Journal of Biomolecular NMR, **15:** 251–264, 1999.
KLUWER/ESCOM

HYPER: A hierarchical algorithm for automatic determination of protein dihedral-angle constraints and stereospecific $C^{\beta}H_2$ resonance assignments from NMR data

Roberto Tejero[a,b], Daniel Monleon[b], Bernardo Celda[a,b], Robert Powers[c] & Gaetano T. Montelione[a,*]

[a]*Center for Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854-5638, U.S.A.*
[b]*Departamento de Química Física, Universidad de Valencia, Dr. Moliner 50, E-46100-Burjassot (Valencia), Spain*
[c]*Department of Structural Biology, Wyeth-Ayerst Research, Pearl River, NY 10965, U.S.A.*

Received 21 September 1999; Accepted 22 September 1999

Key words: automated analysis of NMR data, constraint propagation, grid search, protein NMR

Abstract

A new computer program, HYPER, has been developed for automated analysis of protein dihedral angle values and $C^{\beta}H_2$ stereospecific assignments from NMR data. HYPER uses a hierarchical grid-search algorithm to determine allowed values of ϕ, Ψ, and χ_1 dihedral angles and $C^{\beta}H_2$ stereospecific assignments based on a set of NMR-derived distance and/or scalar-coupling constraints. Dihedral-angle constraints are valuable for restricting conformational space and improving convergence in three-dimensional structure calculations. HYPER computes the set of ϕ, Ψ, and $\chi_1$dihedral angles and $C^{\beta}H_2$ stereospecific assignments that are consistent with up to nine intraresidue and sequential distance bounds, two pairs of relative distance bounds, thirteen homo- and heteronuclear scalar coupling bounds, and two pairs of relative scalar coupling constant bounds. The program is designed to be very flexible, and provides for simple user modification of Karplus equations and standard polypeptide geometries, allowing it to accommodate recent and future improved calibrations of Karplus curves. The C code has been optimized to execute rapidly (0.3–1.5 CPU-sec residue^{-1} using a 5° grid) on Silicon Graphics R8000, R10000 and Intel Pentium CPUs, making it useful for interactive evaluation of inconsistent experimental constraints. The HYPER program has been tested for internal consistency and reliability using both simulated and real protein NMR data sets.

Abbreviations: CPU, central processing unit; CspA, the major cold shock protein from *E. coli*; RNase A, bovine pancreatic ribonuclease A; Z Domain, an engineered immunoglobulin binding domain derived from staphylococcal protein A.

Introduction

Protein structure calculations based on NMR data are significantly improved in quality and speed by the availability of stereospecific assignments at prochiral centers and by the use of dihedral-angle constraints derived from experimental measurements of NOE and scalar coupling data (Güntert et al., 1989, 1991; Nilges et al., 1990). Dihedral-angle constraints limit the conformational space that needs to be searched in structure generation algorithms and can greatly improve the convergence rates (i.e. the number of starting conformations that result in final structures satisfying all of the experimental data). The utility of such dihedral angle constraints is evident from the strong efforts in recent years to develop improved methods for determining homo- and heteronuclear scalar coupling constants in proteins (see for example Bax et al., 1994;

*To whom correspondence should be addressed. E-mail: guy@nmrlab.cabm.rutgers.edu

Biamonti et al., 1994; Vuister et al., 1997). However, such methodology development should be paralleled by improved methods for interpreting scalar coupling constant and NOE data in terms of conformational constraints.

Several strategies and computer programs have already been described for determining $C^{\beta}H_2$ stereospecific assignments and/or associated backbone and side-chain dihedral angle ranges from experimental NMR data (Leach et al., 1977; Zuiderweg et al., 1985; Hyberts et al., 1987; Sobol and Arseniev, 1988; Güntert et al., 1989; Nilges et al., 1990; Clore et al., 1991; Güntert et al., 1991; Polshakov et al., 1995; Gippert et al., 1998) and/or statistical data from the database of protein structures (Gibrat et al., 1991; Kuszewski et al., 1996a,b; Cornilescu et al., 1999). Most of these rely primarily on combined analysis of local distance and scalar coupling constraints. Leach et al. (1977) first described the use of grid searches of dihedral angle space with respect to internuclear constraints for the analysis of proton NOE data. This idea was also inherent in the initial studies of relationships between dihedral angles and experimentally derived intraresidue and sequential distances that form the basis for conventional methods of determining resonance assignments (Billeter et al., 1982; Wüthrich et al., 1984). Sobol and Arseniev (1988) have also explored the use of grid searches in dihedral-angle space to match local polypeptide conformations with proton NOE intensities and cross-relaxation rates assuming limited spin diffusion. Programs in common use today include HABAS (Güntert et al., 1989, 1991) and ANGLESEARCH (Polshakov et al., 1995), which use grid search methods, STEREOSEARCH (Nilges et al., 1990), which searches a database of peptide conformations for structures that satisfy distance and scalar coupling data, and TALOS (Cornilescu et al., 1999), which searches a database of protein sequences, structures, and chemical shifts to identify polypeptide fragments of proteins of known three-dimensional structure and conformations similar to that of the query sequence.

In this paper we describe a new computer program, HYPER, written in the C programming language. Assuming that the NMR data arise from an amino acid residue with a single backbone (ϕ,Ψ) and χ_1 side-chain conformation, HYPER determines ranges of ϕ, Ψ, and χ_1 dihedral angles consistent with absolute and/or relative estimates of up to nine local interproton distances and thirteen homo- and heteronuclear scalar coupling constraints. The program also determines $C^{\beta}H_2$ stereospecific assignments when they are uniquely determined from these data. Rather than using analytical equations (Billeter et al., 1982; Güntert et al., 1989; Polshakov et al., 1995) to define the distances between protons in different conformations, HYPER uses methods from linear algebra developed by Flory and co-workers (Flory, 1969) for statistical evaluation of the configurations of polymer chains to compute the required distances 'on the fly'. HYPER also utilizes a 'hierarchical progressive grid search' algorithm to rapidly define the 'allowed conformational space' from the intersection of conformational spaces determined for the individual constraints. The resulting calculations are very rapid, requiring only a few minutes for analysis of small protein structures on standard laboratory workstations. The output of HYPER includes: (i) allowed ranges of dihedral angles ϕ, Ψ, and χ_1, (ii) families of starting conformations generated randomly within these 'allowed volumes' for use as input to subsequent structure calculations, and (iii) stereospecific assignments uniquely determined from these local data. These constraints and starting conformations enhance the speed and convergence of the subsequent structure generation calculations, resulting in better quality solution structures of proteins.

Computational methods

Calculation of interproton distances

Considering the segment of a polypeptide chain NH-C^{α}H-$C^{\beta}H_2$-C′-NH_{i+1} (shown in Figure 1), one observes that the positions of the five hydrogen atoms (H^N, H^{α}, $H^{\beta 2}$, $H^{\beta 3}$, and H^N_{i+1}) are related by $(5 \times 4)/2 = 10$ interproton distances. One of these (d10 between the $H^{\beta 2}$ and $H^{\beta 3}$ protons) is fixed by the covalent geometry, and is not relevant for this analysis. The other nine interproton distances (d1, d2, ..., d9) are determined by the values of four intervening dihedral angles, ϕ, Ψ, χ_1, and ω. Assuming that the dihedral angle ω is fixed at 0° (cis) or 180° (trans), the remaining three degrees of freedom are determined by nine experimentally accessible distances (d1 through d9). These distances, the corresponding variable names used by the HYPER program, nicknames used for these variables, and the dihedral angles upon which they depend are tabulated in Table 1. Analytical trigonometric relations between some of these distances and the corresponding dihedral angles have been described elsewhere for specific covalent geome-

