Protein Engineering, Design & Selection vol. 18 no. 12 pp. 597–605, 2005
Published online October 24, 2005 doi:10.1093/protein/gzi068

Methods of translating NMR proton distances into their corresponding heavy atom distances for protein structure prediction with limited experimental data

Oscar Hur[1,2] and Kevin Karplus[3]

[1]PO Box 2548, Sunnyvale, CA 94087-0548 and [3]Department of Biomolecular Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064. USA

[2]To whom correspondence should be addressed.
E-mail: ohur1688@alumni.ucsd.edu

This paper proposes a strategy to translate experimental ^{1}H NMR proton distance restraints into their corresponding heavy atom distance restraints for the purpose of protein structure prediction. The relationships between interproton distances and the corresponding heavy atom distances are determined by studying well-resolved X-ray protein structures. The data from the interproton distances of amide protons, α-protons, β-protons and side chain methyl protons are plotted against the corresponding heavy atoms in scatter plots and then fitted with linear equations for lower bounds, upper bounds and optimal fits. We also transform the scatter plots into two-dimensional heat maps and three-dimensional histograms, which identify the regions where data points concentrate. The common interproton distances between amide protons, α-protons, β-protons in α-helices, anti-parallel β-sheets and parallel β-sheets are also tabulated. We have found several patterns emerging from the distance relationships between heavy atom pairs and their corresponding proton pairs. All our upper bound, lower bound and optimal fit results for translating the interproton distance into their corresponding heavy atom distances are tabulated.

Keywords: heavy atom distance/NMR/proton distance

Introduction

From the relationship between proton–proton distances and the corresponding heavy atom–heavy atom (carbon and nitrogen) distances from high-resolution (1.7–1.0 Å) X-ray crystal protein structures, we developed a strategy to translate proton distance restraint data from ^{1}H NMR spectroscopy into corresponding heavy atom distance restraint data for the purpose of combining protein structural prediction with limited experimental data. With the dramatic advances in ^{1}H NMR spectroscopy, much protein structure data determined by ^{1}H NMR has been produced (Bax and Grzesick, 1993). One important parameter obtained experimentally by ^{1}H NMR applications is the distance restraints between two protons (Nilges *et al*., 1988; Clore and Gronenborn, 1998). X-PLOR/CNS and DYANA/DIANA ^{1}H NMR proton distance restraint data contain the information on upper and/or lower distance restraint limits and/or optimal proton distance restraints (Güntert *et al*., 1991, 1997; Brünger, 1992; Schwieters *et al*., 2003). However, many protein structural prediction computational programs, such as Undertaker, can only represent heavy atoms, e.g. carbon, oxygen, nitrogen and sulfur (Karplus *et al*., 2003). We report here the rules for translating ^{1}H NMR proton distance restraints into heavy atom distance restraints. The methods are for the use of limited amounts of NMR data combined with structure prediction for obtaining models more quickly than NMR data alone and more accurately than prediction alone (Bowers *et al*., 2000; Zheng and Doniach, 2002; Hung and Samudrala, 2003; Meiler and Baker, 2003; Li *et al*., 2004).

The translation is done by plotting graphs to calibrate the relationships between the heavy atom distance restraints and the corresponding proton distance restraints from a dataset of 100 high-resolution crystal structure (1.7–1.0 Å) PDB files (Table I) (Word *et al*., 1999a) with all hydrogen atoms added and optimized by Reduce (Word *et al*., 1999b). Two-dimensional scatter graphs and heat maps and three-dimensional histograms of the heavy atom distances versus their corresponding interproton distances are plotted. We also report here that based on the curve fitting, linear equations can yield fairly good approximations for translating proton distance restraints into the corresponding heavy atom distance restraints.

We only calculate the proton distance restraints with at least two rotatable covalent bonds (not counting the rigid peptide C–N bond) between the two corresponding heavy atoms. The short-range proton distance restraints, such as, HN(i)–HA(i), HN(i)–HB(i), HA(i)–HB(i), HA(i)–HA(i + 1) and HA(i)–HN(i + 1), which do not satisfy the above requirement, provide no useful distance restraint information on their corresponding heavy atoms. Torsional angles may be constrained by some of these short-range interproton distances, even though the

Table I. Protein structures PDB files with protons processed, added and optimized by Reduce that were used in the study of distance relationships between the heavy atom pairs and the corresponding proton pairs[a]

1aac	1ads	1aky	1amm	1arb	1aru	1benA[b]	1benB[b]
1bkf	1bpi	1cem	1cka	1cnr	1cnv	1cpcB	1cse
1ctj	1cus	1dad	1dif	1edmB	1etm	1ezm	1fnc
1fus	1fxd	1hfc	1ifc	1igd	1iro	1isuA	1jbc
1kap	1knb	1lam	1lit	1lkk	1lucB	1mctI	1mla
1mrj	1nfp	1nip	1not	1osa	1phb	1php	1plc
1poa	1ptf	1ptx	1ra9	1rcf	1rgeA	1rie	1rpo
1rro	1sgpI	1smd	1snc	1sriA	1tca	1ttaA	1whi
1xic	1xsoA	1xyzA	2ayh	2bopA	2cba	2ccyA	2cpl
2ctc	2end	2er7	2erl	2hft	2ihl	2mcm	2mhr
2msbA	2olb	2phy	2rhe	2rn2	2trxA	2wrp	3b5c
3chy	3ebx	3grs	3lzm	3pte	3sdhA	4fgf	4ptp
5p21	7rsa	8abp	256bA	451c			

[a]A reference dataset of 100 protein structures was chosen on the basis of resolution (1.7 Å or better), crystallographic *R*-value, non-homology and the absence of any unusual problems. Hydrogen atoms were added in standard geometry and, where needed, with rotational optimization of OH, SH and NH_3^+ positions. Side-chain amide orientations were corrected where required by NH van der Waals clashes (Word *et al*., 1999a). The file ID code is followed by the subunit used.
[b]Proteins 1benA and 1benB are from the same PDB file.

corresponding heavy atom distances are not affected (Ramachandran *et al*., 1963).

Materials and methods

The amino acid atom nomenclature used here is in accordance with IUPAC standards (Markley *et al*., 1998). The PDB files dataset (Word *et al*., 1999a) used in this calculation describes protein structures with all protons processed by Reduce to add, optimize and make various minor corrections, including 180° flips of side chains where needed. The dataset includes 100 PDB files with wild-type proteins (Table I). The interproton distances and their corresponding heavy atom distances were calculated for the atom pairs taken from the PDB files (Word *et al*., 1999a). The two-dimensional scatter graphs and heat maps and three-dimensional histograms of the heavy atom distances versus their corresponding interproton distances are plotted. Only the distances that are <6.8 Å are plotted on the graphs because NMR restraints are rarely even that large.

In our calculations, the distances involving diastereotropic protons or methyl protons are calculated as the average of all possible proton pairs because ^{1}H NMR experimental data do not distinguish the three methyl protons. For example, the interproton distance between the HA of Gly and HA of Lys is calculated as the average of the two distances 1HA Gly–HA Lys and 2HA Gly–HA Lys. Diastereotropic α-protons and β-protons include HA of Gly and all HB except Ala, Gly, Ile, Thr and Val. Methyl protons include HB of Ala, HD1 and HG2 of Ile, HD1 and HD2 of Leu, HE of Met, HG2 of Thr, HG1 and HG2 of Val.

