Journal of Biomolecular NMR, **11:** 45–57, 1998.
KLUWER/ESCOM

Chemical shift homology in proteins

Barbara C.M. Potts* & Walter J. Chazin**
Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

Received 1 April 1997; Accepted 28 July 1997

Key words: chemical shift, protein structural homology, resonance assignments, secondary chemical shift, sequence homology

Abstract

The degree of chemical shift similarity for homologous proteins has been determined from a chemical shift database of over 50 proteins representing a variety of families and folds, and spanning a wide range of sequence homologies. After sequence alignment, the similarity of the secondary chemical shifts of C^{α} protons was examined as a function of amino acid sequence identity for 37 pairs of structurally homologous proteins. A correlation between sequence identity and secondary chemical shift rmsd was observed. Important insights are provided by examining the sequence identity of homologous proteins versus percentage of secondary chemical shifts that fall within 0.1 and 0.3 ppm thresholds. These results begin to establish practical guidelines for the extent of chemical shift similarity to expect among structurally homologous proteins.

Introduction

The determination of protein structures by NMR spectroscopy depends largely on distance restraints derived from NOESY spectra. However, a significant effort is devoted toward identifying the chemical shift of each proton in the molecule in order to assign the NOESY cross peaks to specific proton–proton contacts in the polypeptide. In fact, the chemical shift itself contains a very substantial amount of structural information and this parameter is receiving an increasing level of attention as a means to supplement the NOEs for protein structure determination (for a review see Oldfield (1995) and Szilagyi (1995)). Efforts have included the development of approaches to delineate the elements of regular secondary structure (Spera and Bax, 1991; Wishart et al., 1991, 1992; Gronenborn and Clore, 1994; Wishart and Sykes, 1994) and to refine NMR structures directly against not only ^{1}H but also ^{13}C chemical shifts (Ösapay et al., 1994; Kuszewski et al., 1995; Williamson et al., 1995).

It has been recognized for many years that molecules of similar composition and structure will have similar chemical shifts, and that the degree of chemical shift similarity is related to the extent of structural similarity. For example, the similarity of the ^{1}H chemical shifts of one protein to those of a homologous protein or a mutant is often used as an indicator that the newly assigned protein adopts the same global fold. This suggests that the chemical shift could be exploited to assign the resonances of one protein, simply by comparison with existing assignments of another homologous protein. This strategy would expedite the resonance assignment process, thereby reducing the effort required to determine protein structure.

Despite an extensive amount of anecdotal evidence, there is very little information available on how similar one can expect chemical shifts to be. In particular, there have been numerous examples of chemical shift similarities between a wild-type protein and single site mutants (e.g., Wittekind et al. (1989); Carlström and Chazin (1993)), but only few comparisons of more divergent pairs of homologous proteins (e.g., Redfield and Dobson (1990); Potts et al. (1996)). To begin to shed some light on this issue, we have

*Present address: Agouron Pharmaceuticals, 3565 General Atomics Court, San Diego, CA 92121, U.S.A. **To whom correspondence should be addressed.

undertaken an empirical investigation of the chemical shift similarities between homologous proteins. Our strategy involved defining the extent of chemical shift similarity for over 50 pairs of structurally homologous proteins, then determining the correlation between chemical shift similarity and sequence homology, a relationship we term 'chemical shift homology'. The secondary chemical shifts are utilized for these comparisons, but no additional contributions to differences in chemical shifts are accounted for, because we are interested in the practical possibilities of utilizing a previously assigned protein to generate the resonance assignments for a homologue.

Materials and Methods

Amino acid sequence alignment and percentage sequence identity

A total of 52 unique protein sequences were employed in this study, with sequence alignments made for those proteins known to be structurally homologous. Of these proteins, 37 pairs were selected to generate the diverse range of protein structures and degrees of chemical shift homologies used for the general analysis (Table 1). A further 19 pairs were generated for the family-specific analysis of eight EF-hand calcium-binding proteins (Table 2).

The amino acid sequences were aligned based on combined information from published references (Eklund et al., 1991; Kretsinger et al., 1991; Sykes, 1991; Kuriyan and Cowburn, 1993; Skelton et al., 1995) and three-dimensional (3D) structures obtained from the Protein Databank (PDB) accession codes: 9pcy (plastocyanin, French bean); 1plc (plastocyanin, poplar); 5azu (azurin); 1rec (recoverin); 2bca (calbindin D_{9k}, bovine, calcium-loaded); 1clb (calbindin D_{9k}, bovine, apo); 1hid (HPr, *B. subtilis*); 1hdn (HPr, *E. coli*); 3trx (thioredoxin, human); 1trx (thioredoxin, *E. coli*); 1lz1 (lysozyme, human); 1hel and 1hwa (lysozyme, hen); 135l (lysozyme, turkey); 1hun (MIP-1β, human); 1rto (RANTES); 1dvh (cytochrome c553, *D. vulgaris*); 1cch (cytochrome c551, *P. stutzeri*); 1mb3 (c-Myb repeat 1); 1mbg (c-Myb repeat 2); 1mbj (c-Myb repeat 3); 1hsq (SH3 domain of phospholipase C-γ); 1srl (SH3 domain of Src); 2pnb (SH2 domain of phosphatidylinositol 3′-kinase p85-α subunit); 2pld (SH2 domain of phospholipase C-γ1). The sequence alignment derived from examination of the 3D structures does not always coincide with the 'best-fit' sequence alignment, i.e., the alignment which would give rise to the highest percentage of sequence identity. In these cases, the sequence alignment derived from the 3D structures was used for our analysis.

