^{1}H-NMR Stereospecific Assignments by Conformational Data-base Searches

MICHAEL NILGES,[‡] G. MARIUS CLORE,* and ANGELA M. GRONENBORN*

Laboratory of Chemical Physics, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institute of Health, Bethesda, Maryland 20892

SYNOPSIS

A search procedure is described for making stereospecific assignments at prochiral centers in proteins on the basis of nuclear Overhauser enhancement and coupling constant data derived from nmr experiments. A data base comprising torsion angles, associated ^{1}H-^{1}H coupling constants and interproton distances is searched by a computer algorithm for sets of values that match the experimental data within specified error limits. Two different data bases are used. The first is a *crystallographic* data base derived from 34 well-refined crystal structures; the second is a *systematic* data base derived from conformations of a short peptide fragment with idealized geometry by systematically varying the ϕ, ψ, and χ_1 torsion angles. Both approaches are tested for β-methylene groups with model data obtained from 20 crystal structures. The results for the two methods are similar though not identical, so that a combination of the two methods appears to be useful. With an appropriate choice of error estimates, around 80% of the β-methylene groups could be assigned in the test calculations. In addition, results with experimental nmr data indicate that a similar percentage of stereospecific assignments can be made in practical situations.

INTRODUCTION

For the majority of nmr-derived protein structures determined to date, it is generally the case that the backbone atomic rms distributions about the mean coordinate positions are of the order of 1–2 Å.[1-3] Recent studies on four proteins, BDS-I,[4,5] hirudin,[6] tendamistat,[7] and the C-terminal domain of cellobiohydrolase I,[8] have shown that a significant improvement in the definition of nmr structures can be obtained by making use of stereospecific restraints at prochiral centers, and in particular, at β-methylene groups. In the absence of stereospecific assignments, interproton distance restraints derived from nuclear Overhauser enhancement (NOE) measurements that involve β-methylene protons are usually incorporated into the calculations by means of either a pseudoatom[9] or center averaging.[10] As a result, the information content of the data is significantly reduced as the distance restraints have to be weakened by a comparatively large correction term (of the order of 1 Å in the case of β-methylene groups).

Under suitable circumstances, stereospecific assignments of β-methylene protons can be obtained from a qualitative analysis of NOE and coupling constant data.[11-13] This is usually only applicable to completely unambiguous situations and fails to provide a quantitative measure of the correctness of the assignment. An alternative strategy is to try obtaining the stereospecific assignments automatically during the structure calculation itself, thereby avoiding the problem of making stereospecific assignments prior to computing the structures. This can be accomplished either by using r^{-6} distance averaging, which is heavily weighted toward the smaller distance,[14] or by making the assignments arbitrarily and allowing the two protons to ex-

CCC 0006-3525/90/040813-10 $04.00
Biopolymers, Vol. 29, 813–822 (1990)

*Authors to whom correspondence should be addressed.

[‡]Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511.

change places during the calculation[15] such that the conformation with the lower energy is chosen. To date, however, no systematic study of these approaches and their possible errors exists.

In this paper we describe an approach toward the problem of stereospecific assignments in which the experimental intraresidue and sequential interresidue NOEs and $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants are matched with those calculated for conformations present in a data base. The procedure is carried out for the two alternative assignments, and by comparing the results the correct assignments, as well as allowed ranges for ϕ, ψ, and χ_1, can be determined. Two data bases are employed. The first is a crystallographic data base of high-resolution protein structures inspired by the work of Kraulis and Jones,[16] who showed that the wealth of conformational data present in crystal structures is potentially very useful for nmr structure determinations. The second is a data base comprising the complete ϕ, ψ, and χ_1 conformational space of a short peptide fragment with idealized geometry. A similar approach to the use of the systematic data base has recently been proposed by Güntert et. al.[17]

Model calculations performed with both systematic and crystallographic data bases are described. Model NOE distance data and coupling constants are obtained from 20 crystal structures with a crystallographic R factor less than 20% and a nominal resolution of 2.0 Å or better. Different sets of input data are generated, and the influence of the nature of the input data and its completeness on the number of assignments is assessed. In addition, the results of calculations using experimental nmr data on the C-terminal domain of cellobiohydrolase I are presented.

CALCULATIONAL STRATEGY

The computer program STEREOSEARCH described in this paper attempts to make a stereospecific assignment on the basis of the following experimental data:

1. the coupling constants $^3J_{NH\alpha}$, $^3J_{\alpha\beta2}$, and $^3J_{\alpha\beta3}$, which are related to the dihedral angles via the following Karplus equations[18-20]:

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

$$^3J_{\alpha\beta} = [9.5\cos^2(\theta) - 1.6\cos(\theta) + 1.9] \quad (2)$$

where θ is χ_1 for the $\beta3$ proton, $\chi_1 - 120°$ for $\beta2$;

2. the intraresidue NOE connectivities $d_{\alpha N}(i, i)$, $d_{\alpha\beta2}(i, i)$, $d_{\alpha\beta3}(i, i)$, $d_{\beta2N}(i, i)$, $d_{\beta3N}(i, i)$; and
3. the sequential NOE connectivities $d_{\beta2N}(i, i + 1)$, $d_{\beta3N}(i, i + 1)$, $d_{\alpha N}(i, i + 1)$, $d_{NN}(i, i + 1)$.