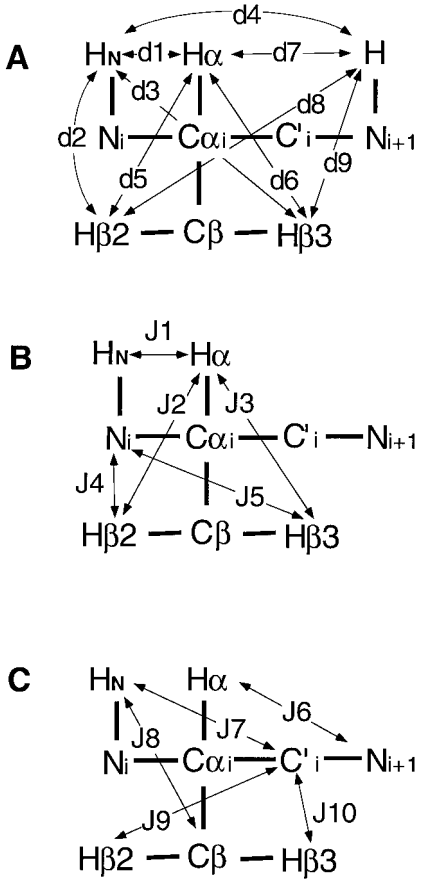

Figure 1. Schematic diagrams showing (A) 9 distances, and (B, C) 10 scalar coupling constants that can be defined as constraint input for the HYPER program. Three additional J constraints, J11, J12 and J13, described in Table 1 are also handled in the current implementation.

tries (Billeter et al., 1982; Wüthrich et al., 1984; Polshakov et al., 1995). The HYPER program computes the subset of required distance versus dihedral angle relationships numerically in the course of the conformational analysis. The conformational spaces of ϕ, Ψ, and χ_1 (for fixed values of ω) are then grid searched in a hierarchical fashion to determine the set of dihedral angles that result in internuclear distances dj′ within the upper and lower bound values of dj defined in the input.

The method for dynamic calculation of interproton distances follows from standard linear algebra (Flory, 1969). Using this basic approach, algorithms within HYPER rapidly compute the dependence of distances d1, . . . , d9 on dihedral angles ϕ, Ψ and χ_1 (and ω if necessary) for defined values of bond lengths and bond angles provided as input. Default values of bond lengths and bond angles are those of the AMBER potential function (Weiner et al., 1986). A related approach is used in the molecular mechanics program ECEPP (Momany et al., 1975; Némethy et al., 1992) and its derivatives DISMAN (Braun and Go, 1985), DIANA (Güntert et al., 1991), DYANA (Güntert et al., 1997) and FANTOM (Schaumann et al., 1990) to construct polypeptide atomic coordinates from sets of dihedral angles. This approach is designed to be readily generalizable in future versions of HYPER to provide analysis of interproton constraints spanning more than three or four dihedral angles.

Analysis of vicinal scalar coupling data

A summary of some vicinal coupling-constant data types that can be used as input to the HYPER program is shown on the segments of polypeptide chain in Figures 1B and 1C. In addition to the 10 vicinal coupling constants indicated in Figure 1, the program is also designed to handle J11 = ^{3}J(^{13}C′-^{13}C$^{\beta}_{i+1}$), J12 = ^{3}J(^{13}C′-^{13}C$^{\gamma}$), and J13 = ^{3}J(^{13}C′ − ^{13}C′$_{i+1}$) coupling constants. Vicinal scalar coupling constants are computed by the HYPER program using Karplus relationships of the generic form $^3J = A\cos^2\theta + B\cos\theta + C$ (Karplus, 1959, 1963), where θ is the torsion angle between the coupled nuclei, corresponding to a standard dihedral angle plus a phase shift that depends on the specific pair of nuclei involved. A, B, and C are constants particular to the specific scalar coupling interaction, which can be provided by the user for each of the 13 coupling constants listed in Tables 1 and 2. Default values of these coupling constants and phase shifts are presented in Table 2.

Evaluation of relative constraints

Another class of useful conformational constraints that can be extracted from experimental NMR data includes relative values of distances or vicinal scalar coupling constants involving the two β-methylene protons. In the current implementation, the four *relative constraints* that are handled are: rd23, the ratio of the d(H^N-H$^{\beta2}$) / d(H^N-H$^{\beta3}$) distances; rd56, the ratio of the d(H$^{\alpha}$-H$^{\beta2}$) / d(H$^{\alpha}$-H$^{\beta3}$) distances; rJ23, the ratio of the ^{3}J(H$^{\alpha}$-H$^{\beta2}$) / ^{3}J(H$^{\alpha}$-H$^{\beta3}$) coupling constants; and rJ45, the ratio of the ^{3}J(^{15}N-H$^{\beta2}$) / ^{3}J(^{15}N-H$^{\beta3}$) coupling constants.

The relative J constraints rJ23 = ^{3}J(H$^{\alpha}$-H$^{\beta2}$) / ^{3}J(H$^{\alpha}$-H$^{\beta3}$) are interpreted by the HYPER program like any other Karplus-type relationship, as indicated in Figure 2A. The interpretation of relative constraints rJ45 = ^{3}J(^{15}N-H$^{\beta2}$) / ^{3}J(^{15}N-H$^{\beta3}$), however, is com-

Table 1. Definitions of constraints used by HYPER

Constraint	Variable name in HYPER[a]	Nickname in HYPER[a]	Dihedral angle(s)[b]	Order of evaluation[c]	Classification in HYPER[c]
$d(H^N\text{-}H^\alpha)$	d1	dhn-ha	ϕ	16	strong
$d(H^N\text{-}H^{\beta 2})$	d2	dhn-hb2	ϕ,χ_1	20	weak
$d(H^N\text{-}H^{\beta 3})$	d3	dhn-hb3	ϕ,χ_1	21	weak
$d(H^N\text{-}H^N_{i+1})$	d4	dhn-hn	ϕ,χ_1	22	weak
$d(H^\alpha\text{-}H^{\beta 2})$	d5	dha-hb2	χ_1	17	strong
$d(H^\alpha\text{-}H^{\beta 3})$	d6	dha-hb3	χ_1	18	strong
$d(H^\alpha\text{-}H^N_{i+1})$	d7	dha-hn	Ψ	19	strong
$d(H^N_{i+1}\text{-}H^{\beta 2})$	d8	dhb2-hn	Ψ,χ_1	23	weak
$d(H^N_{i+1}\text{-}H^{\beta 3})$	d9	dhb3-hn	Ψ,χ_1	24	weak
$^3J(H^N\text{-}H^\alpha)$	J1	jhn-ha	ϕ	1	strong
$^3J(H^\alpha\text{-}H^{\beta 2})$	J2	jha-hb2	χ_1	2	strong
$^3J(H^\alpha\text{-}H^{\beta 3})$	J3	jha-hb3	χ_1	3	strong
$^3J(N\text{-}H^{\beta 2})$	J4	jn-hb2	χ_1	4	strong
$^3J(N\text{-}H^{\beta 3})$	J5	jn-hb3	χ_1	5	strong
$^3J(H^\alpha\text{-}N_{i+1})$	J6	jha-n	Ψ	6	strong
$^3J(H^N\text{-}C')$	J7	jhn-c	ϕ	7	strong
$^3J(H^N\text{-}C^\beta)$	J8	jhn-cb	ϕ	8	strong
$^3J(C'\text{-}H^{\beta 2})$	J9	jc-hb2	χ_1	9	strong
$^3J(C'\text{-}H^{\beta 3})$	J10	jc-hb3	χ_1	10	strong
$^3J(C'\text{-}C^\beta_{i+1})$	J11	jc-cb	ϕ	11	strong
$^3J(C'\text{-}C^{\gamma 1})$	J12	jc-cg1	χ_1	12	strong
$^3J(C'\text{-}C'_{i+1})$	J13	jc-c	ϕ	13	strong
Relative distance d2/d3	rd23	rdhn-hb	χ_1	25	strong
Relative distance d5/d6	rd56	rdha-hb	ϕ,χ_1	26	weak
Relative intensity J2/J3	rJ23	rJha-hb	χ_1	14	strong
Relative intensity J4/J5	rJ45	rJhn-hb	χ_1	15	strong

[a]Constraints are defined in input files using either the variable name or the alternate variable nickname.
[b]Dihedral angle(s) restrained by the corresponding constraint.
[c]Constraints are evaluated hierarchically in the HYPER program, with strong constraints that provide the most restriction of conformational space evaluated first and the weaker constraints that do not significantly restrain conformational space evaluated last.

plicated by two features that are illustrated in Figure 2B. While $^3J(^{15}N\text{-}H^\beta)$ values can range from about -5 to $+1$ (DeMarco et al., 1978), the relative values of these coupling constants often come from relative intensity measurements in coherence transfer experiments (Archer et al., 1991) that do not provide information about the relative signs of coupling constants. For this reason, HYPER assumes that only the *absolute value* of the relative constraint, |rJ45|, is derived from the experimental measurements. An additional complication results from the singularities observed for values of $^3J(^{15}N\text{-}H^{\beta 3}) = 0$ Hz (Figure 2B). Accordingly, HYPER interprets rJ45 relative constraints according to the following rules:

If upper-bound value $|rJ45| > 2.5$ units,
any grid value conformation with computed $|rJ45|'$ less than the specified lower-bound value of |rJ45| is excluded from the allowed space.