The percentage of sequence identity was calculated from the aligned sequences. If the sequence of one protein was longer at either the N- or C-terminus, then the extra residues were ignored in the sequence comparison. However, if the sequence of one protein was longer in the middle, then gaps were inserted to align the two sequences, and the extra residues were included in the total residue count for the pair. This results in a lower percentage of sequence identity than would be obtained if these extra residues were ignored.

NMR resonance assignment data

A total of 56 unique sets of $C^{\alpha}H$ resonance assignments were employed in this study. Of these, 52 are from different proteins and the remainder arise from inclusion of the apo and Ca^{2+}-loaded states of some of the Ca^{2+}-binding proteins. No changes have been made to any of the assignment lists, even in cases where we believe that certain values are in error. The data were primarily obtained from the BioMagResBank (Seavey et al., 1991), in some cases using the NIH WWW Database Gateway to the BioMagResBank. Additional data were transcribed from published reports or received directly from the authors, as indicated in the tables. The data were corrected as necessary according to Wishart et al. (1995) so that all of the chemical shifts are referenced as closely as possible to the same standard compound, DSS at 0.00 ppm. The secondary chemical shift was calculated for each $C^{\alpha}H$ resonance with the program Sshift (Gippert, 1995), which employs the random coil chemical shift values from Wishart et al. (1991), with averaged $C^{\alpha}H$ values for glycines.

The rmsd between the $C^{\alpha}H$ secondary chemical shifts was calculated for each of the protein pairs. If insertion of a gap was required in order to align the sequences, or if a chemical shift data point was unavailable, then the corresponding resonance was omitted from all chemical shift comparisons but not from the calculation of percent sequence identity.

Results and Discussion

Although most biomolecular NMR spectroscopists are convinced of the wealth of structural information stored in the chemical shift parameter, to our knowledge there have been no attempts to develop estimates

Table 1. Chemical shift homologies of selected proteins

	Pairs of homologous proteins[a]	Number of residues[a]	% sequence identity	Number of residues compared[b]	Chemical shift[c] rmsd	% shifts[c] within 0.3 ppm	% shifts[c] within 0.1 ppm	Reference[d]
1	Plastocyanin, French bean	99	52.5	97	0.25	79.3	48.5	B,1
	Plastocyanin, algae	97						B,2
2	Plastocyanin, French bean	99	83.8	97	0.12	95.9	80.4	B,1
	Plastocyanin, spinach	99						B,3
3	Plastocyanin, French bean	99	78.8	99	0.15	94.9	68.7	B,1
	Plastocyanin, poplar	99						4
4	Plastocyanin, French bean	99	18.2	91	0.54	39.6	13.2	B,1
	Azurin	128						5
5	Calbindin D_{9k}, bovine, $(Ca^{2+})_2$	76	88.2	76	0.09	98.7	90.8	B,6
	Calbinin D_{9k}, porcine, $(Ca^{2+})_2$	78						B,7
6	N-domain of troponin C, $(Ca^{2+})_2$	90	65.0	80	0.14	96.3	57.5	8
	Calmodulin, $(Ca^{2+})_2$ (N-term.)	148						B,9
7	Calcyclin, rabbit, apo	90	40.0	86	0.31	70.9	31.4	10
	S100β, rat, apo	91						11
8	Calmodulin, apo (C-term.)	148	18.4	70	0.33	62.9	28.6	12
	Calcyclin, rabbit, apo	90						10
9	Recoverin, apo (C-term.)	202	17.0	74	0.38	71.6	35.6	13
	Calbindin D_{9k}, bovine, apo	76						14
10	HPr, *S. aureus*	87	31.8	85	0.41	58.8	28.2	B,15
	HPr, *E. coli*	85						B,16
11	HPr, *S. aureus*	87	61.6	86	0.32	73.3	32.6	B,15
	HPr, *B. subtilis*	87						B,17
12	Thioredoxin, human	105	26.9	103	0.38	71.8	28.2	B,18
	Thioredoxin, *E. coli*	108						19
13	Trypsin-inhibitor K	57	42.1	57	0.27	77.2	31.6	G,20
	BPTI	58						G,21
14	Trypsin-inhibitor E	57	43.9	55	0.27	70.9	38.2	G,22
	BPTI	58						G,21
15	Trypsin-inhibitor E	57	61.4	55	0.20	83.6	54.5	G,22
	Trypsin-inhibitor K	57						G,20
16	Lysozyme, human	130	60.0	124	0.25	79.0	50.8	23
	Lysozyme, hen	129						B,24
17	Lysozyme, turkey	129	93.0	124	0.07	99.2	92.7	B,25
	Lysozyme, hen	129						B,24