To ensure computational and programming efficiency the assignments are made in a two-step procedure. In the first step, which has to be carried out only once, a data base is created that contains sets of values for the three dihedral angles ϕ, ψ, and χ_1, and the associated coupling constants and interproton distances listed above. The entries in this data base are then compared with the data derived from an nmr experiment in the second step. Two different data bases have been used: a crystallographic data base and a systematic one.

The *crystallographic* data base was generated by obtaining the necessary conformational data from the 34 crystal structures listed in Table I. The crystal structures were taken from coordinates deposited in the Brookhaven Protein Data Bank and hydrogen atoms added using the HBUILD feature of the program X-PLOR.[21] Torsion angles and interproton distances were then measured, and the coupling constants calculated using the Karplus equations given above. For each residue with a β-methylene group found in the crystal structures an entry was made into the data base, irrespective of the type of the side chain. At present, it contains data for 3410 nonproline residues.

The *systematic* data base was generated as follows: The three dihedral angles ϕ, ψ, and χ_1 of the central residue of a "tripeptide" with idealized geometry was systematically varied in steps of 10°. The tripeptide is somewhat truncated insofar as the first residue comprises only its carbonyl group (which is necessary to define ϕ), the third only the amide group, which is all that is needed to define ψ and to incorporate the sequential NOEs listed above. Serine was chosen as the second residue as it has the smallest side chain. (Note that an assignment of glycine α-methylene protons is not possible with this method.) Of the 36^3 (46656) possible conformations, a small number with very severe nonbonded contacts are excluded. These mainly comprise those centered around $\phi = 0$, $\psi = 0$. The criteria for exclusion was a value of greater than 3.5 for the quadratic van der Waals repulsion term given by $\sum_{ij}[(vdw_{ij}^2 - d_{ij}^2)/d_{ij}^2]^2$, where vdw_{ij} is the sum of the van der Waals radii of two nonbonded atoms, i and j, and d_{ij} their actual sepa-

Table I Crystal Structures Used for the Crystallographic Data Base

Protein	PDB Code[a]	(Ref.)	Nominal Resolution	Crystal *R* Factor	No. Methylene Groups[b]
Cytochrome C 551	451C	(24)	1.6	18.7	44
Actinidin	2ACT	(25)	1.7	17.1	112
Adenylate kinase	3ADK	(26)	2.1	19.3	121
Penicillopepsin	2APP	(27)	1.8	13.6	181
Azurin	2AZA	(28)	1.8	15.7	78
Phospholipase A2	1BP2	(29)	1.7	17.1	93
Carbonic anhydrase	2CAB	(30)	2.0	19.3	165
Cytochrome C	1CCR	(31)	1.5	19.0	67
Carboxypeptidase	5CPA	(32)	1.54	19.0	190
Cytochrome P450	2CPP	(33)	1.63	19.0	253
Crambin	1CRN	(34)	1.5	11.4	19
L7/L12 ribosomal protein	1CTF	(35)	1.7	17.4	35
Citrate synthase	2CTS	(36)	2.0	16.1	279
Dihydrofolate reductase	3DFR	(37)	1.7	15.2	94
Erythrocruorin	1ECD	(38)	1.4	19.0	76
Elastase	3EST	(39)	1.65	16.9	135
Flavodoxin	4FXN	(40)	1.8	20.0	85
Gluthatione peroxidase	1GP1	(41)	2.0	17.1	116
Hemogiobin (α chain)	2HHB	(42)	1.74	16.0	84
Insulin (α and β chains)	1INS	(43)	1.5	17.9	36
Human lysozyme	1LZ1	(44)	1.5	17.7	84
T4 lysozyme	2LZM	(45)	1.7	19.3	105
Myohemerythrin	2MHR	(46)	1.7	15.8	80
Ovomucoid 3D domain	2OVO	(47)	1.5	19.9	37
Papain	9PAP	(48)	1.65	16.1	122
Plastocyanin	1PCY	(49)	1.6	17.0	60
Pancreatic trypsin inhibitor	5PTI	(50)	1.0	20.0	36
Bence-Jones immunoglobulin	2RHE	(51)	1.6	14.9	64
Ribonuclease A	5RSA	(52)	2.0	15.9	83
Rubredoxin	5RXN	(53)	1.2	11.5	33
Proteinase A	2SGA	(54)	1.5	12.6	83
Thermolysin	3TLN	(55)	1.6	21.3	179
Trypsin	1TPP	(56)	1.4	19.1	134
Ubiquitin	1UBQ	(57)	1.8	17.6	47
Total					3410

[a]All crystal structures are taken from coordinates deposited in the Brookhaven Protein Data Bank.[58]

[b]The count does not include prolines and residues for which χ_1 was not defined due to missing coordinates. Residue Glu 1 in 2RHE is probably a (very disordered) pyrollidone carboxylic acid and is also excluded.

ration. (Note this is only calculated for nonbonded interactions and hence excludes 1–2 and 1–3 interactions; the minimum value of this function is 0 and the maximum value found is 8507.5.) The resulting systematic data base contains 37547 conformations. The principal reason for excluding ~ 20% of the possible conformations is solely to increase the computational efficiency, as test calculations with all 36^3 conformations yield virtually identical results to those with the reduced set.