If upper-bound value $|rJ45| \leq 2.5$ units,
any grid value conformation with computed $|rJ45|'$ less than the specified lower-bound value of |rJ45| or greater than the specified upper-bound value of |rJ45| is excluded from the allowed space.

These rules simply exclude use of the upper-bound |rJ45| relative constraints for conformations with χ_1 values near the singularities indicated in Figure 2B. In addition, the code includes a third rule to handle the pathological case:

Table 2. Default coefficients and phase shifts used in Karplus equations of the HYPER program

Coupling constant	Variable in HYPER	Dihedral angle	Phase shift	A	B	C	Reference
$^3J(H^N\text{-}H^\alpha)$	J1	ϕ	-60	7.09	-1.42	1.55	Hu and Bax, 1997
$^3J(H^\alpha\text{-}H^{\beta 2})$	J2	χ_1	-120	9.40	-1.40	1.60	Kopple, 1973
$^3J(H^\alpha\text{-}H^{\beta 3})$	J3	χ_1	0	9.40	-1.40	1.60	Kopple, 1973
$^3J(N\text{-}H^{\beta 2})$	J4	χ_1	$+120$	-4.40	1.20	0.10	DeMarco et al., 1978
$^3J(N\text{-}H^{\beta 3})$	J5	χ_1	-120	-4.40	1.20	0.10	DeMarco et al., 1978
$^3J(H^\alpha\text{-}N_{i+1})$	J6	Ψ	$+60$	-0.88	-0.61	-0.27	Wang and Bax, 1995
$^3J(H^N\text{-}C')$	J7	ϕ	$+180$	4.29	-1.01	0.00	Hu and Bax, 1997
$^3J(H^N\text{-}C^\beta)$	J8	ϕ	$+60$	3.06	-0.74	0.13	Hu and Bax, 1997
$^3J(C'\text{-}H^{\beta 2})$	J9	χ_1	0	7.20	-2.04	0.60	Fischman et al., 1980
$^3J(C'\text{-}H^{\beta 3})$	J10	χ_1	$+120$	7.20	-2.04	0.60	Fischman et al., 1980
$^3J(C'\text{-}C^\beta)$	J11	ϕ	-120	1.28	1.02	0.30	Hu and Bax, 1998
$^3J(C'\text{-}C^{\gamma 1})$	J12	χ_1	$+60$	2.02	0.67	0.59	Henning et al., 1996
$^3J(C'\text{-}C'_{i+1})$	J13	ϕ	0	1.36	-0.93	0.60	Hu and Bax, 1997

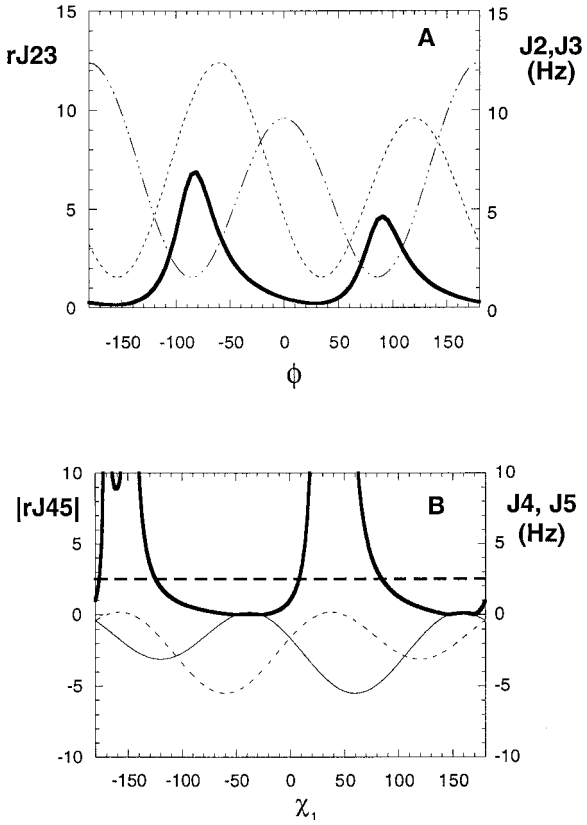

Figure 2. Plots of selected vicinal coupling constants and relative coupling constants. (A) ----, J2 = $^3J(H^\alpha\text{-}H^{\beta 2})$; –···–, J3 = $^3J(H^\alpha\text{-}H^{\beta 3})$; ——, rJ23 = J2/J3. (B) ----, J4 = $^3J(N\text{-}H^{\beta 2})$; ——, J5 = $^3J(H^N\text{-}H^{\beta 3})$; ——, |rJ45| = |J4/J5|. A dashed line is indicated for |rJ45| = 2.5 Hz, for reasons explained in the text.

If lower-bound value |rJ45| > 8.5 units,
ignore |rJ45| constraint.

Hierarchical grid search

HYPER uses a progressive hierarchical conformational grid search to identify the ϕ, Ψ, χ_1 space and stereospecific $C^\beta H_2$ assignments that are consistent with upper- and lower-bound distance and vicinal coupling constant constraints. The search is hierarchical in that constraints that can greatly limit this conformational space are searched first, and the resulting allowed space is used to guide subsequent searches of this solution space to identify subspaces consistent with weaker constraints. The constraints listed in Table 1 are defined as 'strong' or 'weak' based on how much typical experimental data obtained for these constraints restrict the conformational space of the corresponding dihedral angles. These qualitative classifications are based on our experience in HYPER calculations with simulated data, and heuristic knowledge of the precision of the experimental measurements. In the HYPER algorithm, the 'strong' constraints (J1, . . . , J13 coupling-constant constraints and distances d1, d5, d6 and d7) that are available are evaluated first to restrict the allowed ϕ, Ψ, and χ_1 values to narrow regions. In searching the allowed space of any given constraint, the program only evaluates those ϕ, Ψ, χ_1 values which are consistent with the previous constraints that have been evaluated. In this way, initial constraint evaluations guide the search of subsequent constraint solution spaces.

This limits the number of vector matrix computations that are required in the search, particularly since the more computationally intensive analyses of distance constraints which depend on two dihedral angles (i.e. d2, d3, d4, d8, and d9 in Table 1) are all classified as 'weak' and are evaluated only after the conformational space is narrowed by the analysis of all of the available scalar-coupling constraints and the 'strong' distance constraints.

In our experience, for some data sets the solution space consistent with a large number of distance and scalar coupling constraints is quite large, and can be identified quickly using a coarse sampling of ϕ, Ψ, χ_1 space, while for other data sets the solution space is very narrow, requiring fine grid gradations in order to identify the allowed values. For this reason, the hierarchical search, in which the first constraints analyzed restrict the space to be searched in the evaluation of subsequent (generally weaker) constraints, is also progressive in so far as the spacing of the grid search can be interactively reduced until either a solution space consistent with all of the constraints is identified, or the search reaches a limit of resolution defined by the user. The resulting solution space need not be continuous.