Table 1. Continued

	Pairs of homologous proteins[a]	Number of residues[a]	% sequence identity	Number of residues compared[b]	Chemical shift[c] rmsd	% shifts[c] within 0.3 ppm	% shifts[c] within 0.1 ppm	Reference[d]
18	EETI-II CMTI-I	28 29	71.4	28	0.16	85.7	64.3	G,26 G,27
19	CMTI-I CMTI-III	29 29	96.6	29	0.07	100.0	86.2	G,27 G,28
20	α-Neurotoxin, *D. polypepis* cobratoxin, *Naja naja atra*	60 62	54.8	54	0.38	75.9	35.2	B,29 B,30
21	RANTES hMIP-1β	68 69	48.5	68	0.20	88.2	50.0	31 32
22	EGF, human EGF, murine	53 53	69.8	53	0.20	88.7	56.6	B,33 B,34
23	EGF domain, coagulation factor X EGF domain, coagulation factor IX	42 43	59.5	41	0.22	82.9	53.7	B,35 B,36
24	EGF domain, coagulation factor IX EGF, murine	43 53	31.9	42	0.29	69.0	23.8	B,36 B,34
25	Cytochrome b5, calf Cytochrome b5, pig	82 82	92.7	80	0.11	98.8	96.3	B,37 B,37
26	Cytochrome c551, *P. stutzeri* Cytochrome c551, *P. aeruginosa*	82 82	68.3	79	0.22	88.6	54.4	38 B,39
27	Cytochrome c551, *P. stutzeri* Cytochrome c553, *D. vulgaris*	82 79	21.1	63	0.67	54.0	20.6	38 40
28	Zinc finger, Xfin-31B Zinc finger, Xfin-31C	25 25	92.0	25	0.19	88.0	68.0	G,41 G,41
29	Zinc finger, Xfin-31B Zinc finger, ZFY switch	25 30	32.0	25	0.30	64.0	20.0	G,41 42
30	Zinc finger, TFIIIA-2 Zinc finger, Xfin-31	30 25	29.6	25	0.28	76.0	32.0	43 44
31	Zinc finger, Xfin-31 Zinc finger, Xfin-31B	25 25	92.0	25	0.29	84.0	72.0	44 G,41
32	c-Myb (R1 domain) c-Myb (R2 domain)	157 157	48.1	48	0.23	85.4	43.8	45 45
33	c-Myb (R2 domain) c-Myb (R3 domain)	157 157	30.8	51	0.37	80.4	35.3	45 45
34	c-Myb (R2,R3 domains) R2/R3 domain of b-Myb	157 110	84.0	106	0.36	94.3	75.5	45 46

Table 1. Continued

	Pairs of homologous proteins[a]	Number of residues[a]	% sequence identity	Number of residues compared[b]	Chemical shift[c] rmsd	% shifts[c] within 0.3 ppm	% shifts[c] within 0.1 ppm	Reference[d]
35	SH3 domain, Src	64	32.3	61	0.49	62.3	21.3	B,47
	SH3 domain, PLC-γ	72						48
36	SH3 domain, Src	64	29.3	55	0.45	61.8	20.0	B,47
	SH3 domain, Drk	59						49
37	N-terminal SH2 domain, p85-α	106	22.4	87	0.33	64.4	25.3	50
	C-terminal SH2 domain, PLC-γ1	103						51

[a] The entry for 'Number of residues' corresponds to the construct used to acquire the NMR data. When a domain from a larger protein was used in our analysis, the domain was identified in parenthesis next to the name of the protein. For example, in comparison #6 the isolated N-domain of troponin C was expressed and used for NMR analysis, whereas the data for the N-terminal domain of calmodulin were obtained from an analysis of the intact protein.

[b] The number of C^{α} proton resonances compared is limited by insertions/gaps in the sequence alignment and/or unassigned resonances.

[c] Secondary chemical shifts of C^{α} protons.