While the systematic data base has the definite advantage that all the possible conformations in ϕ, ψ, and χ_1 torsion angle space are searched, small deviations from idealized geometry are not taken into account. The crystallographic data base, on the other hand, which includes such deviations, is restricted to the relevant regions of the conformational space; that is to say, conformations that exist in known crystal structures. Thus, the overwhelming majority of the entries in the crystallo-

graphic data base have a negative ϕ torsion angle and most χ_1 torsion angles are close to the staggered rotamer positions. Indeed, only 103 conformations have values of $\phi > 0°$, 37 of which belong to Asn residues. In addition, only 687 of the 3410 χ_1 values deviate by more than 20° from the staggered rotamer positions and the angular rms deviation from these positions is 17.9°. These findings are in complete agreement with those of Richardson[22] and Ponder and Richards.[23]

In the second step the experimental constraints are compared with the entries in one of the data bases. These constraints are entered into the search program in the form of conditional statements; the program then simply checks the conformations in the data base for which these conditions are satisfied. Two kinds of conditions can be used. The first comprises *absolute* distance and/or coupling constant constraints. The distance or coupling constant constraint can be set smaller, larger, or equal to a particular value within a specified error. Thus for any given constraint an upper bound, a lower bound, or both upper and lower bounds can be specified. In the present study, only upper bounds were used for the distances. The second kind consists of *relative* distance constraints. Two distances can be constrained to be different by a specified amount, or to be equal within a specified error. As is shown below, these relative distance constraints are crucial for a high success rate of the method.

Additionally, the search can be restricted to certain ranges of the torsion angles ϕ, ψ, or χ_1. This is useful, for example, in the special case of prolines where χ_1 is constrained to the region around $0 \pm 30°$. It should also be noted that stereo-specific assignment of β-methylene protons of proline is trivial as the Hα proton is always closer to the Hβ3 proton than to the Hβ2.

The stereospecific assignment is obtained as follows. Experimental constraints are entered assuming an arbitrary assignment of the β-protons. The data base is searched for both possible assignments. Whenever all the experimental constraints are matched within the specified errors by an entry in the data base, the corresponding ϕ, ψ, and χ_1 torsion angles are stored separately for each assignment. The correct assignment can be determined unambiguously if the data base contains only conformations for one of the two possible assignments. Alternatively, a looser, more statistical, interpretation can be used in which a stereospecific assignment is made if there are many more conformations for one assignment than for the other. The latter statistical interpretation of the results proved to be quite useful for the crystallographic data base (see below).

Even if no stereospecific assignment is possible, the results of the search usually contains valuable information in the form of the torsion angle ranges that are consistent with the experimental data. These ranges are also obtained for residues with only a single β proton such as threonine, isoleucine, and valine. (Note that the β-proton of Thr and Ile is equivalent to β2 and that of Val to β3 in the IUPAC nomenclature.)

RESULTS AND DISCUSSION

The stereospecific assignment strategy was tested with model NOE distance and coupling constant data derived from the conformation of 1414 non-proline β-methylene groups in 20 crystal structures with crystallographic R factor less than 20% and a nominal resolution of 2.0 Å or better (a subset of Table I).

The coupling constants $^3J_{NH\alpha}$, $^3J_{\alpha\beta2}$, and $^3J_{\alpha\beta3}$ were calculated from the torsion angles ϕ and χ_1 via the Karplus equations given in Eqs. (1) and (2), and an error estimate of ± 2 Hz was used in all calculations.

The interproton distances were classified into *absolute* distance classes corresponding to a factor of approximately three in NOE intensity, allowing for some spin diffusion: namely, *very strong* (< 2.1 Å), *strong* (< 2.6 Å), *medium* (< 3.3 Å), and *weak* (< 4.0 Å). (Without the spin diffusion correction, these class limits would be 2.3, 2.7, 3.3, and 4.0 Å; thus the effect of the correction is to slightly reduce the number of distances in the very strong and strong classes.) The class limits are typical of those employed in nmr structure calculations,[1-10,14] and as the NOE intensity is dependent on r^{-6}, NOEs become virtually undetectable for interproton distances greater than 4 Å. The error estimates chosen for the upper limits were 0.4 Å for the very strong class, and 0.2 Å for the other classes. *No* explicit lower distance bounds were used.