The data points are plotted according to their secondary structures, defined in the PDB files (Word *et al*., 1999a). The two-dimensional scatter graphs for heavy atom distances versus their corresponding interproton distances are plotted in red, green and blue representing α-helices, β-sheets and loops, respectively. These graphs are also fitted with linear equations for the optimal fits, upper bounds and lower bounds (see Table V). The optimal fits are determined by linear regression while the upper bounds and lower bounds are obtained visually. Moreover, two-dimensional heat maps and three-dimensional histograms are created to determine the spots with the highest data point concentrations. The inverse distance weighting interpolation method (Equation 1) (Shepard, 1968; McLain, 1976) is used in the smoothing transformation to transform the scatter graphs into two-dimensional heat maps and three-dimensional histograms.

$$Z = \sum_{i=1}^{n} W_i Z_i \Big/ \sum W_i$$

$$W_i = \left[(x_i - x)^2 + (y_i - y)^2 \right]^{-2} \qquad (1)$$

Results and discussion

Common interproton distances obtained by X-PLOR/CNS and DYANA/DIANA (Güntert *et al*., 1991, 1997; Brünger, 1992; Schwieters *et al*., 2003) were plotted against their corresponding heavy atom distances (Figures 1–5). The interactions between two adjacent residues show different patterns from the interactions between the residues that are more than one residue apart. Therefore, separate graphs are plotted for the interactions between two adjacent residues and all other interactions. The scatter plot in Figure 1 shows the three rotamers at β-carbons for α-helices (gray dots) and anti-parallel β-sheets (black dots). Part A of Figures 2–5 show the scatter plots according to their secondary structures: α-helices (red dots), β-sheets (green dots) and loops (blue dots). All plots show the heavy atom distances on the *y*-axis and the interproton distances on the *x*-axis.

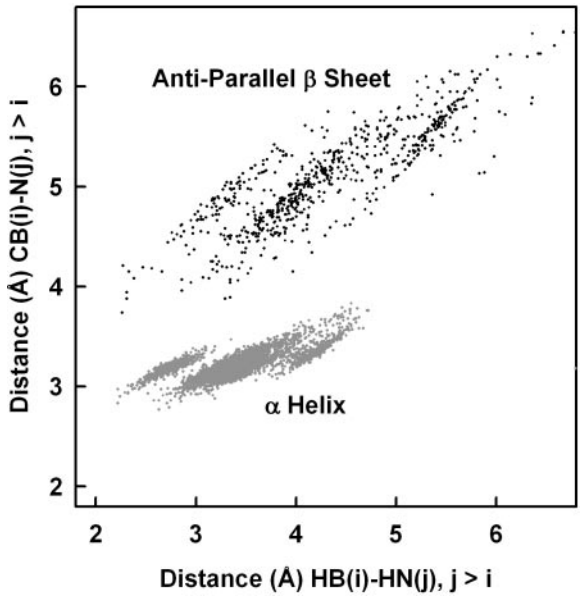
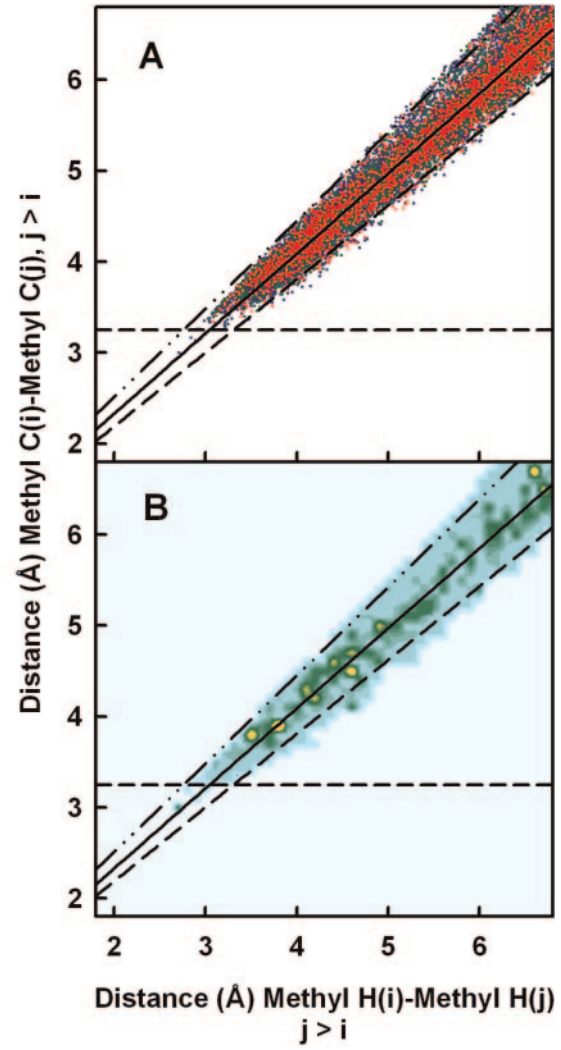
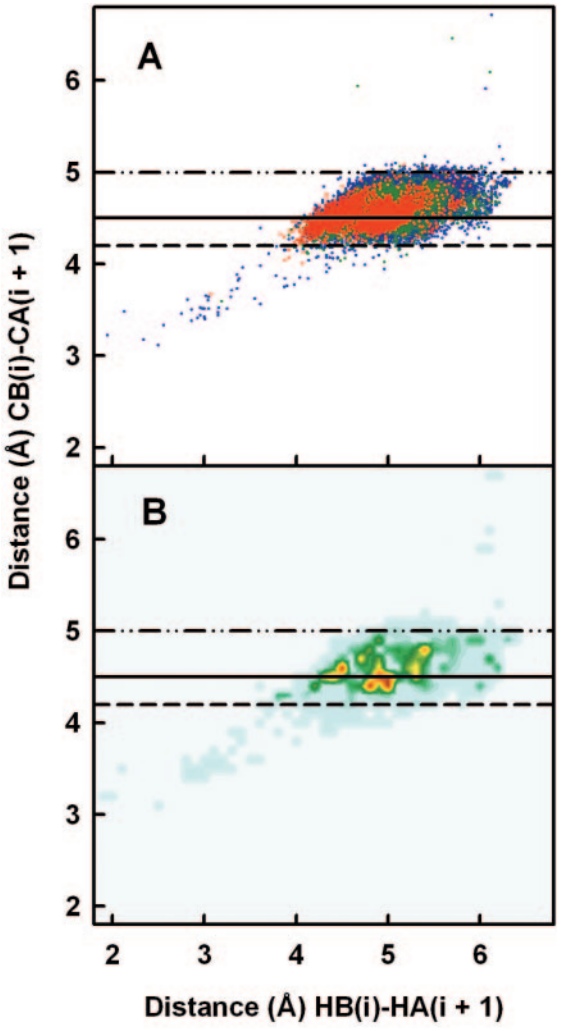


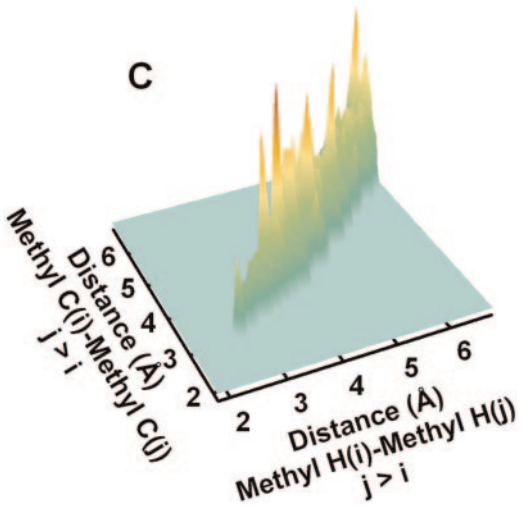
Fig. 1. Different rotamers observed at β-carbons in the distance relationship of CB(*i*)–N(*j*) versus HB(*i*)–HN(*j*). The α-helix and anti-parallel β-sheet interactions are represented in dark gray and black, respectively. The α-helix interactions are between two adjacent residues, HB(*i*)–HN(*j*), when $j = i + 1$. The anti-parallel β-sheet interactions are between two residues of adjacent strands, HB(*i*)–HN(*j*). In anti-parallel β-sheets, amide nitrogen and carbonyl oxygen of residue (*i*) are hydrogen bonded to carbonyl oxygen and amide nitrogen of residue (*j*), respectively.