Determination of stereospecific $C^{\beta}H_2$ resonance assignments

Several of the distance and scalar-coupling constraints listed in Tables 1 and 2 involve stereochemically distinct $H^{\beta 2}$ and $H^{\beta 3}$ atoms. HYPER assumes that these stereospecific assignments are not initially available. By convention, the input constraint file is created *assuming* that the upfield $H^{\beta u}$ and downfield $H^{\beta d}$ methylene resonances correspond to the $H^{\beta 2}$ and $H^{\beta 3}$ atoms, respectively. The progressive hierarchical grid search is executed with this assumption, and either a solution space (**S1**) is identified or the set of constraints are determined to be mutually inconsistent. Next, the program automatically flips these assignments for all of the corresponding d and J constraints (i.e. in the second cycle $H^{\beta u}$ is assigned to $H^{\beta 3}$ and $H^{\beta d}$ is assigned to $H^{\beta 2}$) and the progressive grid search is repeated. Again, the result of the grid search is either an allowed solution space (**S2**) or a determination that the set of constraints are inconsistent. If consistent solution spaces (**S1** and **S2**) are obtained for both sets of stereospecific assignments, the program determines that the data are insufficient to determine stereospecific $C^{\beta}H_2$ assignments, and identifies the solution space as the union of solution spaces **S1** and **S2**. If, however, one of the sets of stereospecific assignments results in an inconsistent solution, while the other results in an allowed solution space, the allowed space and corresponding set of stereospecific assignments are identified to be correct. Finally, if both sets of stereospecific assignments produce no consistent solution space, the program determines that there is no single conformation that is consistent with the set of constraints. Such inconsistencies can result from inaccurate estimates of distance or vicinal coupling constraints, or from the effects of ensemble averaging due to multiple backbone and/or side-chain conformations that are in rapid equilibrium on these NMR timescales. The user is then in a position to evaluate which subsets of the constraints for that particular residue are self-consistent, which may be helpful in identifying ensemble-averaging effects and/or incorrect constraints. For glycine residues, the kinds of NMR data handled in the current implementation do not generally provide stereospecific assignments of the prochiral $H^{\alpha 2}$ and $H^{\alpha 3}$ resonances. The current implementation of the program also does not provide stereospecific assignments of prochiral isopropyl methyl groups of valine and leucine residues.

Input

An example of a HYPER input file is shown in Appendix A. The input to HYPER includes: (i) a knowledge base of Karplus equation coefficients used to define the Karplus relationships for the various coupling constants, (ii) a second knowledge base containing values of standard peptide bond lengths and bond angles, and (iii) a listing of experimental constraints for each of one or more residues that includes the mean, upper-bound, and lower-bound values of distance constraints (d1, . . . , d9), scalar coupling constants (J1, . . . , J13), relative distance constraints (rd23 and rd56), and relative scalar coupling constant constraints (rJ23 and rJ45). The average value of each constraint is not used by the program. HYPER calculations assume accurate (though not necessarily precise) measurements of distance and J constraints (or of relative distance or J constraints). Distance constraints are derived from NOESY and/or ROESY spectra using complete relaxation matrix analysis (Borgias and James, 1990) wherever possible, and scalar coupling constants are obtained from the emerging family of NMR experiments that provide accurate estimates of homo- and heteronuclear coupling constants (for reviews see Bax et al., 1994; Biamonti et al., 1994; Vuister et al., 1997).

Relative constraints rd23 and rd56 can be obtained from NOESY- or ROESY-type 2D or heteronuclear 3D spectra recorded with short mixing times (Clore et al., 1991), relative constraints rJ23 can be obtained from TOCSY-type 2D or 3D spectra recorded with short mixing times (Clore et al., 1991; van Duynhoven et al., 1992; Fogolari et al., 1993; Constantine et al., 1994) and relative values rJ45 can be obtained from HNHB-type experiments (Archer et al., 1991). Relative constraints for a given ϕ, Ψ, χ_1 grid point are computed in different ways, depending on the relationship between the corresponding distance or J values indicated in the input file. HYPER provides for two types of relative constraint relationships: (i) ratios of distances or scalar coupling constants, defined in the input as rd23r, rd56r, rJ23r, and rJ45r, and (ii) magnitudes of differences between pairs of distances or constraints, defined in the input as rd23l, rd56l, rJ23l, and rJ45l for differences in distance or J coupling values less (l) than a given value or as rd23g, rd56g, rJ23g, and rJ45g for differences in distance or J coupling values greater (g) than a given value. Examples of relative constraints defined as ranges of ratios and as relative magnitudes are given in Appendix A, together with comments explaining how the corresponding input is interpreted by the HYPER program.

Output

The output of HYPER for each residue analyzed includes (i) ranges of the values of ϕ, Ψ, and χ_1 that are consistent with the complete set of input upper- and lower-bound constraints, (ii) stereospecific assignments of $C^{\beta}H_2$ resonances if these were unambiguously defined from the input data, and (iii) sets of ϕ, Ψ, χ_1 values *randomly selected within the allowed solution space*, to be used as defining starting conformations for structure generation calculations. By default, this third kind of output is suppressed and is only generated when required.

The solution spaces derived by HYPER will not in general correspond to a single continuous, rectangular region of ϕ, Ψ, χ_1 space. However, as ranges of dihedral angles are reported, the output corresponds to the smallest rectangular area that includes all of the allowed space (together with some regions of unallowed space). Although it would be possible to output a more precise description of the solution space, we do not currently have methods to utilize such data in structure generation programs except in the form of upper and lower bounds on dihedral angles. The HYPER program does keep track of the true (generally non-rectangular) solution space, and it is this space that is used to generate the randomly selected sets of ϕ, Ψ, χ_1 values for use as starting structures in structure-generation calculations.

Software distribution

Copies of the program suitable for execution on Silicon Graphics IRIX and Pentium LINUX platforms, together with representative input and output files, are available upon request from the authors (http://www-nmr.cabm.rutgers.edu/).

Results

Simulated constraints: Ideal secondary structures

In order to verify the self-consistency of HYPER calculations, a series of 10 regular polypeptide chains, each containing three serine residues (Ser_3), were constructed using the molecular modeling program INSIGHT-II (Molecular Simulations, Inc.). In each chain, the ϕ and Ψ values of each residue were set to identical values corresponding to one of the common secondary structures outlined in Table 3. Side-chain χ_1 values were set to -60, 180, or $+60$ degrees in each of these regular structures. Distances d1, ... , d9 were then measured for the central residue in each of these polypeptide chains using interactive computer graphics. Upper and lower bound constraints were then generated from these values by adding and subtracting, respectively, 0.2 Å to each of these values. Scalar coupling constants J1, ... J13 were also computed for each structure, and upper and lower bound scalar-coupling constraints were generated by adding and subtracting 0.2 Hz, respectively, from each value. Relative constraint ratios rd23r, rd56r, rJ23r, and rJ45r were also computed for these conformers; the corresponding uncertainties were ± 0.2 units for relative distances and ± 0.4 units for relative J values. A series of HYPER calculations were then carried out for these 10 residues, each subject to a set of 26 pairs of upper- and lower-bound constraints derived from one of the 10 secondary structural elements listed in Table 3. While such precise and complete constraint sets cannot be obtained with real data, this data set provides a rigorous test of the internal consistency of the HYPER code. The grid search was carried out at a fixed resolution of 2°. The resulting HYPER analyses provided allowed ranges (shown in Table 3) for each of these 10 constraint sets for the correct stereospecific

assignment of $C^{\beta}H_2$ constraints, and no consistent solutions when using the incorrect stereospecific assignments of $C^{\beta}H_2$ constraints. In each case, the target dihedral angle values from which the constraints were derived are included in the HYPER solution space. These benchmark calculations demonstrate that the set of distances and scalar coupling values computed within the HYPER program are fully self-consistent to high resolution for the 10 specific points in ϕ, Ψ, χ_1 space indicated for the secondary structural elements in Table 3. In addition, the successful characterization of these narrow solution spaces demonstrates that the internal distance and scalar coupling calculations of the HYPER program are also consistent with geometries and distances of molecules generated with the Insight-II molecular graphics program.