[d] References: B, accessed entry in the BioMagResBank; G, accessed entry using the NIH WWW Database Gateway to BioMagResBank; (1) Chazin, W.J. and Wright, P.E. (1988) *J. Mol. Biol.*, **202**, 623–636; (2) Moore, J., Chazin, W.J., Powls, R. and Wright, P.E. (1988) *Biochemistry*, **27**, 7806–7816; (3) Driscoll, P., Hill, H. and Redfield, C. (1987) *Eur. J. Biochem.*, **170**, 279–292; (4) G. Gippert, S. Koide and P.E. Wright, personal communication; (5) van de Kamp, M., Canters, G.W., Wijmenga, S.S., Lommen, A., Hilbers, C.W., Nar, H., Messerschmidt, A. and Huber, R. (1992) *Biochemistry*, **31**, 10194–10207; (6) Kördel, J., Forsén, S. and Chazin, W.J. (1989) *Biochemistry*, **28**, 7065–7074; (7) Drakenberg, T., Hofmann, T. and Chazin, W.J. (1989) *Biochemistry*, **28**, 5946–5954; (8) Gagne, S.M., Tsuda, S., Li, M.X., Chandra, M., Smillie, L.B. and Sykes, B.K. (1994) *Protein Sci.*, **3**, 1961–1974; (9) Ikura, M., Kay, L. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667; (10) Potts, B.C.M., Carlström, G., Okazaki, K., Hidaka, H. and Chazin, W.J. (1996) *Protein Sci.*, **5**, 2162–2174; (11) Amburgey, J.C., Abildgaard, F., Starich, M.R., Shah, S., Hilt, D.C. and Weber, D.J. (1995) *J. Biomol. NMR*, **6**, 171–179, A. Drohat and D. Weber, personal communication; (12) A. Bax, personal communication; (13) Ames, J.B., Tanaka, T., Stryer, L. and Ikura, M. (1994) *Biochemistry*, **33**, 10743–10753; (14) Skelton, N.J., Forsén, S. and Chazin, W.J. (1990) *Biochemistry*, **29**, 5752–5761; (15) Kalbitzer, H., Neidig, K. and Hengstenberg, W. (1991) *Biochemistry*, **30**, 11186–11192; (16) van Nuland, N., Dijk, A.V., Dijkstra, K., Hoesel, F.V., Scheek, R. and Robillard, G. (1992) *Eur. J. Biochem.*, **203**, 483–491; (17) Wittekind, M., Reizer, J. and Klevit, R. (1990) *Biochemistry*, **29**, 7191–7200; (18) Forman-Kay, J., Clore, G.M., Driscoll, P., Wingfield, P., Richards, F. and Gronenborn, A. (1989) *Biochemistry*, **28**, 7088–7097; (19) Dyson, H.J., Holmgren, A. and Wright, P.E. (1989) *Biochemistry*, **17**, 7074–7087; (20) Keller, R.M., Baumann, R., Hunziker-Kwik, E., Joubert, F.J. and Wüthrich, K. (1983) *J. Mol. Biol.*, **163**, 623–646; (21) Wagner, G., Bruhwiler, D. and Wüthrich, K. (1987) *J. Mol. Biol.*, **196**, 227–231; (22) Arseniev, A.S., Wider, G., Joubert, F.J. and Wüthrich, K. (1982) *J. Mol. Biol.*, **159**, 323–351; (23) Redfield, C. and Dobson, C.M. (1990) *Biochemistry*, **29**, 7201–7214; (24) Redfield, C. and Dobson, C.M. (1988) *Biochemistry*, **27**, 122–135; (25) Bartik, K., Dobson, C.M. and Redfield, C. (1993) *Eur. J. Biochem.*, **215**, 255–266; (26) Heitz, A., Chiche, L., Le-Nguyen, D. and Castro, B. (1989) *Biochemistry*, **28**, 2392–2398; (27) Holak, T.A., Gondol, D., Otlewski, J. and Wilusz, T. (1989) *J. Mol. Biol.*, **210**, 635–648; (28) Krishnamoorthi, R., Gong, Y., Lin, C.S. and Van der Velde, D. (1992) *Biochemistry*, **31**, 898–904; (29) Labhardt, A., Hunziker-Kwik, E. and Wüthrich, K. (1988) *Eur. J. Biochem.*, **177**, 295–305; (30) Yu, C., Lee, C., Chuang, L., Shei, Y. and Wang, C. (1990) *Eur. J. Biochem.*, **193**, 789–799; (31) Skelton, N.J., Aspiras, F., Ogez, J. and Schall, T.J. (1995) *Biochemistry*, **34**, 5329–5342; (32) Lodi, P.J., Garrett, D.S., Kuszewski, J., Tsang, M.L.-S., Weatherbee, J.A., Leonard, W.J., Gronenborn, A.M. and Clore, G.M. (1994) *Science*, **263**, 1762–1767; PDB accession code 1hum.mr; (33) Cooke, R., Tappin, M., Campbell, I., Kohda, D., Miyake, T., Fuwa, T., Miyazawa, T. and Inagaki, F. (1990) *Eur. J. Biochem.*, **193**, 807–815; (34) Montelione, G., Wüthrich, K. and Scheraga, H. (1988) *Biochemistry*, **27**, 2235–2243; (35) Selander, M., Persson, E., Stenflo, J. and Drakenberg, T. (1990) *Biochemistry*, **29**, 8111-8118; (36) Huang, L., Cheng, H., Pardi, A., Tam, J. and Sweeney, W. (1991) *Biochemistry*, **30**, 7402–7409; (37) Guiles, R., Altman, J., Lipka, J., Kuntz, I. and Waskell, L. (1990) *Biochemistry*, **29**, 1276–1289; (38) Cai, M., Bradford, E.G. and Timkovich, R. (1992) *Biochemistry*, **31**, 8603–8612; (39) Detlefsen, D., Thanabal, V., Pecoraro, V. and Wagner, G. (1990) *Biochemistry*, **29**, 9377–9486; (40) Marion, D. and Guerlesquin, F. (1992) *Biochemistry*, **31**, 8171–8179; (41) Mortishire-Smith, R.J., Lee, M.S., Bolinger, L. and Wright, P.E. (1992) *FEBS Lett.*, **296**, 11–45; (42) Kochoyan, M., Havel, T.F., Nguyen, D.T., Dahl, C.E., Keutmann, H.T. and Weiss, M.A. (1991) *Biochemistry*, **30**, 3371–3386; (43) Liao, X., Clemens, K., Cavanagh, J., Tennant, L. and Wright, P.E. (1994) *J. Biomol. NMR*, **4**, 433–454; (44) Lee, M.S., Cavanagh, J. and Wright, P.E. (1989) *FEBS Lett.*, **254**, 159–164; (45) Y. Nishimura, personal communication; (46) Carr, M.D., Wollborn, U., McIntosh, P.B., Frenkiel, T.A., McCormick, J.E., Bauer, C.J., Klempnauer, K.-H. and Feeney, J. (1996) *Eur. J. Biochem.*, **235**, 721–735; (47) Yu, H., Rosen, M. and Schreiber, S. (1993) *FEBS Lett.*, **324**, 87–92; (48) D. Kohda and F. Inagaki, personal communication; (49) Zhang, O., Kay, L.E. Oliver, J.P. and Forman-Kay, J.D. (1994) *J. Biomol. NMR*, **4**, 845–858; (50) Hensmann, M., Booker, G.W., Panayotou, G., Boyd, J., Linacre, J., Waterfield, M. and Campbell, I.D. (1994) *Protein Sci.*, **3**, 1020–1030; (51) A. Singer and J. Forman-Kay, personal communication.