The data points from α-helices are superimposed on top of those from β-sheets, which are also superimposed on top of those from loop secondary structures. These scatter plots are fitted with linear equations for optimal fits (——, black solid lines), upper limits (–···–, black lines with dots) and lower limits (– – –, black dashed lines) (part A of Figures 2–5). The resulting lines are listed in Table V. In addition, for clarity, Figure 4D shows a separate scatter plot exclusively for the data points of α-helices for the distance relationship between the heavy atom pairs and the corresponding proton pairs of HN(*i*)–HN($i + j$), when $j \geqslant 2$. The optimal fit line for α-helix interactions, $y = x$, is plotted as a solid green line in Figure 4D.

Part B of Figures 2–5 and part E of Figure 4 show the two-dimensional heat maps using inverse distance weighting interpolation (Equation 1) to represent the distribution density of the corresponding scatter plots of part A of Figures 2–5 and part D of Figure 4, respectively. Also, for clarity, the relationship between heavy atom distances and the corresponding interproton distances of HA(*i*)–HN($i + j$) are shown in two different two-dimensional heat maps for α-helices and β-sheets in Figure 5D and E, respectively. In addition, part C of Figures 2–5 and part F of Figure 4 show part B of 2–5 and part E of Figure 4, respectively, in three-dimensional histograms. From these two-dimensional heat maps and three-dimensional histograms, the relative distributions of the data







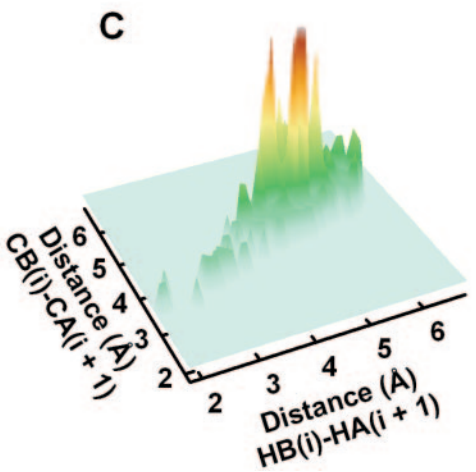


Fig. 2. The determination of the optimal fit, the lower bound and the upper bound for the calculation of the heavy atom distances, methyl carbon(*i*)–methyl carbon(*j*), from the corresponding interproton distances, methyl proton(*i*)–methyl proton(*j*). (**A**–**C**) show the relationships when $j > i$. In the scatter graph of (A), α-helix, β-sheet and loop data points are shown in red, green and blue, respectively. The two-dimensional heat map and three-dimensional histogram are shown in (B) and (C), transformed from (A) with Equation 1. Also, the optimal fit, the lower limit and the upper limit are shown in —— line, ---- line and – ·· – line, respectively. The optimal fit is determined to be $y = 0.57 + 0.88x$ (Table V).

Fig. 3. Determination of the optimal fit, the lower bound and the upper bound for the calculation of the heavy atom distances, CB(*i*)–CA(i + 1), from the corresponding interproton distances, HB(*i*)–HA(i + 1). In the scatter graph of (**A**), α-helix, β-sheet and loop data points are shown in red, green and blue, respectively. The two-dimensional heat map and three-dimensional histogram are shown in (**B**) and (**C**), transformed from (A) using Equation 1. Also, the optimal fit, the lower limit and the upper limit are shown in —— line, ---- line and – ·· – line, respectively. The optimal fit is determined to be $y = 4.50$. The slopes for upper limit, lower limit and optimal fit are zero (Table V).