The absolute constraints were supplemented by *relative* distance constraints for certain pairs of distances (e.g., the intraresidue distances from the β2 and β3 protons to the HN proton). If one of the two distances was more than either 1.0 or 0.5 Å larger than the other, it was constrained to be at least 0.6 and 0.1 Å larger, respectively, during the search; otherwise the larger distance was allowed to be up to 0.4 Å smaller during the search. Thus

distance *differences* were classified into three classes (1, 0.5, and 0 Å) with an error estimate of 0.4 Å. Whenever both distances in the crystal structure were larger than 3 Å, no relative distance condition was added. It should be noted that the use of relative distance constraints is not equivalent to the use of explicit lower distance bounds. In the latter case, the lower limit on a particular distance is fixed, while in the former the lower limit is variable and depends upon the relative values of two distances.

Three sets of calculations were performed with different input data. In the first set, absolute distance constraints were used together with relative ones; in the second set, only the absolute distance constraints were used, while in the third set only the relative distance constraints were employed. Relative constraints were used for distance pairs from proton i to the two β-protons of a single β-methylene group; that is to say from H$\alpha(i)$ to H$\beta2(i)$/H$\beta3(i)$, from HN(i) to H$\beta2(i)$/H$\beta3(i)$, and from HN$(i+1)$ to H$\beta2(i)$/H$\beta3(i)$. Coupling constants were used in all calculations. The second and third set of calculations were performed only to investigate the relative contribution of the different kinds of constraints to the success of an assignment, *not* to model an input data set that could realistically be derived from experimental data. It should be noted that the relative distances used can be obtained with a greater precision than absolute distance estimates and they are not affected by motional averaging as the geometric relationship between the two β-methylene protons of a single methylene group is constant. The distance classification that was used for the absolute distance constraints in the form of upper bounds only, corresponds to large differences in NOE intensity and the error estimates are relatively large. Thus, the first set of constraints corresponds to a data set that can be easily obtained experimentally.

As the model data were obtained from 20 structures that were also used for generating the crystallographic data base, a special data base had to be generated for each of these 20 test structures that did *not* contain data derived from the structure itself in order to remove any bias from the results.

Table II summarizes the results obtained from three sets of calculations. Columns C1 and C2 list the assignments obtained with the crystallographic data base using two different evaluation methods, and column S list the results with the systematic data base. In the first evaluation method (used for C1 and S) a stereospecific assignment is only made if the input data is consistent with conformations for only one of the two possibilities. A second interpretation of a more statistical type was used only with the crystallographic data base (column C2). In this case a stereospecific assignment is made when at least ten times as many conformations are found for one possibility than for the other, providing a minimum of 20 conformations consistent with the input data was present in the data base. A prerequisite for using this data base with a statistical interpretation is that the data base should not be weighted toward one or a few crystal structures. For this reason, we used only one of the chains of hemoglobin, for example, and incorporated proteins with the same fold only if they had sufficiently different sequences.

With the first set of input data, which models realistic experimental data, approximately 80% of the stereospecific assignments could be obtained. Leaving out either the absolute or the relative constraints reduces the number of assigned β-methylene groups by 20–30%. Our experience with the program is that the relative constraints are generally very important for a successful stereospecific assignment. This is due to the fact that absolute distance estimates are entered as upper bounds only. Thus, for example, consider the $\chi_1 = -60°$ and 180° staggered rotamers that are characterized by one large (> 10 Hz) and one small (< 5 Hz) $^3J_{\alpha\beta}$ coupling constant. In this case, the two alternative assignments can be distinguished by the two intraresidue connectivities $d_{\beta 3N}$ and $d_{\beta 2N}$, which would be classified as *weak* and *strong*, respectively, for $\chi_1 = -60°$, and *strong* and *strong* for $\chi_1 = 180°$.[10] Consequently, the use of upper bounds can only distinguish the $\chi_1 = 180°$ conformation from the $\chi_1 = -60°$ conformation but *not* vice versa. Relative distance constraints keep the distinction between the two distances without explicitly introducing lower bounds. In the absence of absolute distance estimates, on the other hand, the program cannot distinguish between conformations with values of 2.5 and 3.0 Å, say, from those with values of 3.0 and 3.5 Å. Note, however, that it was possible to assign 67% of the cases with the crystallographic data base using relative distance estimates *alone* (see Table II). This reflects the fact that the crystallographic data base mainly samples the most relevant regions of the conformational space (i.e., values of χ_1 around the staggered rotamer positions and values of ϕ between $-30°$ and $-180°$).