Table 2. Chemical shift homology among EF-hand calcium-binding proteins

	Pairs of homologous proteins[a]	Number of residues[a]	% sequence identity	Number of residues compared[b]	Chemical shift[c] rmsd	% shifts[c] within 0.3 ppm	% shifts[c] within 0.1 ppm	Reference[d]
38	S100β, bovine, apo	91	95.6	85	0.23	90.6	77.6	52
	S100β, rat, apo	91						11
5	Calbindin D_{9k}, bovine, $(Ca^{2+})_2$	76	88.2	76	0.09	98.7	90.8	B,6
	Calbindin D_{9k}, porcine, $(Ca^{2+})_2$	78						B,7
39	Calmodulin, apo (N-term.)	148	65.0	80	0.13	93.8	67.5	12
	N-domain of troponin C, apo	90						8
6	Calmodulin, $(Ca^{2+})_2$ (N-term.)	148	65.0	80	0.14	96.3	57.5	B,9
	N-domain of troponin C, $(Ca^{2+})_2$	90						8
7	Calcyclin, rabbit, apo	90	40.0	86	0.31	70.9	31.4	10
	S100β, rat, apo	91						11
40	Calcyclin, rabbit, apo	90	37.7	76	0.26	75.0	35.5	10
	Calbindin D_{9k}, bovine, apo	76						B,14
41	Calbindin D_{9k}, bovine, apo	76	36.7	74	0.36	68.9	27.0	B,14
	S100β, bovine, apo	91						52
42	Calmodulin, apo (C-term.)	148	27.0	69	0.32	71.0	30.4	12
	Calbindin D_{9k}, bovine, apo	76						B,14
43	Calmodulin, $(Ca^{2+})_2$ (C-term.)	148	27.0	70	0.33	71.4	27.1	B,9
	Calbindin D_{9k}, bovine, $(Ca^{2+})_2$	76						B,6
44	Calbindin D_{9k}, bovine, apo	76	26.3	72	0.32	72.2	26.4	B,14
	N-domain of troponin C, apo	90						8
45	Calbindin D_{9k}, bovine, $(Ca^{2+})_2$	76	26.3	72	0.36	65.3	31.9	B,6
	N-domain of troponin C, $(Ca^{2+})_2$	90						8
46	Calmodulin, apo (N-term.)	148	21.1	72	0.30	69.4	36.1	12
	Calbindin D_{9k}, bovine, apo	76						B,14
47	Calmodulin, $(Ca^{2+})_2$ (N-term.)	148	21.1	72	0.36	68.1	29.2	B,9
	Calbindin D_{9k}, bovine, $(Ca^{2+})_2$	76						B,6
48	Calcyclin, rabbit, apo	90	18.3	77	0.31	70.1	26.0	10
	Calmodulin, apo (N-term.)	148						12
8	Calcyclin, rabbit, apo	90	18.4	70	0.33	62.9	28.6	10
	Calmodulin, apo (C-term.)	148						12
9	Recoverin, apo (C-term.)	202	17.0	74	0.38	71.6	32.4	13
	Calbindin D_{9k}, bovine, apo	76						B,14
49	Recoverin, $(Ca^{2+})_2$ (C-term.)	202	17.0	72	0.40	69.4	19.4	53
	Calbindin D_{9k}, bovine, $(Ca^{2+})_2$	76						B,6

Table 2. Continued

	Pairs of homologous proteins[a]	Number of residues[a]	% sequence identity	Number of residues compared[b]	Chemical shift[c] rmsd	% shifts[c] within 0.3 ppm	% shifts[c] within 0.1 ppm	Reference[d]
50	Recoverin, apo (N-term.)	202	15.8	73	0.33	65.7	35.6	13
	Calbindin D_{9k}, bovine, apo	76						B,14
51	Recoverin, $(Ca^{2+})_2$ (N-term.)	202	15.8	72	0.41	59.7	22.2	53
	Calbindin D_{9k}, bovine, $(Ca^{2+})_2$	76						B,6

[a] The entry for 'Number of residues' corresponds to the construct used to acquire the NMR data. When a domain from a larger protein was used in our analysis, the domain was identified in parenthesis next to the name of the protein. For example, in comparison #6 the isolated N-domain of troponin C was expressed and used for NMR analysis, whereas the data for the N-terminal domain of calmodulin were obtained from an analysis of the intact protein.
[b] The number of C^{α} proton resonances compared is limited by insertions/gaps in the sequence alignment and/or unassigned resonances.
[c] Secondary chemical shifts of C^{α} protons.
[d] References: B, accessed entry in the BioMagResBank; (52) Kilby, P.M., Eldik, L.J.V. and Roberts, G.C.K. (1995) *FEBS Lett.*, **363**, 90–96; (53) T. Tanaka and M. Ikura, personal communication. All the other references are from the list provided in footnote d of Table 1.

for the expected level of chemical shift similarity for homologous proteins (chemical shift homology). At the outset, it is critical to recognize the inherent problems associated with the study of protein chemical shifts. These include the following:

(1) The resonance assignments of proteins are made under conditions in which the protein is 'well behaved' in solution, and these conditions will vary from protein to protein (and laboratory to laboratory). In particular, backbone amide protons are notorious for their sensitivity to variations in pH and temperature.

(2) The chemical shifts are not always reported with respect to the same reference standard.

(3) The resonance assignments may be incomplete.

(4) There are a limited number of homologous proteins for which complete chemical shift assignments are available, particularly for ^{15}N and ^{13}C nuclei.