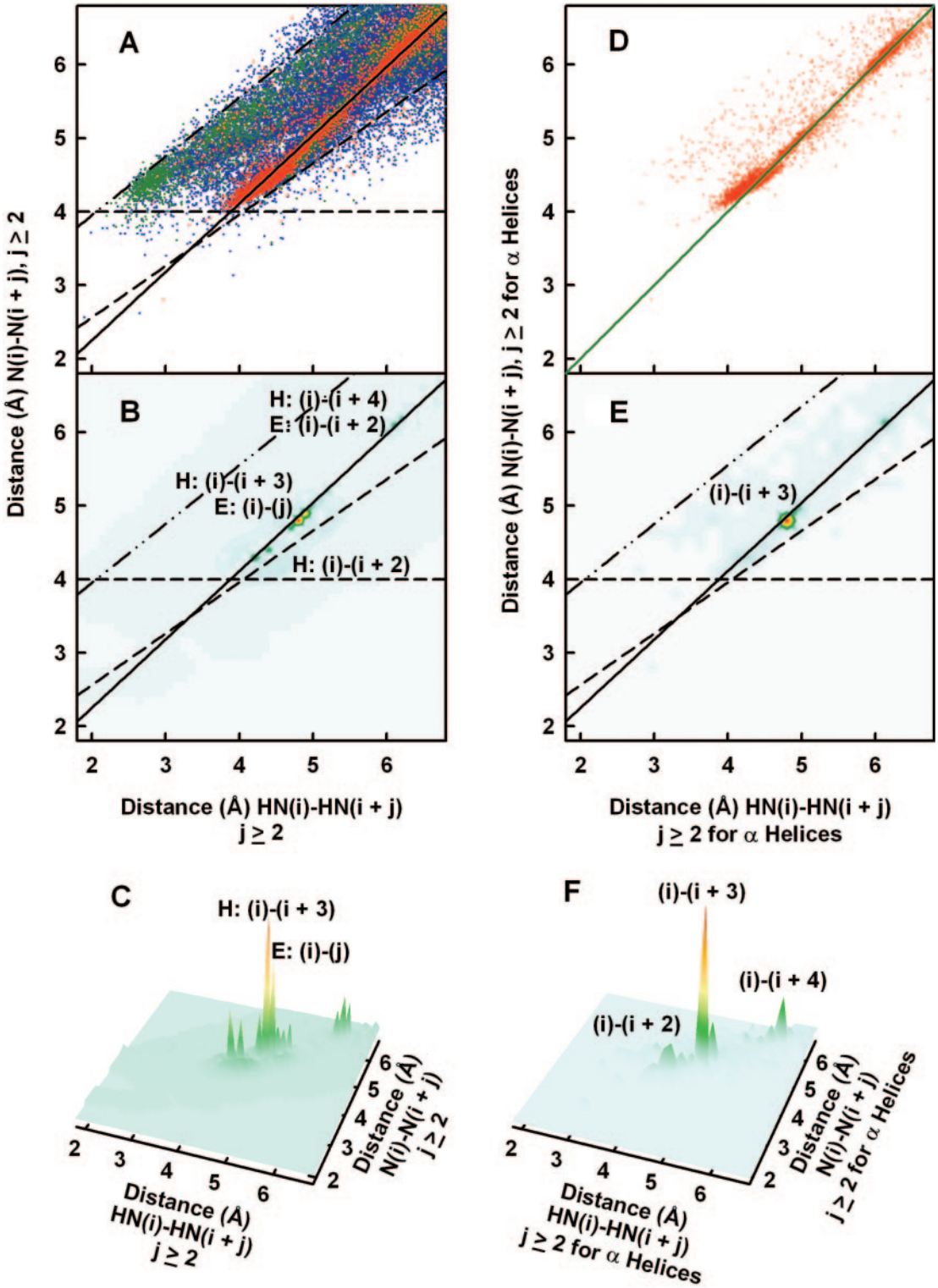


Fig. 4. Determination of the optimal fits, the lower bounds and the upper bound for the calculation of the heavy atom distances, N(i)–N($i + j$), from the corresponding interproton distances, HN(i)–HN($i + j$), when $j \geq 2$. In the scatter graph of (**A**), α-helix, β-sheet and loop data points are shown in red, green and blue, respectively. The two-dimensional heat map of (**B**) and three-dimensional histogram of (**C**) are shown, transformed from (A) using Equation 1. (**D**) shows the scatter plot for the data points that are from α-helix interactions only between the atoms that are more than 1 residue apart. (**E**) and (**F**) show the two-dimensional heat map and three-dimensional histograms, transformed from (D) using Equation 1, respectively. Also, the optimal fit, the lower bounds and the upper bound are shown in —— black line, ---- black lines and – ·· – black line, respectively, in (A–B) and (E). The optimal fit is determined to be $y = 0.39 + 0.93x$ (Table V). The optimal fit exclusively for α-helix data points is determined to be $y = x$ (solid green line, D). In each major data cluster in (B–C) and (E–F), the dominant contributing secondary structure interactions are also indicated on the graphs. H, α-helices. E, β-sheets. (**G**) and (**H**) show the patterns for anti-parallel and parallel β-sheets, respectively. In (G), the amide nitrogen and carbonyl oxygen of residue (i) are hydrogen bonded to carbonyl oxygen and amide nitrogen of residue (j), respectively, in anti-parallel β-sheets. In (H), the amide nitrogen and carbonyl oxygen of residue (i) are hydrogen bonded to carbonyl oxygen of residue ($j - 1$) and amide nitrogen of residue ($j + 1$), respectively, in parallel β-sheets. In each major data cluster in (G–H), the dominant contributing secondary structure interactions are also indicated on the graphs. Also, the optimal fit, the lower bounds and the upper bound are shown in —— black line, ---- black lines and – ·· – black line, respectively, in (G–H). The optimal fit is determined to be $y = 0.39 + 0.93x$ (Table V).

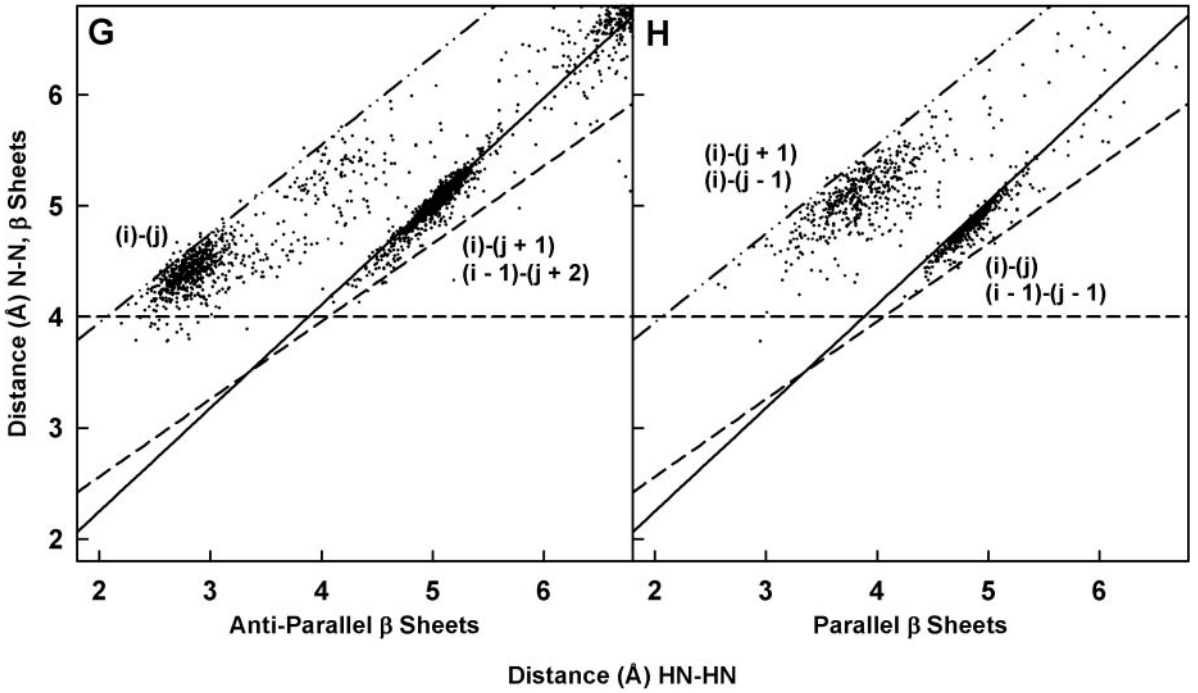


Fig. 4. Continued.

points from the original scatter plots (part A of 2–5 and part D of Figure 4) can be observed. Heavy concentrations of data points form clusters in the two-dimensional heat maps and three-dimensional histograms.

Figure 4G and H and Figure 5F and G are scatter graphs for various anti-parallel and parallel β-sheet interactions. The distances between adjacent residues among amide protons, α-protons and β-protons are listed in Table II for α-helices and β-sheets. The interproton distances between residues in adjacent strands in anti-parallel and parallel β-sheets are listed in Tables III and IV, respectively.

Distinct rotamers on β-carbon observed

In this study, if a β carbon has two β-protons, the two diastereotropic β-proton distances are averaged. In many cases where β-protons are studied, distinct clusters can be observed for each of the rotamers. The most striking example is CB(i)–N(j) vs HB(i)–HN(j) (Figure 1). In Figure 1, the gray dots are for α-helix residues where $j = i + 1$ and the black dots are for anti-parallel β-sheets where the amide nitrogen and carbonyl oxygen of residue (i) form hydrogen bonds with carbonyl oxygen and amide nitrogen of residue j, respectively. In both interactions of the α-helices and anti-parallel β-sheets, three distinct clusters are observed with the center one being the most prominent. In addition, the α-helix interactions of the heavy atom distances versus the corresponding interproton distances of HA(i)–HB(i + 1) (data not shown), HA(i)–HB(i + 2) (data not shown) and HN(i)–HB(i + 1) (data not shown) also show three distinct clusters with the center one being the most intense.

At β-carbons of most amino acids, three rotamers can form. The center prominent cluster is from the *trans* rotamer and the two smaller ones are from the two *gauche* rotamers. The *trans* rotamer is the most favorable configuration owing to its staggered geometry, so its cluster is also the strongest.