Comparing the results obtained with different methods, the crystallographic data base with a

Table II Assigned Residues in the Model Calculations[a]

PDB Code	No. Assigned β-Methylene Groups								
	Absolute and Relative Distance Estimates			Absolute Distance Estimates Only			Relative Distance Estimates Only		
	C1	C2	S[b]	C1	C2	S	C1	C2	S
451C	30	39	36	13	24	21	21	31	16
2ACT	77	93	90	46	64	61	49	71	42
2APP	131	160	154	97	121	118	69	125	77
1CCR	51	57	56	31	37	35	43	50	38
1CRN	15	18	16	6	7	7	10	15	10
1CTF	32	33	31	16	19	18	22	26	18
3DFR	70	80	75	43	57	53	47	62	44
2HHB	70	75	71	26	34	31	57	68	50
1INS	29	34	32	15	19	19	21	28	18
1LZ1	60	73	65	30	43	40	42	55	36
2MHR	62	71	71	25	38	36	50	57	47
2OVO	26	33	32	18	25	25	11	24	15
1PCY	42	45	44	33	37	38	26	36	22
5PTI	25	29	26	18	20	22	16	24	13
2RHE	46	51	52	37	42	45	26	41	26
5RSA	63	65	64	42	47	50	37	51	37
5RXN	21	29	24	17	22	20	15	22	12
2SGA	57	70	69	46	59	54	31	49	35
1TPP	89	114	103	58	86	77	52	90	58
1UBQ	36	37	38	26	26	27	19	27	20
Total	1032	1206	1149	643	827	797	664	952	634
%	73	85	81	45	58	56	47	67	45

[a] Columns S and C1 list the results with the systematic and crystallographic data bases, respectively, in which a stereospecific assignment is *only* made if the input data is consistent with conformations for only *one* of the two possibilities. Column C2 lists the results with the crystallographic data base in which a statistical interpretation has been used such that a stereospecific assignment is made when at least 10 times as many conformations are found for one possibility than for the other, and a minimum of 20 conformations consistent with the input data is present in the crystallographic data base.

[b] The results given are those with the systematic data base containing 37,547 conformations, approximately 80% of the total 36^3 conformations available in a 10° grid search of ϕ, ψ, χ_1 space. Calculations carried out with a systematic data base containing all 36^3 conformations yield virtually identical results: the total number of stereospecifically assigned β-methylene groups is 1133, which corresponds to an assignment success rate of 80%.

statistical interpretation (column C2) consistently produces slightly larger numbers of assignments, and also seems to be less affected by incompleteness of input data. This interpretation appears to be reasonably safe as not a single incorrect assignment was obtained using this method. It should always be borne in mind, however, that formally an assignment based on such a statistical interpretation cannot be regarded as absolutely firm as the minor form is still a possibility.

A comparison of columns C1 and S shows that the number of stereospecific assignments made with the systematic data base is larger than that with the crystallographic one for two of the three input data sets. This implies that although the systematic data base covers a much larger fraction of the ϕ, ψ, and χ_1 torsion angle space, it does not contain conformations that may be realistic given that they are present in the crystallographic data base. This is due to the fact that the systematic data base does not take into account small deviations from idealized geometry. For the same reason, the ranges of torsion angles consistent with the input data are sometimes slightly *larger* with the crystallographic data base than with the systematic one, contrary to what one might expect given the relative sizes of the two data bases. Thus, the sampling in the relevant energetically favorable regions of conformational space appears to be better in the crystallographic data base.

Table III Results Using Experimental NMR Data for the 18 Nonproline Residues Containing Nondegenerate β-Methylene Protons of the C-Terminal Domain of Cellobiohydrolase I[a]