Furthermore, insertion of a gap may be required for the alignment of sequences, making it impossible to use all of the resonance assignments in the comparison. And of course, when comparing residues that are not identical, there are inherent differences in chemical shifts due to the difference in the side chain.

In order to overcome as many of these difficulties as possible, this study was limited to the backbone $C^{\alpha}H$ resonances, every effort was made to correct the chemical shift data to a standard reference compound (Wishart et al., 1995), and the secondary chemical shift (observed chemical shift minus the random coil chemical shift) was employed in the comparison. To remove the uncertainties associated with scoring sequence homology, percentage sequence identity has been utilized as a means to classify the sequence similarity between each protein pair.

Chemical shift similarities

The investigation of chemical shift homology employed 52 proteins in 51 pairwise combinations, 37 were used for the general comparative analysis (Table 1) and 19 for the family-specific analysis of EF-hand calcium-binding proteins (Table 2). Although some specific comparisons of highly homologous or mutant proteins are well documented in the literature (Wittekind et al., 1989; Redfield and Dobson, 1990; Carlström and Chazin, 1993; Potts et al., 1996), we have attempted to generalize the comparison by using proteins representing a variety of families and folds and spanning a wide range of sequence homologies. Within this database of pairs of homologous proteins, the sequence identities ranged from 16 to 97%.

Figure 1 shows a plot of the percentage sequence identity versus chemical shift homology for the 37 diverse protein pairs in Table 1, quantified in terms of the rmsd between the $C^{\alpha}H$ secondary chemical shifts. The rmsd values range from 0.06 to 0.67 ppm and the mean value for this set of comparisons is 0.28 ppm. The correlation between sequence identity and chemical shift homology is reflected in a Spearman rank-order correlation coefficient (Press et al., 1986) of 0.79 for these two parameters. Overall, the data confirm that the chemical shifts of structurally homologous proteins are similar.

In order to provide a more practical guideline for estimating the extent of chemical shift similarity expected for structurally homologous proteins, the

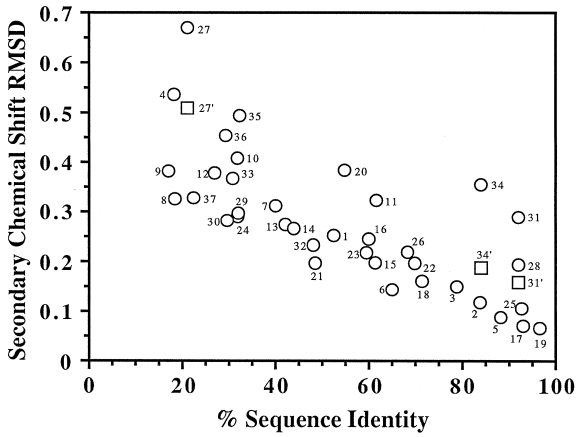

Figure 1. The secondary chemical shift rmsd for C^{α} protons plotted against the percentage of amino acid sequence identity (open circles) for the homologous protein pairs in Table 1. The labels correspond to the numbers in Table 1. Open squares show revised rmsd values for three protein pairs (#27, #31, and #34) after large chemical shift differences are omitted from the calculation, as described in the text.

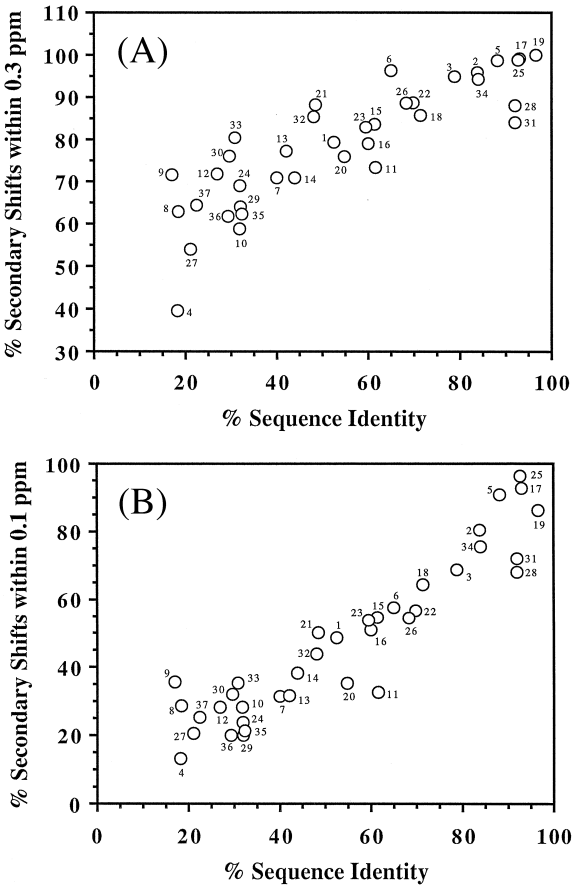

Figure 2. The percentage of C^{α} proton secondary chemical shifts within a 0.3 ppm (A) or 0.1 ppm (B) threshold, plotted against the percentage of amino acid sequence identity for the homologous protein pairs in Table 1. The labels correspond to the numbers in Table 1.

percentage of $C^{\alpha}H$ secondary chemical shifts that fall within a given threshold have been examined relative to the degree of sequence identity. A plot of this relationship is shown in Figure 2A, using a threshold of 0.3 ppm. For all protein pairs in this database with greater than 20% sequence identity, over 50% of the $C^{\alpha}H$ secondary chemical shifts fall within this 0.3 ppm threshold. The percentage rises to >80% when comparing proteins with greater than 65% sequence identity. This analysis provides a strong indication of the correlation between chemical shift and sequence homologies, with a Spearman rank-order correlation coefficient of 0.87 for the full data set in Figure 2A.