Interproton distances of side chains

Figure 2A–C show the graphs for the methyl carbon distances between all methyl carbons versus their corresponding methyl protons ($j > i$). The distances from all three methyl protons are averaged for each methyl carbon because ^{1}H NMR experimental data do not distinguish the three methyl protons. Figure 2A–C show that the data points are concentrated in a narrower range around the optimal fit (solid line) than any other distance relationship (Figures 1 and 3–5), probably because the three interproton distances are averaged, removing most variations due to rotation of the methyl groups. Also, no single prominent cluster of data points is observed. Instead, the data points are spread evenly between the upper limit and lower limit in Figure 2B and C.

Another side chain proton pair, HB(i)–HB(i + j), $j \geqslant 2$ (Table V, graph not shown), also shows a narrow range between its lower and upper bounds. However, the area for the data points in HB(i)–HB($i + j$) is still wider than the region between the lower bound and upper bound in methyl proton(i)–methyl proton(j) in Figure 2A–C. In Figure 2A–C for methyl carbon(i)–methyl carbon(j) versus methyl proton(i)–methyl proton(j), the absence of three rotamers at methyl carbons, due to three methyl protons, causes the data point region to be narrower than those side chain proton distances involved with rotamers, for example, the distance relationship of CB(i)–CB(i + j) and HB(i)–HB(i + j) (data not shown).

Interproton distances between adjacent residues

Figure 3A–C are the graphs for the heavy atom distances versus the corresponding interproton distances of HB(i)–HA(i + 1). The upper and lower bounds are y = 5.00 and 4.20, respectively, that is, the heavy atom distances are independent of the proton distances (Table V). The range for interproton distance of HB(i)–HA(i + 1) is wide, ~4.0–6.0 Å,

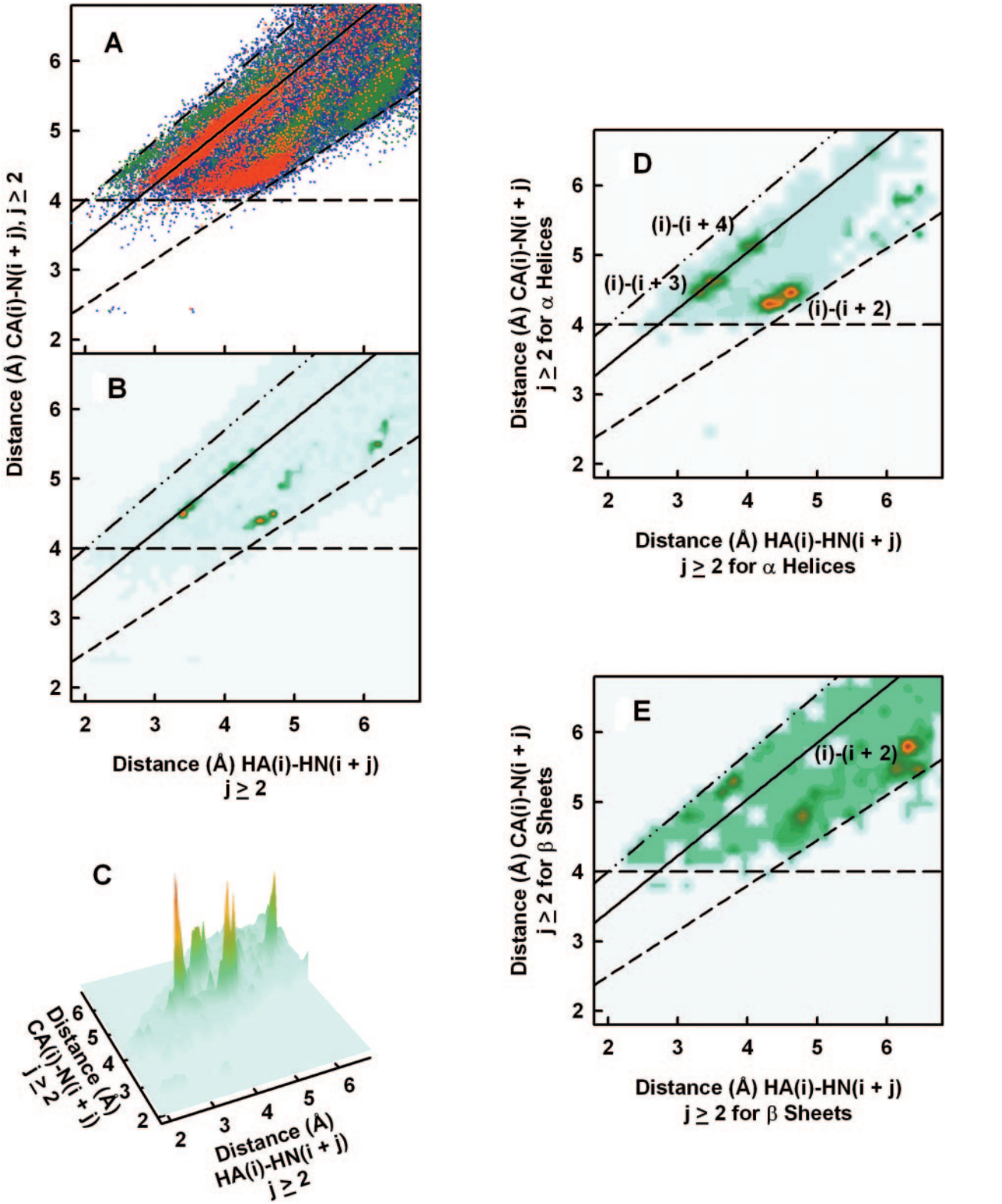


Fig. 5. The determination of the optimal fit, the lower bounds and the upper bound for the calculation of the heavy atom distance, CA(*i*)–N(*i* + *j*), from the corresponding interproton distance, HA(*i*)–HN(*i* + *j*), when $j \geq 2$. In the scatter graph of (**A**), α-helix, β-sheet and loop data points are shown in red, green and blue, respectively. The two-dimensional heat map of (**B**) and three-dimensional histogram of (**C**) are shown, transformed from (A) using Equation 1. (**D**) and (**E**) are heat maps for α-helices interactions and β-sheet interactions when $j \geq 2$, respectively. Also, the optimal fit, the lower bounds and the upper bound are shown in —— line, ---- lines and – ·· – line, respectively, in (A–B) and (D–E). The optimal fit is determined to be $y = 1.80 + 0.81x$ (Table V). In each major data cluster in (B) and (D–E), the dominant contributing secondary structure interactions are also indicated on the graphs. (**F**) and (**G**) show the patterns for anti-parallel and parallel β-sheets, respectively. In (F), the amide nitrogen and carbonyl oxygen of residue (*i*) are hydrogen bonded to carbonyl oxygen and amide nitrogen of residue (*j*), respectively, in anti-parallel β-sheets. In (G), the amide nitrogen and carbonyl oxygen of residue (*i*) are hydrogen bonded to carbonyl oxygen of residue (*j* – 1) and amide nitrogen of residue (*j* + 1), respectively, in parallel β-sheets. Also, the optimal fit, the lower bounds and the upper bound are shown in —— line, ---- lines and – ·· – line, respectively, in (F–G). The optimal fit is determined to be $y = 1.80 + 0.81x$ (Table V). In each major data cluster in (F–G), the dominant contributing secondary structure interactions are also indicated on the graphs.