				Intraresidue NOEs					Interresidue NOEs			Stereospecific Assignment		Torsion angle Ranges[b]			
				$H^{\alpha}(i)$ to		HN(i) to			HN($i+1$) to								
Residue	$^3J_{HN\alpha}$	$^3J_{\alpha\beta a}$	$^3J_{\alpha\beta b}$	$H^{\beta\alpha}(i)$	$H^{\beta b}(i)$	$H^{\beta a}(i)$	$H^{\beta b}(i)$	$H\alpha(i)$	$H^{\alpha}(i)$	$H^{\beta a}(i)$	$H^{\beta b}(i)$	βa	βb	ϕs	ϕc	χ_1s	χ_1c
Gln-2	—	7.5	7.5	—	—	—	—	—	s	w	w	Both		-185 ± 85	-110 ± 65	$-15 \pm 10, 135 \pm 10$	$-15 \pm 10, 135 \pm 10$
Ser-3	7.2	4.5	6.0	s	s	w	s	w	s	w	m	Both		-115 ± 50	-120 ± 45	$50 \pm 5, -115 \pm 10$	$55 \pm 10, -115 \pm 10$
His-4	4.0	10.5	4.5	w	s	s	s	w	—	—	—	$\beta 3$	$\beta 2$	-50 ± 5	60 ± 15	160 ± 5	165 ± 10
Tyr-5	8.0	12.0	4.5	w	s	—	—	s	w	—	—	Both		60 ± 25	65 ± 20	$-50 \pm 15, 170 \pm 15$	$-55 \pm 20, -175 \pm 10$
Gln-7	5.4	3.8	12.5	m	w	s	s	m	s	—	—	$\beta 2$	$\beta 3$	-70 ± 5	-70 ± 15	185 ± 20	180 ± 25
Cys-8	9.9	2.5	3.5	s	s	w	m	w	w	w	m	$\beta 2$	$\beta 3$	-120 ± 35	-120 ± 35	60 ± 15	60 ± 15
Tyr-13	7.2	12.5	4.0	w	m	m	w	w	s	—	w	$\beta 3$	$\beta 2$	-120 ± 55	-115 ± 50	180 ± 15	175 ± 20
Ser-14	9.4	4.5	4.5	m	m	w	w	w	—	—	—	$\beta 2$	$\beta 3$	-120 ± 35	-120 ± 35	60 ± 5	60 ± 15
Cys-19	9.0	3.0	12.0	m	s	w	m	w	s	m	w	$\beta 3$	$\beta 2$	-120 ± 35	-105 ± 30	-85 ± 10	-80 ± 15
Cys-25	5.4	3.5	10.5	s	m	s	s	w	s	w	—	$\beta 2$	$\beta 3$	-70 ± 15	-70 ± 15	175 ± 10	175 ± 10
Gln-26	9.4	4.5	10.5	s	m	w	s	w	s	w	—	$\beta 3$	$\beta 2$	-120 ± 35	-115 ± 30	-40 ± 5	-45 ± 10
Leu-28	—	11.0	3.5	w	s	m	w	—	w	m	m	$\beta 3$	$\beta 2$	-75 ± 70	-80 ± 45	-180 ± 15	175 ± 20
Tyr-31	9.9	4.0	12.0	s	m	w	s	s	w	—	w	$\beta 3$	$\beta 2$	-120 ± 35	-110 ± 25	-60 ± 15	-55 ± 20
Tyr-32	7.2	3.5	12.0	s	w	w	w	w	s	—	—	Both		-120 ± 55	-115 ± 50	$-60 \pm 15, 180 \pm 15$	$-60 \pm 15, 175 \pm 20$
Ser-33	9.9	4.0	11.0	s	m	w	s	w	s	m	w	$\beta 3$	$\beta 2$	-120 ± 35	-120 ± 35	-60 ± 15	-60 ± 25
Gln-34	9.9	10.5	4.5	w	m	s	s	w	s	w	w	$\beta 3$	$\beta 2$	-110 ± 25	-115 ± 30	-165 ± 20	-165 ± 20
Cys-35	6.3	3.0	12.0	s	m	w	s	w	s	w	w	$\beta 3$	$\beta 2$	-115 ± 50	-110 ± 55	-65 ± 20	-65 ± 20
Leu-36	9.9	3.5	11.0	s	w	w	w	w	—	—	—	Both		-120 ± 35	-120 ± 35	$-60 \pm 15, 180 \pm 15$	$-60 \pm 15, 175 \pm 20$

[a] In the data-base search, the coupling constants are assumed to have an accuracy of ± 2 Hz. The NOEs are classified into three distance ranges— < 2.7 Å, < 3.3 Å and < 5.0 Å corresponding to strong (s), medium (m), and weak (w) NOEs.[4-8] No explicit lower bounds are used on the distance restraints. In addition, relative restraints were used for pairs of distances from proton i to the two β-methylene of a β-methylene group. The relative restraints are as follows: the distance corresponding to a strong NOE is assumed to be ≥ 0.5 Å shorter than that corresponding to a weak or absent NOE; the distance corresponding to a strong, medium, or weak NOE is asssumed to be no more than 0.2 Å larger than that corresponding to a medium, weak or absent NOE, respectively. A stereospecific assignment is only made when all the conformations that match with the experimental data have the same stereospecific assignment. The coupling constants were measured from 2D primitive exclusive COSY spectra and the NOEs from 50 ms NOESY spectra recorded in D_2O and H_2O.[8]

[b] ϕs and χ_1s are the torsion angle ranges derived from the systematic data-base search, while ϕc and χ_1c are those from the crystallographic data-base search.

An additional advantage of the crystallographic data base is that the methodology can be extended to longer peptide fragments or beyond the β position of the side chains at almost no additional computational cost. One such attempt was made by including additional sequential interproton distance data [$d_{\alpha N}(i-1,i)$ and $d_{NN}(i-1,i)$]. Only marginal improvements could be achieved, partly because the crystallographic data base appears not to be large enough yet to ensure sufficient sampling for longer peptide fragments.

