable search procedure for additional assignments based on long-range constraints, which will foreseeably consist of several rounds of screening of the three-dimensional structure at different stages of refinement.[18,20,22]

The results obtained in section IV indicate that the comparison in Figures 1–3 and Table III of corresponding structures calculated with or without stereospecific assignments represents an upper limit to the improvements of protein structure determinations that can in practice be expected from stereospecific resonance assignments. This is a consequence of the fact that these test calculations assumed that 100% of the β-methylene groups were assigned, and that stereospecific assignments were also available for selected additional groups of protons. In contrast, the Tables V and VI predict that with the use of experimental NMR data stereospecific assignments for βCH_2 groups can be expected in the extent of approximately 30–80%, depending primarily on the precision of the NMR measurements. An important message from the present study then is that significant further improvement of the precision of protein structure determination by NMR in solution may be achieved through further progress in quantitative measurements of the local conformational constraints.

Acknowledgment. Financial support by the Schweizerischer Nationalfonds (project no. 3.198.85) and by special funds of the ETH Zürich, as well as the use of the facilities of the Zentrum für Interaktives Rechnen (ZIR) of the ETH Zürich, is gratefully acknowledged. We thank Mrs. E. Huber for the careful processing of the typescript.