This data set has also been examined using a threshold of 0.1 ppm, which is likely to be more informative for using chemical shift homology to assist in making sequence-specific assignments. The correlation between sequence identity and chemical shift homology using this lower threshold is surprisingly good (Figure 2B). There is an approximately linear relationship between the percentage of sequence identity and the percentage of the $C^{\alpha}H$ secondary shifts that fall within the 0.1 ppm threshold, with a correlation coefficient of 0.97.

It is important to emphasize that the data shown in Figures 1 and 2 are drawn from a database of structurally homologous proteins. The results are not intended to suggest that any random pair of peptide sequences will have levels of chemical shift homology similar to those in this database. A preliminary comparison of peptides derived from unrelated proteins with sequence similarity at a local level indicated that sequence and chemical shift homology were not correlated. The mean secondary chemical shift rmsd for these data (0.72 ppm) is more than twofold higher than the value (0.28 ppm) obtained for the database of homologous proteins used in this analysis.

Outliers in the chemical shift homology plots

There are several notable examples of homologous protein pairs with secondary chemical shift rmsd values that fall outside the bulk of the data (Figure 1), including pairs of cytochrome c's (pair #27, Table 1), zinc fingers (pair #31), and Myb transcription factors (pair #34). In each case, the discrepancy is largely associated with C^{α} protons that have large ring current effects arising from spatially adjacent residues.

In the case of cytochromes c551 (*P. stutzeri*) and c553 (*D. vulgaris*), the amino acid sequences were aligned based on their 3D structures, yielding 21% sequence identity and an rmsd of 0.67 ppm for the $C^{\alpha}H$ secondary chemical shift. Although the two proteins share a common 3D folding topology, there are significant structural differences. The regular helix containing the methionine ligand to the heme iron in cytochrome c553 is a polyproline helix in cytochrome c551, and one of the helical elements present in cytochrome c551 is absent in cytochrome c553. In addition, all of the loop motifs are of different lengths in the two proteins. These structural differences are correlated with large differences in certain $C^{\alpha}H$ secondary chemical shifts, which occur when there is a very large ring current effect in only one of the two proteins. There are three particularly extreme examples. In cytochrome c551, Gly^{24} in the Ω loop and Ile^{48} within a helix resonate at particularly high field (0.08 and 1.49 ppm, respectively). In contrast, none of the C^{α} protons of the Ω loop or the corresponding helix of cytochrome c553 have such unusual high-field chemical shifts. In cytochrome c553, the $C^{\alpha}H$ chemical shift of the methionine ligand to the heme (Met^{57}) resonates at very high field (1.77 ppm), whereas the chemical shift of the corresponding residue (Met^{61}) in cytochrome c551 is only slightly upfield shifted (3.65 ppm). Each of these cases of very large chemical shift differences likely reflect the relative positioning of the corresponding residues with respect to the heme prosthetic group, which has a very large ring current. Since the secondary chemical shift differences for these residues are so much greater than all others, they dominate the rmsd. When these three outliers are omitted from the comparison, the chemical shift rmsd drops from 0.67 to 0.51 ppm (Figure 1), a value more consistent with the results obtained for other proteins.

A comparison of the DNA binding domains of two members of the Myb transcription factor family also yields an anomalous result on first inspection. Myb DNA binding domains consist of three homologous, tandem repeats termed R1, R2, and R3. A comparison between the individual repeats of c-Myb (i.e., R1 versus R2, R2 versus R3) results in secondary chemical shift rmsd values that are consistent with those of the other protein pairs in Figure 1 (#32, #33). However, a comparison of tandem repeats from the two different Myb proteins, c-Myb-R2R3 and b-Myb-R2R3 (#34), reveals a secondary chemical shift rmsd of 0.36 ppm even though they have a high degree of sequence identity (84%). Since 76% of the secondary shifts fall within 0.1 ppm of one another (Figure 2B), this high rmsd indicates the presence of a limited number of chemical shifts in very large disagreement. Each repeat of c-Myb contains one residue with a high-field shifted C^{α} proton resonance ($\delta < 2$ ppm) (Ogata et al., 1992; Jamin et al., 1993; Y. Nishimura, personal communication). Analysis of the solution structures for the c-Myb individual repeats R1, R2, and R3 indicates that these extraordinary shifts are correlated with a close proximity to the aromatic ring of a tryptophan residue. The tryptophan is conserved in all of these Myb domains; hence, the large ring current effect observed in the c-Myb domains is either not present in b-Myb or the assignments are in error. When the chemical shifts for these protons are not included in the chemical shift comparison, the rmsd drops to 0.19 ppm (Figure 1), in line with the data for the other proteins in the database.

The secondary $C^{\alpha}H$ chemical shift rmsd values for four pairs of zinc finger motifs are included in Figure 1. Chemical shift similarity among zinc finger modules has been discussed previously, and the $C^{\alpha}H$ chemical shift of the conserved leucine residue, Leu^{18}, was shown to be correlated with the presence or absence of an aromatic residue in the small hydrophobic core (Lee et al., 1992; Mortishire-Smith et al., 1992). Our analysis shows that Xfin31 and Xfin31B are 92% identical in sequence but have a relatively high secondary chemical shift rmsd (0.29 ppm). The substitution of Phe^{12} for glycine in Xfin31B removes the ring current on Leu^{18} and, correspondingly, the $C^{\alpha}H$ resonance is shifted downfield by 1.2 ppm relative to wild-type protein. With only 25 amino acid residues per peptide in the comparison, this single large difference among the $C^{\alpha}H$ chemical shifts results in the rather high chemical shift rmsd; omission of this data point brings the rmsd down to 0.16 ppm (Figure 1). In all of the other zinc finger comparisons included in the database (Table 1), residue 12 is either aromatic in both peptides or is mutated in both peptides. As a result, the C^{α} proton chemical shifts of Leu^{18} are comparable in each case, and the resulting chemical shift rmsd values are in agreement with the consensus in Figure 1.