Table III presents the results of the data-base search on experimental nmr data for the C-terminal domain of cellobiohydrolase I, a small protein of 36 residues containing 18 nonproline β-methylene groups with nondegenerate β-methylene chemical shifts.[8] The coupling constants were specified to an accuracy of ± 2 Hz, and the upper limits of the distance restraints were divided into three classes— $< 2.7 < 3.3$, and < 5.0 Å, corresponding to strong, medium, and weak NOEs. These three distance ranges are somewhat less stringent than those used in the model calculations. In addition, relative distance constraints were used for pairs of distances from proton i to the two β-methylene protons of a single β-methylene group. In particular, we assumed that a strong NOE corresponded to a distance at least 0.5 Å shorter than a weak NOE, and that the distance corresponding to a strong, medium, or weak NOE was no more than 0.2 Å larger than that corresponding to a medium, weak, or absent NOE, respectively. Stereospecific assignments could be obtained for 15 of the 18 β-methylene groups. (Note that *no* statistical interpretation of the results of the data-base search was used and that a stereospecific assignment is *only* made if the input data is consistent with conformations for only one of the two possibilities.) There was no difference in the stereospecific assignments obtained with the systematic and crystallographic databases, and the ranges of the ϕ and χ_1 angles derived from the two searches were generally within ± 1 grid point (i.e., $\pm 10°$) of the systematic data base. It is also instructive to examine those cases where a stereospecific assignment could not be made. In two of the case, Gln-2 and Ser-3, it is clear that the $^3J_{\alpha\beta}$ coupling constant data are suggestive of the presence of multiple χ_1 conformations. In the other three cases, a stereospecific assignment was not possible owing to the absence of discriminating intraresidue NOE information from the NH proton to the two β-methylene protons (i.e., these NOEs were either absent or of the same intensity). In this respect, we note that the most important data required for stereospecific assignments are the $^3J_{\alpha\beta}$ coupling constants and the $C^\alpha H - C^\beta H$ and $NH - C^\beta H$ intraresidue NOEs. Further, in those cases where one of the $^3J_{\alpha\beta}$ coupling constants is ≥ 10 Hz, only the larger $^3J_{\alpha\beta}$ coupling constant is required to make a stereospecific assignment as the other $^3J_{\alpha\beta}$ coupling constant has to be ≤ 4 Hz [cf. Eq. (2)].

CONCLUDING REMARKS

In this paper we have compared two different data-base approaches for obtaining stereospecific assignments at prochiral centers and torsion angle ranges. One data base was generated from conformations found in 34 well-refined crystal structures, the other by systematically varying the torsion angles in a short peptide fragment with ideal geometry. Both methods were tested with model data for 1414 β-methylene groups obtained from 20 crystal structures.

Generally, the assignments obtained with both data bases are very similar, although the two sets of assignments do not always overlap entirely in the test calculations. The reason for this is twofold. While the systematic data base contains many conformations that are not found in crystal structures, the crystallographic data base contains a larger variation of conformations within particular torsion angle ranges. As a result, the allowed ranges for the torsion angles are usually a little larger for the crystallographic data base, indicating that the sampling is better in the "relevant" regions of conformational space. Thus a combination of the two methods seems to be useful. While searching a crystallographic data base avoids artifacts introduced by the idealized geometry (too small torsion angle ranges), the systematic data base contains ranges of the conformational space that are not well represented in the crystallographic data base, and thus ensures that an unusual conformation that has not yet been observed in a crystal is not missed. Additionally, the number of conformations with matching parameters in the crystal structures can be used as a measure of the probability of a particular conformation.

We thank Per J. Kraulis, Ad Bax, Paul C. Driscoll, and Lewis E. Kay for useful discussions. This work was supported by the Intramural AIDS Anti-Viral Targeted Program of the Office of the Director of the National Institutes of Health (G.M.C. and A.M.G.). The computer

program STEREOSEARCH described in this paper is available upon request from the authors.