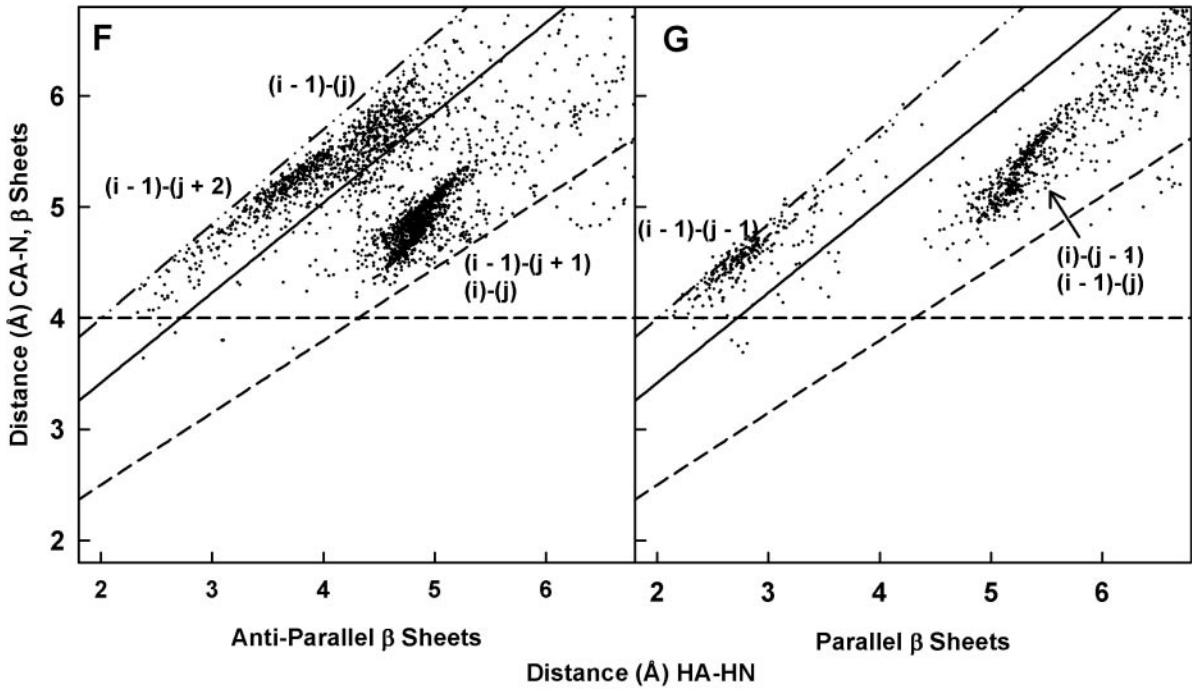


Fig. 5. Continued.

Table II. Average distances and their deviations among amide protons, α-protons and β-protons within adjacent four residues for α-helices and β-sheets determined from well-resolved protein crystal structures

	α-Helices			β-Sheets		
	HA(*i*)	HB(*i*)	HN(*i*)	HA(*i*)	HB(*i*)	HN(*i*)
HA(*i* + 1)	NA	4.71 ± 0.37	5.27 ± 0.17	NA	5.02 ± 0.55	4.93 ± 0.32
HA(*i* + 2)	6.61 ± 0.26	7.87 ± 0.36	7.04 ± 0.25	6.82 ± 0.47	6.77 ± 0.92	8.02 ± 0.66
HA(*i* + 3)	5.62 ± 0.50	7.94 ± 0.65	7.51 ± 0.58	>8.50	>8.50	>8.50
HA(*i* + 4)	6.68 ± 1.15	7.94 ± 1.10	8.77 ± 0.85	>8.50	>8.50	>8.50
HB(*i* + 1)	6.04 ± 0.34	5.72 ± 0.60	5.39 ± 0.51	5.03 ± 0.56	6.31 ± 0.66	6.22 ± 0.64
HB(*i* + 2)	6.09 ± 0.63	7.66 ± 0.60	6.12 ± 0.58	7.77 ± 0.78	7.24 ± 1.18	8.36 ± 0.94
HB(*i* + 3)	3.89 ± 1.32	6.16 ± 1.11	5.92 ± 0.67	>8.50	>8.50	>8.50
HB(*i* + 4)	6.26 ± 1.29	6.90 ± 1.44	7.20 ± 0.70	>8.50	>8.50	>8.50
HN(*i* + 1)	NA	3.36 ± 0.40	2.80 ± 0.31	NA	3.86 ± 0.74	4.29 ± 0.50
HN(*i* + 2)	4.44 ± 0.35	5.46 ± 0.38	4.33 ± 0.34	6.09 ± 0.63	6.15 ± 0.83	6.64 ± 0.61
HN(*i* + 3)	3.55 ± 0.43	5.54 ± 0.37	4.88 ± 0.33	>8.50	>8.50	>8.50
HN(*i* + 4)	4.29 ± 0.50	5.68 ± 0.40	6.35 ± 0.93	>8.50	>8.50	>8.50

NA, not available.

compared with the range of the corresponding heavy atom distance of CB(*i*)–CA(*i* + 1), ~4.2–5.0 Å. In addition, the distance relationships for interproton distances of HA(*i*)–HB(*i* + 1) (data not shown), HB(*i*)–HN(*i* + 1) (data not shown), HN(*i*)–HN(*i* + 1) (data not shown), HN(*i*)–HA(*i* + 1) (data not shown) and the distances of the corresponding heavy atoms also show that the upper and lower bounds have slopes of zero (Table V).

Second, the α-helix interactions (red dots) are concentrated into a more compact region than the β-sheet interactions (green dots) in Figure 3A. This can be explained by the fact that α-helices are more rigid structurally than β-sheets. The hydrogen bonding pattern in β-sheets allows strands to twist and bend. Hence clusters of α-helix interactions are more constrained than those of β-sheets interactions.

Linear relationship for α-helix interactions in the interproton distances of HN(i)–HN(i ± j) when j ⩾ 2

One major cluster and two minor clusters, mostly from α-helix interactions (red dots in Figure 4A and D), are observed along the optimal fit line (solid black line in Figure 4A, B and E, Table V) for the distance relationship between heavy atoms and the corresponding HN(*i*)–HN(*i* + *j*) when $j \geq 2$. The solid black line in Figure 4A, B and E, $y = 0.39 + 0.93x$ (Table V), is the optimal fit for the interactions from all secondary structures; whereas the solid green line in Figure 4D, $y = x$, is the optimal fit for the interactions only from α-helices. These two optimal fits are very similar to each other. In addition, the major cluster in the center gives a much higher intensity than the other two smaller clusters (Figure 4B, C, E and F). Data points from β-sheet (green dots) and loop (blue dots) interactions

Table III. Average distances and their deviations of proton pairs among amide protons, α-protons and β-protons for anti-parallel β-sheets determined from well-resolved protein crystal structures[a]