Simulated constraints: Derived from a high resolution crystal structure

A second test of the self-consistency of the internal representations of constraints in the HYPER program involved back-calculating dihedral angle ranges from distance and scalar coupling data derived from the coordinates of a protein X-ray crystal structure. For this work we selected the 1.26-Å resolution joint X-ray and neutron diffraction crystallographic structure of bovine pancreatic ribonuclease A (RNase A), a 124-residue protein containing a mixture of α-helical, β-strand, and non-regular secondary structure (Wlodawer et al., 1988). Interproton distances d1, . . . , d9 were computed for 124 residues, and 0.5 Å was added and subtracted to create upper and lower bound distance constraints for HYPER input. Scalar coupling constants corresponding to constraints J1, . . . , J13 were then computed from the corresponding dihedral angles derived from the crystal structure, and 1.0 Hz was added and subtracted from each of these to create the corresponding upper and lower bound scalar coupling constraints. Relative constraint ratios rd23r, rd56r, rJ23r, and rJ45r were also computed, assuming uncertainties of ± 0.5 units and ± 2 units, respectively. Residues 93 and 114 are cis prolines (Wlodawer et al., 1988) and the input file was modified accordingly by designating library 'cis pro' {omega 0.0} for the corresponding residue.

A series of HYPER calculations were then carried out for 124 residues of RNase A, each subject to the distance and scalar-coupling constant constraints derived from the crystal structure. A total of 2733 pairs of upper and lower bound constraints (22 constraints per residue) were evaluated. The grid search was carried out at a resolution of 5°. The resulting HYPER analyses provided allowed ranges of ϕ, Ψ, and χ_1 dihedral angles together with stereospecific assignments for all 86 pairs of $C^{\beta}H_2$ constraints. However, even with the complete set of simulated data the HYPER program could not distinguish stereospecific assignments for Gly $C^{\alpha}H_2$ resonances. When the correct stereospecific constraints were used, no inconsistent solutions were obtained for this complete and precise ideal data set derived from a single conformation of RNase A. HYPER solution spaces for a representative segment of the protein are shown in Figure 3. For every residue, the dihedral angle values of the crystal structure from which the constraints were derived are included in the HYPER solution space. These benchmark calculations demonstrate, again, that the set of distances and scalar coupling values computed within the HYPER program are fully self-consistent to high resolution for the 124 specific points in ϕ, Ψ, χ_1 space corresponding to this single conformer of RNase A.

Real experimental NMR constraints: The Z domain of protein A

Having demonstrated the internal consistency of distance and scalar-coupling constraints within the HYPER program, we next tested the program with real experimental data obtained for the 58-residue Z domain of staphylococcal protein A (Tashiro et al., 1997). A total of 246 d1, . . . , d9, J1, J4, J5 and rJ23 upper-/lower-bound constraints are available for 56 residues (4.4 constraint pairs per residue) of Z domain. Distance constraints were derived from full relaxation matrix calculations, and were assumed to have a precision of ± 0.5 Å. $^3J(H^N\text{-}H^{\alpha})$ scalar coupling constants were determined from 2D HSQC-J (Neri et al., 1990; Billeter et al., 1992) and HMQC-J (Kuboniwa et al., 1994) spectra, and were estimated to have precisions of ± 1 Hz. Vicinal $^3J(^{15}N\text{-}H^{\beta})$ coupling constants were derived with an estimated uncertainty of ± 1 Hz from 2D homonuclear NOESY and 3D PFG [^{15}N] HSQC-NOESY spectra using ^{15}N-enriched Z domain without ^{15}N decoupling during the two proton evolution periods (Montelione et al., 1989). Relative values of vicinal $^3J(H^{\alpha}\text{-}H^{\beta})$ coupling constants were estimated with an uncertainty of ± 2 units from ratios of $H^{\alpha}\text{-}H^{\beta}$ cross-peak intensities in a 2D TOCSY spectrum (Constantine et al., 1994) recorded with a mixing time of 15 ms.

A series of HYPER calculations were then carried out for the 56 residues of Z domain for which appropriate constraints were available, using a 5° grid

Table 3. Consistency checks for HYPER calculations on 10 ideal standard secondary structure polypeptide conformations of (L-Ser)$_3$

Conformation	ϕ	Ψ	χ_1	Ranges[a] ϕ		Ranges[a] Ψ		Ranges[a] χ_1	
				min	max	min	max	min	max
Right-handed alpha-helix	−57	−47	−60	−58	−56	−52	−44	−60	−60
Antiparallel beta-sheet	−139	135	180	−140	−138	128	140	180	180
Parallel beta-sheet	−119	113	60	−122	−116	106	120	60	60
Type I beta-bend position 2	−60	−30	−60	−60	−60	−32	−28	−60	−60
Type I beta-bend position 3	−90	0	180	−90	−90	0	0	180	180
Type II beta-bend position 2	−60	120	60	−60	−60	110	128	60	60
Type II beta-bend position 3	80	0	−60	78	80	0	0	−60	−60
Collagen triple helix	−78	−149	60	−78	−78	−154	−146	60	60
Right-handed 3_{10} helix	−49	−26	−60	−50	−48	−26	−26	−60	−60
Left-handed alpha-helix	57	47	180	56	58	36	54	180	180

[a] All calculations were carried out using a grid search resolution of 2°.

search. Z domain contains no glycine residues. The resulting solution spaces are indicated in Figure 4. HYPER provided constraints on 55, 46, and 23 dihedral angles ϕ, Ψ, and χ_1, respectively. In addition, stereospecific assignments were obtained for 5 out of a total of 46 non-degenerate β-methylene proton pairs in the corresponding input file. For one residue, His18, inconsistent solutions were obtained for both choices of stereospecific assignments, suggesting dynamic averaging as discussed below. Constraints and stereospecific assignments derived from a similar earlier version of HYPER have been used to generate high resolution three-dimensional structures of Z domain (Tashiro et al., 1997).

Real experimental NMR constraints: The major cold shock protein (CspA) from E. coli

We have also applied HYPER for analysis of real experimental data available for the 70-residue major cold shock protein of *E. coli* (CspA) (Feng et al., 1998). A total of 342 d1, . . . , d9, J1, J4, J5 and rJ23 upper/lower-bound constraints are available for 53 residues (6.4 constraints per residue) of CspA. As for Z domain, these distance constraints were assumed to have a precision of ± 0.5 Å, and $^3J(H^N\text{-}H^\alpha)$ scalar coupling constants were assumed to have precisions of ± 1 Hz. Relative values of vicinal $^3J(H^\alpha\text{-}H^\beta)$ coupling constants were estimated with an uncertainty of ± 2 units from ratios of $H^\alpha\text{-}H^\beta$ cross peaks as described above for Z domain (Feng et al., 1998). HYPER provided constraints on 43, 39, and 40 dihedral angles ϕ, Ψ, and χ_1, respectively and stereospecific assignments for 9 out of a total of 40 $C^\beta H_2$ methylene groups in the corresponding input file. The resulting constraints and stereospecific assignments have been used to refine the three-dimensional structure of CspA (Newkirk et al., 1994; Feng et al., 1998).

Execution speeds

The C code for HYPER has been compiled for SG R4000, R8000, and R10000 processors operating under the IRIX operating system, and on Intel Pentium processors operating under the LINUX operating system. Some CPU processing times (s/residue) for these simulated and real constraint data sets are summarized in Table 4. HYPER has also been compiled under SunOS 5.X on SUN workstations and for the DOS/Windows9X operating systems (performance data not shown). In general, the program runs very fast on all of the platforms tested. These execution times ($\sim$1 min for complete analysis of a 100-residue protein; $\sim$0.5 s/residue) are sufficiently fast to provide real-time interactive analysis of a constraint data set. This is important, as it allows the user the flexibility to evaluate inconsistent sets of constraints in order to identify subsets which are mutually consistent in an interactive fashion.