REFERENCES

1. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
2. Clore, G. M. & Gronenborn, A. M. (1987) *Protein Engin.* **1**, 275.
3. Clore, G. M. & Gronenborn, A. M. (1989) *CRC Crit. Rev. Biochem.*, **24**, 479.
4. Driscoll, P. C., Gronenborn, A. M., Beress, L. & Clore, G. M. (1989) *Biochemistry* **28**, 2188.
5. Driscoll, P. C., Gronenborn, A. M. & Clore, G. M. (1988) *FEBS Lett.* **243**, 223.
6. Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Köhler, S. & Gronenborn, A. M. (1989) *Biochemistry* **28**, 2601.
7. Kline, A. D., Braun, W. & Wüthrich, K. (1988) *J. Mol. Biol.* **204**, 675.
8. Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989) *Biochemistry*, **28**, 7421.
9. Wüthrich, K., Billeter, M. & Braun, W. (1983) *J. Mol. Biol.* **169**, 949.
10. Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1986) *EMBO J.* **5**, 2729.
11. Hyberts, S., Märki, W. & Wagner, G. (1987) *Eur. J. Biochem.* **164**, 625.
12. Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol.* **196**, 611.
13. Arseniev, A., Schultze, P., Wörgötter, E., Braun, W., Vasák, M., Kägi, J. & Wüthrich, K. (1988) *J. Mol. Biol.* **201**, 637.
14. Clore, G. M., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1986) *J. Mol. Biol.* **191**, 523.
15. Weber, P. L., Morrison, R. & Hare, D. R. (1988) *J. Mol. Biol.* **204**, 483.
16. Kraulis, P. J. & Jones, T. A. (1988) *Proteins Struct. Funct. Gen.* **2**, 188.
17. Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989) *J. Am. Chem. Soc.* **111**, 3997.
18. Karplus, M. (1963) *J. Am. Chem. Soc.* **85**, 2870.
19. DeMarco, A., Llinás, M. & Wüthrich, K. (1978) *Biopolymers* **17**, 617.
20. Pardi, A., Billeter, M. & Wüthrich, K. (1984) *J. Mol. Biol.* **180**, 741.
21. Brünger, A. T. & Karplus, M. (1988) *Proteins* **4**, 148.
22. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167.
23. Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193**, 775.
24. Matsuura, Y., Takano, T. & Dickerson, R. E. (1982) *J. Mol. Biol.* **156**, 389.
25. Baker, E. N. & Dodson, E. J. (1980) *Acta Crystallogr. Sect. A* **36**, 559.
26. Dreusicke, D., Karplus, P. A. & Schulz, G. E., to be published.
27. James, M. N. G. & Sielecki, A. R. (1983) *J. Mol. Biol.* **163**, 299.
28. Norris, G. E., Anderson, B. F. & Baker, E. N. (1986) *J. Am. Chem. Soc.* **108**, 2784.
29. Dijkstra, B. W., Kalk, K. H., Hol, W. G. J. & Drenth, J. (1981) *J. Mol. Biol.* **147**, 97.
30. Kannan, K. K., Ramanadham, M. & Jones, T. A. (1984) *Ann. NY Acad. Sci.* **429**, 49.
31. Ochi, H., Hata, Y., Tanaka, N., Kakudo, M., Sakurai, T., Aihara, S. & Morita, Y. (1983) *J. Mol. Biol.* **166**, 407.
32. Rees, D. C., Lewis, M. & Lipscomb, W. N. (1983) *J. Mol. Biol.* **168**, 367.
33. Poulos, T. L., Finzel, B. C. & Howard, A. J. (1987) *J. Mol. Biol.* **195**, 687.
34. Hendrickson, W. A. & Teeter, M. M. (1981) *Nature* **290**, 107.
35. Leijonmarck, M. & Liljas, A. (1987) *J. Mol. Biol.* **195**, 555.
36. Remington, S., Wiegand, G. & Huber, R. (1982) *J. Mol. Biol.* **158**, 111.
37. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982) *J. Biol. Chem.* **257**, 13650.
38. Steigemann, W. & Weber, E. (1979) *J. Mol. Biol.* **127**, 309.
39. Meyer, E., Cole, G., Radahakrishan, R. & Epp, O., to be published.
40. Smith, W. W., Burnett, R. M., Darling, G. D. & Ludwig, M. L. (1977) *J. Mol. Biol.* **156**, 389.
41. Epp, O., Ladenstein, R. & Wendel, A. (1983) *Eur. J. Biochem.* **133**, 51.
42. Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. (1984) *J. Mol. Biol.* **158**, 111.
43. Dodson, E. J., Dodson, G. G., Hodgkin, D. C. & Reynolds, C. D. (1979) *Can. J. Biochem.* **57**, 469.
44. Artymiuk, P. J. & Blake, C. C. F. (1981) *J. Mol. Biol.* **152**, 737.
45. Weaver, L. H. & Matthews, B. W. (1987) *J. Mol. Biol.* **193**, 189.
46. Sheriff, S., Hendrickson, W. A. & Smith, J. L. (1987) *J. Mol. Biol.* **197**, 273.
47. Bode, W., Epp, O., Huber, R., Laskowski, M. Jr. & Ardelt, W. (1985) *Eur. J. Biochem.* **147**, 387.
48. Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A. & Drenth, J. (1984) *J. Mol. Biol.* **179**, 233.
49. Guss, J. M. & Freeman, H. C. (1983) *J. Mol. Biol.* **169**, 521.
50. Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984) *J. Mol. Biol.* **180**, 307.
51. Furey, W., Jr., Wang, B. C., Yoo, C. S. & Sax, M. (1983) *J. Mol. Biol.* **167**, 661.
52. Wlodawer, A., Borkakoti, N., Moss, D. S. & Howlin, B. (1986) *Acta Crystallogr. Sect. B* **42**, 379.
53. Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1980) *J. Mol. Biol.* **138**, 615.
54. James, M. N. G., Sielecki, A. R., Brayer, G. D., Delbaere, L. T. J. & Bauer, C.-A. (1980) *J. Mol. Biol.* **144**, 43.

55. Holmes, M. A. & Matthews, B. W. (1982) *J. Mol. Biol.* **160**, 623.
56. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983) *Acta Crystallogr. Sect. B* **39**, 480.
57. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987) *J. Mol. Biol.* **194**, 531.
58. 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. (1977) *J. Mol. Biol.* **112**, 535.

Received December 5, 1988
Accepted May 3, 1989