	HA(i)	HB(i)	HN(i)	HA($i-1$)	HB($i-1$)	HN($i-1$)
HA($j-1$)	5.87 ± 0.94	5.81 ± 0.94	4.77 ± 0.78	>8.50	>8.50	>8.50
HA(j)	7.30 ± 0.42	6.23 ± 0.97	4.83 ± 0.33	5.65 ± 0.51	7.12 ± 0.98	8.27 ± 0.66
HA($j+1$)	6.31 ± 0.64	5.97 ± 0.94	3.71 ± 0.48	2.68 ± 0.67	4.72 ± 0.74	4.91 ± 0.58
HA($j+2$)	>8.50	>8.50	>8.50	6.32 ± 0.65	8.45 ± 0.94	7.75 ± 0.66
HB($j-1$)	7.13 ± 1.09	7.38 ± 1.19	5.67 ± 0.97	>8.50	>8.50	>8.50
HB(j)	6.22 ± 0.89	4.61 ± 1.26	4.23 ± 0.77	5.62 ± 0.70	7.35 ± 1.14	7.99 ± 1.04
HB($j+1$)	8.55 ± 0.82	8.18 ± 1.00	5.90 ± 0.66	4.84 ± 0.89	6.31 ± 1.05	6.87 ± 0.93
HB($j+2$)	>8.50	>8.50	>8.50	5.99 ± 0.92	8.09 ± 1.12	6.82 ± 1.07
HN($j-1$)	8.25 ± 1.06	7.96 ± 1.26	6.90 ± 0.91	>8.50	>8.50	>8.50
HN(j)	4.98 ± 0.63	4.37 ± 0.94	3.04 ± 0.62	4.68 ± 0.62	5.72 ± 0.99	7.00 ± 0.76
HN($j+1$)	7.79 ± 0.45	6.83 ± 0.94	5.05 ± 0.28	5.02 ± 0.54	6.81 ± 0.86	7.55 ± 0.61
HN($j+2$)	>8.50	>8.50	>8.50	3.76 ± 0.69	5.85 ± 0.76	5.07 ± 0.60

[a]Residues i and j are hydrogen bonded to each other in adjacent anti-parallel β-strands. The amide nitrogen and carbonyl oxygen atoms of residue i are hydrogen bonded to the carbonyl oxygen and amide nitrogen atoms of residue j in anti-parallel β-sheet formation, respectively.

Table IV. Average distances and their deviations of proton pairs among amide protons, α-protons and β-protons in parallel β-sheets determined from well-resolved protein crystal structures[a]

	HA(i)	HB(i)	HN(i)	HA($i-1$)	HB($i-1$)	HN($i-1$)
HA($j-2$)	>8.50	>8.50	>8.50	4.67 ± 0.57	6.88 ± 0.66	5.67 ± 0.83
HA($j-1$)	7.93 ± 0.48	6.96 ± 0.88	5.17 ± 0.25	4.87 ± 0.40	6.92 ± 0.82	7.43 ± 0.47
HA(j)	4.96 ± 0.40	3.87 ± 0.95	2.90 ± 0.46	4.57 ± 0.49	6.15 ± 0.92	7.05 ± 0.51
HA($j+1$)	7.93 ± 0.52	7.58 ± 0.80	6.35 ± 0.48	>8.50	>8.50	>8.50
HB($j-2$)	>8.50	>8.50	>8.50	6.16 ± 0.89	8.49 ± 1.14	7.48 ± 1.06
HB($j-1$)	7.40 ± 0.66	7.17 ± 0.86	4.69 ± 0.58	3.62 ± 0.87	5.13 ± 1.06	6.02 ± 0.91
HB(j)	7.21 ± 0.83	5.67 ± 1.26	5.35 ± 0.65	6.79 ± 0.59	>8.50	>8.50
HB($j+1$)	6.74 ± 0.94	7.11 ± 0.85	5.76 ± 0.80	>8.50	>8.50	>8.50
HN($j-2$)	>8.50	>8.50	>8.50	6.82 ± 1.01	>8.50	8.07 ± 0.94
HN($j-1$)	6.43 ± 0.59	5.89 ± 0.87	3.94 ± 0.53	2.91 ± 0.60	5.25 ± 0.64	4.93 ± 0.51
HN(j)	7.47 ± 0.33	6.27 ± 0.91	4.85 ± 0.18	5.46 ± 0.40	7.31 ± 0.88	8.20 ± 0.37
HN($j+1$)	5.20 ± 0.50	4.92 ± 1.13	3.89 ± 0.49	>8.50	>8.50	>8.50

[a]Residues i and j are in adjacent parallel β-strands. The amide nitrogen atom of residue i is hydrogen bonded to the carbonyl oxygen atom of residue $j-1$. Also, the carbonyl oxygen atom of residue i is hydrogen bonded to the amide nitrogen of residue $j+1$.

are distributed over a wider area than α-helices (red dots) (Figure 4A).

The most prominent cluster is from the interactions of α-helix residue pairs of HN(i)–HN($i+3$) and anti-parallel and parallel β-sheets interactions (Figure 4B and C). Moreover, Figure 4D and F show that the concentration for α-helix data points of HN(i)–HN($i+3$) is much more dense than the concentrations of data points for HN(i)–HN($i+2$) or HN(i)–HN($i+4$) from the same secondary structure. This pattern suggests that the alignments of HN(i)–HN($i+3$) and N(i)–N($i+3$) is more rigid than others because N(i) and N($i+3$) are on the same plane. In α-helices, O($i-1$) and N($i+3$) form a hydrogen bond. Moreover, the peptide bond between O($i-1$) and N(i) cannot be freely rotated.

The three clusters for α-helices, HN(i)–HN($i+2$), HN(i)–HN($i+3$) and HN(i)–HN($i+4$), are on the optimal fit line for α-helices, $y = x$ (green solid line in Figure 4D). This indicates that, at least for α-helices, the distances between amide proton pairs and between the corresponding amide nitrogen pairs are essentially the same when they are two, three or four residues apart. This linear relationship is due to the sp^2 hybridization-like nature of amide nitrogen. The sp^2 hybridization-like nature forces amide nitrogen to be trigonal planar with its lone pair interacting with electron deficient carbonyl carbon in the resonance form. All amide protons in α-helices point in the same direction. No other interaction shows this kind of linear relationship in our study.

Non-linear relationship for α-helix interactions in the interproton distances of HA(i)–HN(i ± j) when j ⩾ 2

Figure 5A shows the scatter plot for the heavy atom distances versus their corresponding interproton distances of HA(i)–HN($i+j$) when $j \geqslant 2$. In Figure 5B and C, five clusters of data points are observed. Two clusters, HA(i)–HN($i+3$) and HA(i)–HN($i+4$), which are formed mostly by the interactions from α-helices (Figure 5D, red dots in Figure 5A), are found along the optimal fit, $y = 1.80 + 0.81x$ (solid line in Figure 5A, B, D and E, Table V). Another cluster in Figure 5B–D, HA(i)–HN($i+2$), located around the intersection of the two lower bound limits (dash line), is also from interactions in α-helices (red dots in Figure 5A).