Discussion

Reliability

These tests of the HYPER program on both simulated and real data sets demonstrate that the equations describing various absolute and relative distance and J constraints within the HYPER program are self-consistent and reliable to high precision. In particular,

Table 4. Execution times for HYPER analysis[a] of simulated and real NMR constraint data sets on different platforms

CPU	Clock speed (MHz)	RNase A (s-residue^{-1})	Z domain (s-residue^{-1})	CspA (s-residue^{-1})
SG R4000[b]	100 (IP22)	1.58	4.54	6.22
SG R8000[c]	75 (IP21)	0.89	2.39	2.86
SG R10000[d]	194 (IP25)	0.28	0.78	0.86
Intel Pentium II[e]	266	0.41	0.95	1.36

[a]All calculations carried out with a grid search resolution of 5°.
[b]Calculations carried out on a single processor of an SG Indigo workstation operating under UNIX.
[c]Calculations carried out on a single processor of an SG Challenge server operating under UNIX.
[d]Calculations carried out on a single processor of an SG Onyx workstation operating under UNIX.
[e]Calculations carried out on a single processor of an IBM personal computer operating under LINUX.

the program has been tested with simulated constraints derived from ideal secondary structure conformations assuming that all 26 constraint types were available. HYPER also provided small, but finite, solution spaces for all of the 124 residues of RNase A from complete and highly precise constraint data sets derived from the crystal structure. For the ideal secondary structure conformers and RNase residues, HYPER also provided correct stereospecific assignments for all $C^{\beta}H_2$ methylene proton pairs. Self-consistent solution spaces and some $C^{\beta}H_2$ stereospecific assignments were also obtained using real data for Z domain and CspA. The resulting dihedral angle constraints output by the HYPER program and stereospecific assignments have been used together with the complete set of NOE and other NMR data to derive high-quality three-dimensional structures of Z domain (Tashiro et al., 1997) and CspA (Feng et al., 1998). Thus, the current implementation of the HYPER program provides reliable estimates of the solution spaces, dihedral angle constraints, and stereospecific assignments of extensive sets of data derived from real NMR experiments.

In addition to providing local conformational constraints, the HYPER program also provides an automatic and unbiased approach for determining stereospecific $C^{\beta}H_2$ resonance assignments. Using very complete and relatively precise simulated data, the program consistently determines correct stereospecific H^{β} resonance assignments for *every* residue containing a $C^{\beta}H_2$ group. However, with the real data sets described in this paper, only 14 out of 86 $C^{\beta}H_2$ pairs ($\sim$15%) were stereospecifically assigned. This performance rate is somewhat lower than that obtained using more heuristic methods, which generally assume standard rotamer states for side-chain χ_1 conformations. While such information could be included in the HYPER analysis, our current view, based on the results with simulated data, is that higher rates of stereospecific $C^{\beta}H_2$ assignment frequencies can best be achieved by obtaining more complete and precise sets of distance and scalar-coupling constant estimates than those described here for Z domain and CspA.

Flexibility

The HYPER program has been designed to provide significant flexibility in handling constraints derived from NOE and scalar coupling constant measurements. In particular, as Karplus relationships for the various dihedral angles handled by the program continue to be refined, the program provides for modifications of the Karplus coefficients by simple editing of the input files. Moreover, Karplus coefficients and polypeptide covalent geometry can be defined individually for each residue in the sequence, allowing specification of residue-specific Karplus curves or covalent geometry. HYPER also provides methods for handling relative distance and J constraints involving atoms X (= HN, N, C′ or H^{α}) and pairs of methylene $C^{\beta}H_2$ protons, including estimates of the ratios of these constraints or constraints on sizes of differences between X-$H^{\beta 2}$ and X-$H^{\beta 3}$ distances or J values. In ad-

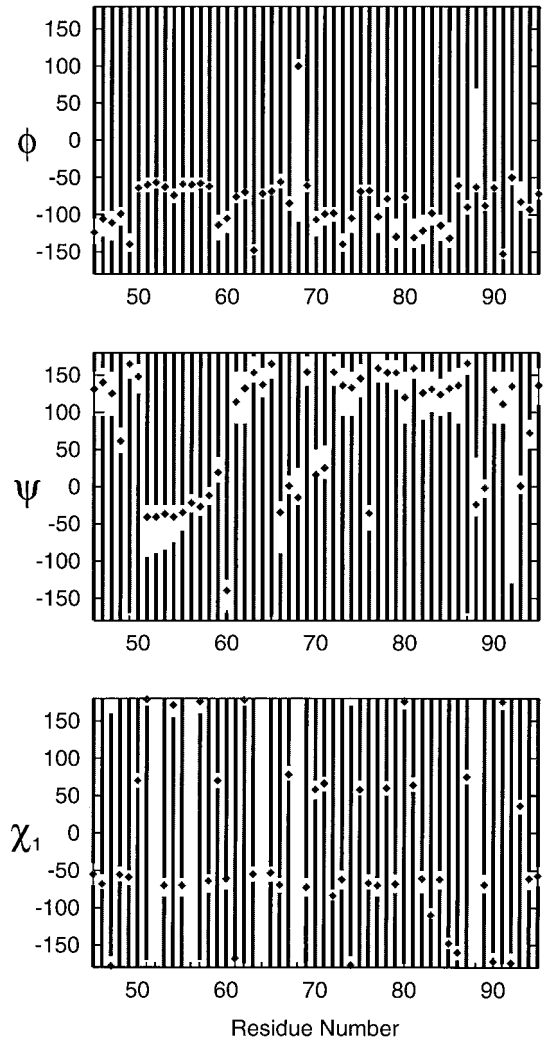

Figure 3. Histogram representation of the output of HYPER for simulated data generated from the X-ray crystal structure of bovine pancreatic ribonuclease A (Wlodawer et al., 1988; pdb entry 7RSA). Black bars represent the excluded dihedral space indicated in the output from HYPER. Diamonds represent the values of ϕ, Ψ, χ_1 calculated from the X-ray crystal structure coordinates.

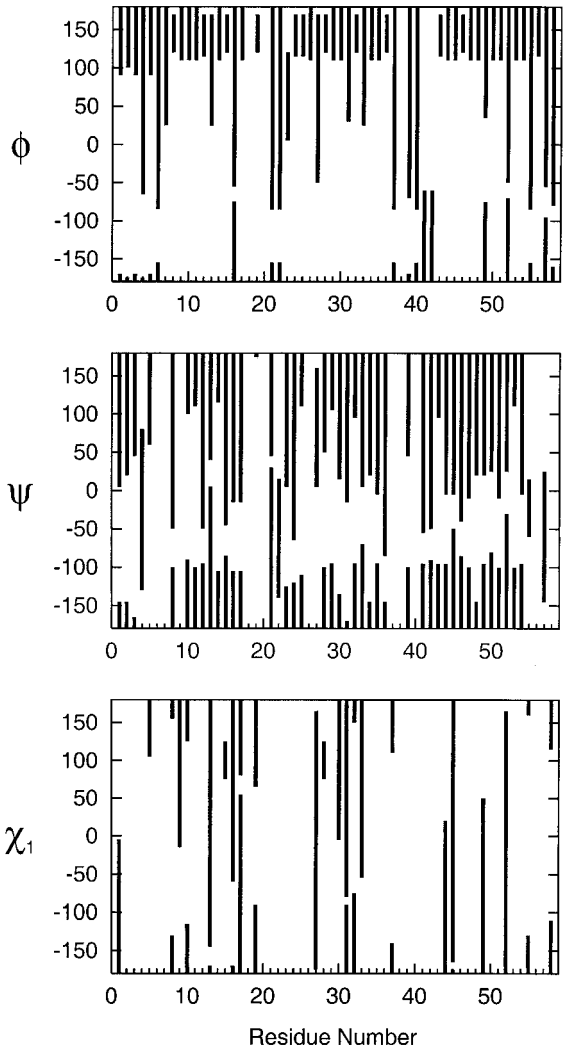

Figure 4. Histogram representation of the output of HYPER for real experimental data generated for the Z domain of staphylococcal protein A (Tashiro et al., 1997). Black bars represent the excluded dihedral space indicated in the output from HYPER.

dition, the program also handles constraints involving diastereotopic Gly $C^{\alpha}H_2$ atoms, Pro residues which do not have intraresidue constraints involving amide HN atoms, and cis peptide bond conformations.