In contrast to N(i)–N($i+j$) versus HN(i)–HN($i+j$) in Figure 4A–E, the three data clusters from α-helix interactions are not aligned linearly in Figure 5D. The positions of the clusters show that the interproton distance and the corresponding heavy atom distance for α-helices are not the same. We believe that this difference is due to the different hybridizations in amide nitrogen and α-carbon. The hybridization in α-carbon is sp^3 with tetrahedral geometry. This tetrahedral geometry forces the α-carbon to protrude out of the α-helical face and

Table V. Optimal fits and lower and upper bounds for the translation of distances of various proton pairs between amide protons, α-protons and β-protons into the corresponding heavy atom pairs determined in the study of the distance relationships[a]

Proton pair	Optimal fit	Lower bounds[b]	Upper bound
HA(i)–HB(i + 1)	$y = 4.90$	$y = 4.40$	$y = 5.00$
HB(i)–HA(i + 1)	$y = 4.50$	$y = 4.20$	$y = 5.00$
HN(i)–HN(i + 1)	$y = 1.55 + 0.45x$	$y = 2.60$	$y = 3.75$
HN(i)–HA(i + 1)	$y = 0.50 + 0.70x$	$y = 4.05$	$y = 4.95$
HB(i)–HN(i + 1)	$y = 1.85 + 0.40x$	$y = 2.90$	$y = 4.00$
HB(i)–HB(i + 1)	$y = 3.35 + 0.37x$	$y = 4.50$	$y = 3.00 + 0.50x$
HN(i)–HB(i + 1)	$y = 2.25 + 0.50x$	$y = 4.50$	$y = 2.35 + 0.60x$
MeH(i)–MeH(j), $j > i$	$y = 0.57 + 0.88x$	$y = 0.57 + 0.81x$ $y = 3.25$	$y = 0.57 + 0.97x$
HA(i)–HN($i + j$), $j \geqslant 2$	$y = 1.80 + 0.81x$	$y = 1.20 + 0.65x$ $y = 4.00$	$y = 2.30 + 0.85x$
HA(i)–HA($i + j$), $j \geqslant 2$	$y = x$	$y = 2.80 + 0.35x$ $y = 4.00$	$y = 2.20 + x$
HA(i)–HB($i + j$), $j \geqslant 2$	$y = 2.30 + 0.60x$	$y = 3.35 + 0.20x$ $y = 3.90$	$y = 2.40 + 0.80x$
HB(i)–HN($i + j$), $j \geqslant 2$	$y = -0.20 + 1.1x$	$y = 1.75 + 0.50x$ $y = 3.90$	$y = 2.30 + 0.80x$
HB(i)–HA($i + j$), $j \geqslant 2$	$y = 1.90 + 0.75x$	$y = 2.60 + 0.35x$ $y = 3.80$	$y = 1.80 + 0.95x$
HB(i)–HA($i + j$), $j \geqslant 2$	$y = 1.60 + 0.70x$	$y = 1.00 + 0.65x$ $y = 3.50$	$y = 1.80 + 0.90x$
HN(i)–HN($i + j$), $j \geqslant 2$	$y = 0.39 + 0.93x$	$y = 1.16 + 0.70x$ $y = 4.00$	$y = 2.35 + 0.80x$
HN(i)–HA($i + j$), $j \geqslant 2$	$y = 0.10 + x$	$y = 2.70 + 0.35x$ $y = 4.00$	$y = 2.5 + 0.80x$
HN(i)–HB($i + j$), $j \geqslant 2$	$y = 0.60 + x$	$y = 1 + 0.65x$ $y = 3.80$	$y = 2.35 + 0.80x$

[a]The optimal fits and lower and uppers bounds are determined from the interactions of all secondary structures.
[b]Lower bounds may have two fits for the same proton pair when the fits with a slope of zero may be applied to shorter interproton distances (usually <4 Å) and the fits with slope $\neq 0$ may be applied to longer interproton distances (usually >4 Å).

the α-protons to point outwards in different directions. In addition, the distance relationship between heavy atom pairs and their corresponding proton pairs of HA(i)–HA($i + j$) when $j \geqslant 2$ (data not shown) also show this non-linear relationship.

Conclusion

First, in many examples involving distances from β-protons, distinct rotamers can be observed in α-helices and β-sheets. For example, in the relationship between the distances between heavy atoms and between the corresponding β-protons and amide protons, three distinct clusters, representing each rotamer at the β-carbon, can be observed. The largest cluster is from the *trans* rotamer and the two smaller clusters are from the two *gauche* rotamers.

Second, we found that in short-range interproton distances of HB(i)–HA(i + 1), HA(i)–HB(i + 1), HB(i)–HN(i + 1), HN(i)–HN(i + 1) and HN(i)–HA(i + 1), the lower and upper bounds for the translation have slopes of zero. This suggests that the range allowed for heavy atom distances from the translation of their interproton distances is very small. We conclude that interproton distances that are less than six heavy atoms away are independent of their corresponding heavy atom distances.

Finally, in the distance relationship of N(i)–N(i + j) and HN(i)–HN(i + j), all three (j = 2, 3 or 4) α-helix data clusters are aligned linearly along the line of $y = x$. The linear relationship indicates that the distances of both N(i)–N(i + j) and HN(i)–HN(i + j) are essentially the same, if they are part of an α-helix. In contrast, the distance relationship of CA(i)–N(i + j) and HA(i)–HN(i + j) shows the three α-helix clusters aligned non-linearly and the distances CA(i)–N(i + j) and HA(i)–HN(i + j) are not the same. This difference can be attributed to the different hybridizations on amide nitrogen (sp^2-like trigonal planar geometry) and on α-carbon (sp^3 tetrahedral geometry). In α-helix secondary structure, all amide protons point in the same direction towards the N-terminus. In contrast, α-protons of α-helices point outwards, away from α-helices.

References

Bax,A. and Grzesiek,S. (1993) *Acc. Chem. Res.*, **26**, 131–138.
Bowers,P.M., Strauss,C.E.M. and Baker,D. (2000) *J. Biomol. NMR*, **18**, 311–318.
Brünger,A.T. (1992) *X-PLOR Version 3.1: a System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
Clore,G.M. and Gronenborn,A.M. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 5891–5898.
Güntert,P., Braun,W. and Wüthrich,K. (1991) *J. Mol. Biol.*, **217**, 517–530.
Güntert,P., Mumenthaler,C. and Wüthrich,K. (1997) *J. Mol. Biol.*, **273**, 283–298.
Hung,L.H. and Samudrala,R. (2003) *Nucleic Acids Res.*, **31**, 3296–3299.
Karplus,K., Karchin,R., Draper,J., Casper,J., Mandel-Gutfreund,Y., Diekhans,M and Hughey,R. (2003) *Proteins: Struct. Funct. Genet.*, **53**, Suppl 6, 491–496.
Li,W., Zhang,Y. and Skolnick,L. (2004) *Biophys. J.*, **87**, 1241–1248.
Markley,J.L., Bax,A., Arata,Y., Hilbers,C.W., Kaptein,R., Sykes,B.D., Wright,P.E. and Wüthrich,K. (1998) *J. Mol. Biol.*, **280**, 933–952.
McLain,D.H. (1976) *Comput. J.*, **19**, 178–181.
Meiler,J. and Baker,D. (2003) *Proc. Natl Acad. Sci. USA*, **100,** 15404–15409.
Nilges,M., Clore,G.M. and Gronenborn,A.M. (1988) *FEBS Lett.*, **229**, 317–324.
Ramachandran,G.N., Ramakrishnan,C. and Sasisekharan,V. (1963) *J. Mol. Biol.*, **7**, 95–99.
Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) *J. Magn. Reson.*, **160**, 66–74.
Shepard,D. (1968) In *Proceedings of 23rd ACM National Conference*, pp. 517–524.
Word,J.M., Lovell,S.C., La Bean,T.H., Taylor,H.C., Zalis,M.E., Presley,B.K., Richardson,J.S. and Richardson,D.C. (1999a) *J. Mol. Biol.*, **285**, 1711–1733.
Word,J.M., Lovell,S.C., Richardson,J.S. and Richardson,D.C. (1999b) *J. Mol. Biol.*, **285**, 1735–1747.
Zheng,W. and Doniach,S. (2002) *J. Mol. Biol.*, **316**, 173–187.

Received August 17, 2005; accepted September 8, 2005

Edited by Marius Clore