An additional feature of the program is the ability to define prior knowledge about ranges of allowed values of dihedral angles as input to the grid search analysis. Information that can be used to restrict the local conformational space searched by HYPER includes chemical shift data (Celda et al., 1995; Kuszewski et al., 1995, 1996b; Pearson et al., 1995; Cornilescu et al., 1999), data on one-bond heteronuclear coupling constants (Mierke et al., 1992; Vuister et al., 1992, 1993) and three-bond isotope-shift effects, information about relative bond vector orientations from comparisons of zero- and multiple-quantum relaxation rates (Reif et al., 1997), information about local energetics defined by steric overlap or conformational database information (Gibrat et al., 1991; Kuszewski et al., 1996b), or other kinds of prior knowledge derived from more sophisticated conformational energy calculations.

Dynamic averaging

One of the principal assumptions of the current version of HYPER is that the constraints to be analyzed are derived from a single static molecular conformation. This is a standard assumption of most structure generation programs available today. However, even folded globular protein structures are in fact an ensemble of conformations, often interconverting on a timescale that is fast compared to the chemical shift and coupling constant timescales. The static target assumption is often an acceptable approximation for backbone conformations and for buried side-chain conformations, while surface side-chain atoms and backbone atoms of surface loops generally adopt multiple conformations that are ensemble-averaged in NMR measurements. Several algorithms and programs have been developed to extract the individual conformers contributing to one or more ensemble-averaged NMR parameters (see for example Torda et al., 1990; Schmitz et al., 1993; Pearlman, 1994; Constantine et al., 1995; Bonvin and Brünger, 1996). In such approaches it is valuable to have criteria for identifying local regions of the protein structure exhibiting ensemble averaging on the basis of inconsistencies in the constraints when attempts are initially made to fit them to a single conformer.

The HYPER program detects ensemble-averaging effects when they result in an inconsistent set of local distance and vicinal coupling constant constraints. In this case, the program yields no consistent solution space for both tests of $C^{\beta}H_2$ (or $C^{\alpha}H_2$) stereospecific assignments. In these circumstances, the program can be used to evaluate for a particular residue all permutations of subsets of the corresponding constraints. Since HYPER execution times are quite fast ($\sim$ 1 s/residue), all possible constraint subsets can be evaluated for self-consistency quite quickly. In this way, the program can identify subsets of self-consistent constraints and/or identify consistently violated constraints that may have been measured inaccurately. With this data in hand, the user is in a position to decide if specific constraints should be reanalyzed from the experimental data, or if the residue in question should be modeled assuming ensemble averaging between multiple conformations. Since such dynamics are often restricted to side-chain flexibility, it is sometimes appropriate in these situations to rerun the HYPER analysis for this specific residue assuming a rigid backbone conformation using only the subset of self-consistent constraints between backbone atoms.

Future directions

One interesting extension of the HYPER process involves hierarchical conformational searching of higher dimensional spaces involving multiple residues; e.g. constraints imposed by identified $i \rightarrow i + 3$ hydrogen bonds of α-helices, in which the search of higher dimensional spaces is guided and limited by the allowed spaces characterized by searches of lower-dimensional spaces. Efforts are currently in progress to explore the applicability of hierarchical grid searches in the analysis of higher dimensional spaces restrained by medium- and long-range distance constraints.

Acknowledgements

We thank G. Elkins, W. Feng, D. Jin, B. Lyons and M. Tashiro for useful comments and suggestions, and R. Klein for editorial comments on the manuscript. This work was supported by grants from The University of Valencia for partial support of R.T., The Generalitat Valenciana for partial support of D.M., The National Institutes of Health (GM-47014), the National Science Foundation (MCB-9407569), a Research Excellence Award from the New Jersey Commission on Science and Technology, Valencian Grant GV96-D-CN-09-140-1 and a Camille Dreyfus Teacher-Scholar Award.

References

Archer, S.J., Ikura, M., Torchia, P.A. and Bax, A. (1991) *J. Magn. Reson.*, **95**, 636–641.

Billeter, M., Braun, W. and Wüthrich, K. (1982) *J. Mol. Biol.*, **155**, 321–346.

Billeter, M., Neri, D., Otting, G., Qian, Y.Q. and Wüthrich, K. (1992) *J. Biomol. NMR*, **2**, 257–273.

Borgias, B.A. and James, T.L. (1990) *J. Magn. Reson.*, **87**, 475–487.

Braun, W. and Go, N. (1985) *J. Mol. Biol.*, **186**, 611–626.

Celda, B., Biamonti, C., Arnau, M.J., Tejero, R. and Montelione, G.T. (1995) *J. Biomol. NMR*, **5**, 161–172.

Clore, G.M., Bax, A. and Gronenborn, A.M. (1991) *J. Biomol. NMR*, **1**, 13–22.

Constantine, K.L., Friedrichs, M.S. and Müller, L. (1994) *J. Magn. Reson.*, **B104**, 62–68.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

DeMarco, A., Llinas, M. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2727–2742.

Feng, W., Tejero, R., Zimmerman, D.E., Inouye, M. and Montelione, G.T. (1998) *Biochemistry*, **37**, 10881–10896.

Fischman, A.J., Live, D.H., Wyssbrod, H.R., Agosta, W.E. and Cowburn, D. (1980) *J. Am. Chem. Soc.*, **102**, 2533–2539.

Flory, P.J. (1969) *Statistical Mechanics of Chain Molecules*, John Wiley, New York, NY.

Fogolari, F., Esposito, G., Cauci, S. and Viglino, P. (1993) *J. Magn. Reson.*, **A102**, 49–57.
Gibrat, J.-F., Robson, B. and Garnier, J. (1991) *Biochemistry*, **30**, 1578–1586.
Gippert, G.P., Wright, P.E. and Case, D.A. (1998) *J. Biomol. NMR*, **11**, 241–263.
Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.
Güntert, P., Qian, P.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. (1991) *J. Mol. Biol.*, **217**, 531–540.
Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.
Henning, M., Maurer, M., Schwalbe, H., Diener, A. and Griesinger, C. (1996) *Poster No. MP131*, XVIIth ICMRBS Meeting, Keystone, CO.
Hu, J.-S. and Bax, A. (1997) *J. Am. Chem. Soc.*, **119**, 6360–6368.
Hu, J.-S. and Bax, A. (1998) *J. Biomol. NMR*, **11**, 199–203.
Hyberts, S.G., Märki, W. and Wagner, G. (1987) *Eur. J. Biochem.*, **164**, 625–635.
Karplus, M. (1959) *J. Chem. Phys.*, **30**, 11–15.
Karplus, M. (1963) *J. Am. Chem. Soc.*, **85**, 2870–2871.
Kopple, K. (1973) *Biopolymers*, **12**, 672.
Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) *J. Biomol. NMR*, **4**, 871–878.
Kuszewski, J., Qin, J., Gronenborn, A. and Clore, G.M. (1995) *J. Magn. Reson.*, **B106**, 92–96.
Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1996a) *Protein Sci.*, **5**, 1067–1080.
Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1996b) *J. Magn. Reson.*, **B112**, 79–81.
Leach, S.J., Némethy, G. and Scheraga, H.A. (1977) *Biochem. Biophys. Res. Commun.*, **75**, 207–215.
Mierke, D.F., Grdadolnik, S.C. and Kessler, H. (1992) *J. Am. Chem. Soc.*, **114**, 8283–8284.
Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A. (1975) *J. Phys. Chem.*, **79**, 2361–2380.
Montelione, G.T., Winkler, M.E., Rauenbuehler, P. and Wagner, G. (1989) *J. Magn. Reson.*, **82**, 198–204.
Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S. and Scheraga, H.A. (1992) *J. Phys. Chem.*, **96**, 6472–6484.
Neri, D., Otting, G. and Wüthrich, K. (1990) *J. Am. Chem. Soc.*, **112**, 3663–3665.
Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, D., Inouye, M. and Montelione, G.T. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 5114–5118.
Nilges, M., Clore, G.M. and Gronenborn, A.M. (1990) *Biopolymers*, **29**, 813–822.
Pearson, J.G., Wang, J., Markley, J.L., Le, H. and Oldfield, E. (1995) *J. Am. Chem. Soc.*, **117**, 8823–8829.
Polshakov, V.I., Frenkiel, T.A., Birdsall, B., Soteriou, A. and Feeney, J. (1995) *J. Magn. Reson.*, **B108**, 31–43.
Reif, B., Henning, M. and Griesinger, C. (1997) *Science*, **276**, 1230–1233.
Schaumann, T., Braun, W. and Wüthrich, K. (1990) *Biopolymers*, **29**, 679–694.
Sobol, A.G. and Arseniev, A.S. (1988) *Bioorganicheskaia Khimiia*, **14**, 997–1013.
Tashiro, M., Tejero, R., Zimmerman, D.E., Celda, B., Nilsson, B. and Montelione, G.T. (1997) *J. Mol. Biol.*, **272**, 573–590.
van Duynhoven, J.P.M., Goudriaan, J., Hilberts, C.W. and Wijmenga, S.S. (1992) *J. Am. Chem. Soc.*, **114**, 10055–10056.
Vuister, G.W., Delaglio, F. and Bax, A. (1992) *J. Am. Chem. Soc.*, **114**, 9674–9675.
Vuister, G.W., Delaglio, F. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 67–80.
Vuister, G.W., Tessari, M., Karimi-Nejad, Y. and Whitehead, B. (1997) In *Biological Magnetic Resonance 17* (Eds., Krishna, N.R. and Berliner, L.J.), Kluwer Academic, New York, NY.
Wang, A.C. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 1810–1813.
Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. (1986) *J. Comput. Chem.*, **7**, 230–252.
Wlodawer, A., Svensson, L.A., Sjolin, L. and Gilliland, G. (1988) *Biochemistry*, **27**, 2705–2717.
Wüthrich, K., Billeter, M. and Braun, W. (1984) *J. Mol. Biol.*, **180**, 715–740.
Zuiderweg, E.R.P., Boelens, R. and Kaptein, R. (1985) *Biopolymers*, **24**, 601–611.

Appendix A

Example of a sample input file for HYPER. The parser routines inside HYPER recognize **djlist**, **library**, **karplus**, and **Residue** as keywords in addition to the names of the distances (d1 – d9 and their nicknames), scalar coupling constants (J1 – J13 and their nicknames), and relative values of some d and J constraints as discussed in the text. Text following two dashes, --, is considered a comment and consequently ignored by the parser. All data related to **Residue**, **karplus**, **library** and **djlist** keywords must be included between brackets as shown in the sample of input.

```
-- Example of HYPER input.
-- Comments are designated by TWO dashes (--)

-- KARPLUS equation Knowledge Base.
-- A comment is required between single quotes.

karplus ' J coupling eqs data '
{
    --                                  Phase    Phase
    --  Eq.#    C      B      A       shift    shift        Explanation of eqn.
    -- -----  -----  -----  -----     -----    ----- ----------------------------
           1   1.55  -1.42   7.09    -60.0      0.0        -- (HN -HA Phi-60)
           2   1.60  -1.40   9.40   -120.0      0.0        -- (HA -HB* Chi-120,Chi)
           3   0.10   1.20  -4.40    120.0   -120.0        -- (HN -HB* Chi+-120)
           4  -0.27  -0.61  -0.88     60.0      0.0        -- (HA -HNi+1 Psi+60)
           5   0.00  -1.01   4.29    180.0      0.0        -- (HN -C' Phi+180)
           6   0.13  -0.74   3.06     60.0      0.0        -- (HN -CB Phi+60)
           7   0.60  -2.04   7.20      0.0    120.0        -- (HN -HB* Chi,Chi+120)
           8   0.30   1.02   1.28   -120.0      0.0        -- (C'-CB Phi-120)
           9   0.60  -0.93   1.36      0.0      0.0        -- (C'-C'i+1 Phi)
          10   0.59   0.67   2.02     60.0      0.0        -- (C'-CG1 Chi+60)
}

-- LIBRARY with data for polypeptide geometry.
-- A comment is required between single quotes.

library ' Standard peptide geometry '
{
        --      BOND/ANGLE      VALUES  EXPLANATION AND DEAFULT VALUE
        --      ----------     -------- --------------------------------
                ln*ca            1.449   --Bond N-CA      :(def: 1.459)
                ln*c             1.335   --Bond N-C       :(def: 1.335)
                ln*hn            1.010   --Bond N-H       :(def: 1.010)
                lca*cb           1.526   --Bond CT-CT     :(def: 1.536)
                lca*c            1.522   --Bond CT-C      :(def: 1.522)
                lc*h             1.090   --Bond CT-HC     :(def: 1.110)
                h*n*ct           118.4   --Angle H-N-CT   :(def: 118.4)
                n*ct*hc          109.5   --Angle N-CT-HC  :(def: 109.5)
                n*ct*ct          109.7   --Angle N-CT-CT  :(def: 109.7)
                n*ct*c           110.1   --Angle N-CT-C   :(def: 110.1)
                ct*ct*hc         109.5   --Angle CT-CT-HC:(def: 109.5)
                ct*c*n           116.6   --Angle CT-C-N   :(def: 116.6)
                c*n*h            119.8   --Angle C-N-H    :(def: 119.8)
}

djlist ' A comment is required between these single quotes. '
{
  Residue_ALA_02              --  -rt-alpha-helix_-57_-47_-60
        {
        --   library ' cis peptide ' { omega  0.0 }   -- to handle CIS peptide

                --       CODE      VALUE    UPPER    LOWER      NICKNAMES
                --      -------    ------   ------   ------     ----------
                        dhn-ha      2.86     3.06     2.66          -- d1
                        dhn-hb2     2.51     2.71     2.31          -- d2
                        dhn-hb3     3.61     3.81     3.41          -- d3
                        dhn-hn      2.80     3.00     2.60          -- d4
                        dha-hb2     3.09     3.29     2.89          -- d5
                        dha-hb3     2.50     2.70     2.30          -- d6
                        dha-hn      3.59     3.79     3.39          -- d7
                        dhb2-hn     2.59     2.79     2.39          -- d8
                        dhb3-hn     3.67     3.87     3.47          -- d9
                        jhn-ha      2.84     3.04     2.64          -- j1
                        jha-hb2    12.40    12.60    12.20          -- j2
                        jha-hb3     3.25     3.45     3.05          -- j3
                        jn-hb2     -0.40    -0.20    -0.60          -- j4
                        jn-hb3     -5.50    -5.30    -5.70          -- j5
                        jha-n      -6.09    -5.89    -6.29          -- j6
                        jhn-c       3.26     3.46     3.06          -- j7
                        jhn-cb      2.99     3.19     2.79          -- j8
                        jc-hb2      1.38     1.58     1.18          -- j9
                        jc-hb3      1.38     1.58     1.18          -- j10
                        jc-cb       2.52     2.72     2.32          -- j11
                        jc-cg       3.28     3.68     2.88          -- j12
                        jc-c        0.57     0.97     0.17          -- j13
                        rdhn-hbr    0.70     0.90     0.50          -- rd23r
                        rdha-hbr    1.24     1.44     1.04          -- rd56r
                        rjha-hbr    3.82     4.22     3.42          -- rj23r
                        rjhn-hbr    0.07     0.47    -0.33          -- rj45r
        }
}
```