The other cases of relatively large secondary chemical shift rmsd values occur for protein pairs that appear to approach the lower limit of structural homology that will give rise to chemical shift homology. For example, consider the blue copper proteins azurin and plastocyanin, type I cuprodoxins from classes I and III, respectively, that both adopt a Greek key β barrel topology (Adman, 1991). In comparing the homolo-

gies of these two proteins, their sequence identity is 18% while the chemical shift rmsd is 0.54 ppm. The structural differences between these two proteins include differences in the geometry and number of the ligands about the copper center (Adman, 1991; Kofman et al., 1996) and a variable polypeptide strand outside of the β barrel (Adman, 1991; Sykes, 1991). The variable strand consists of an ∼30-residue 'flap' in azurin as opposed to ∼20 residues that contain the 'acidic patch' in plastocyanin, and there is no correspondence in the 3D structures. Although plastocyanin and azurin do have some large specific chemical shift differences (e.g., copper ligands His^{37} (δ 5.93) versus His^{46} (δ 7.25); Phe^{14} (δ 4.98) versus Phe^{15} (δ 6.20); Asn^{31} (δ 5.02) versus His^{35} (δ 6.51) in plastocyanin and azurin, respectively), only 40% of the chemical shifts fall within the 0.3 ppm threshold, and only 13% within the 0.1 ppm threshold (Figure 2, #4). Thus, the high chemical shift rmsd is not solely related to the few isolated cases of large chemical shift differences. This level of chemical shift divergence appears to mark the point at which the chemical shift homology concept begins to break down due to an insufficient degree of structural homology. These results suggest that the ability to make resonance assignments by comparison will be restricted to instances where the structural homology is extensive. Further analysis will be required to determine whether specific sequence and structural homology thresholds can be defined.

The examples discussed above illustrate some potential difficulties that may be encountered in chemical shift homology analysis and resonance assignment by comparison. The results also imply that the threshold-based comparisons (Figure 2) provide a better measure of overall chemical shift homology than the secondary chemical shift rmsd. Since the rmsd measurement includes all possible proton pairs, isolated cases of large Δδs can result in a large chemical shift rmsd, which might be misinterpreted as an indicator of lower than expected structural homology. When a high percentage of resonances fall within the 0.3 ppm threshold but the chemical shift rmsd is relatively high, overall structural homology is likely but with some important local structural perturbations. Consequently, a combination of positive indications of chemical shift homology from these two parameters provides the strongest evidence of structural homology.

Chemical shift homology in EF-hand calcium-binding proteins

EF-hand calcium-binding proteins represent one of the largest families of homologous proteins that have been characterized by solution state NMR spectroscopy. With the abundance of spectroscopic and structural information available, this family is a prime candidate for detailed studies of chemical shift homology in proteins. Consequently, 19 chemical shift comparisons were made for EF-hand calcium-binding proteins (Table 2), including five pairs from Table 1. This secondary database includes examples of (i) the same protein from two species (#5, bovine versus porcine calbindin D_{9k}; #38, bovine versus rat S100β); (ii) proteins within a subfamily, including members of the S100 subfamily (#40, calcyclin versus calbindin D_{9k}; #41, S100β versus calbindin D_{9k}; #7, calcyclin versus S100β) and the calmodulin subfamily (#6, calmodulin versus troponin C); and (iii) comparisons across subfamilies but within the same superfamily (e.g., #42, calbindin D_{9k} versus calmodulin). Thus, a wide range of sequence homologies, from 16% to 96% amino acid sequence identity, is included in this analysis.

The $C^{\alpha}H$ secondary chemical shift rmsd values for the EF-hand calcium-binding proteins range from 0.09 to 0.41 ppm, with a mean value of 0.30 ppm. This value is similar to the mean value of 0.28 ppm obtained for all protein pairs in Table 1, even though the EF-hand calcium-binding protein data are weighted toward low sequence homology, with 15 of the 19 protein pairs exhibiting sequence identities of $<40\%$. Among the four comparisons of proteins with $>65\%$ sequence identity, $>90\%$ of the $C^{\alpha}H$ chemical shifts fall within the 0.3 ppm threshold and $>50\%$ within the 0.1 ppm threshold. Among the 15 other protein pairs with 16-40% sequence identity, $>60\%$ of the $C^{\alpha}H$ chemical shifts fall within the 0.3 ppm threshold and $>19\%$ within the 0.1 ppm threshold. Together, these results show a very high degree of chemical shift homology among the proteins within this family, which is presumably a reflection of their close correspondence in both secondary and tertiary structure.

Chemical shift assignment by comparison

The analysis of chemical shift homology can be used to assess the prospects for making chemical shift assignments by comparison to a highly homologous protein which is already assigned. In this context, the results obtained using the 0.1 ppm threshold are most informative, as the looser 0.3 ppm threshold is not particularly discriminating. For example, the two

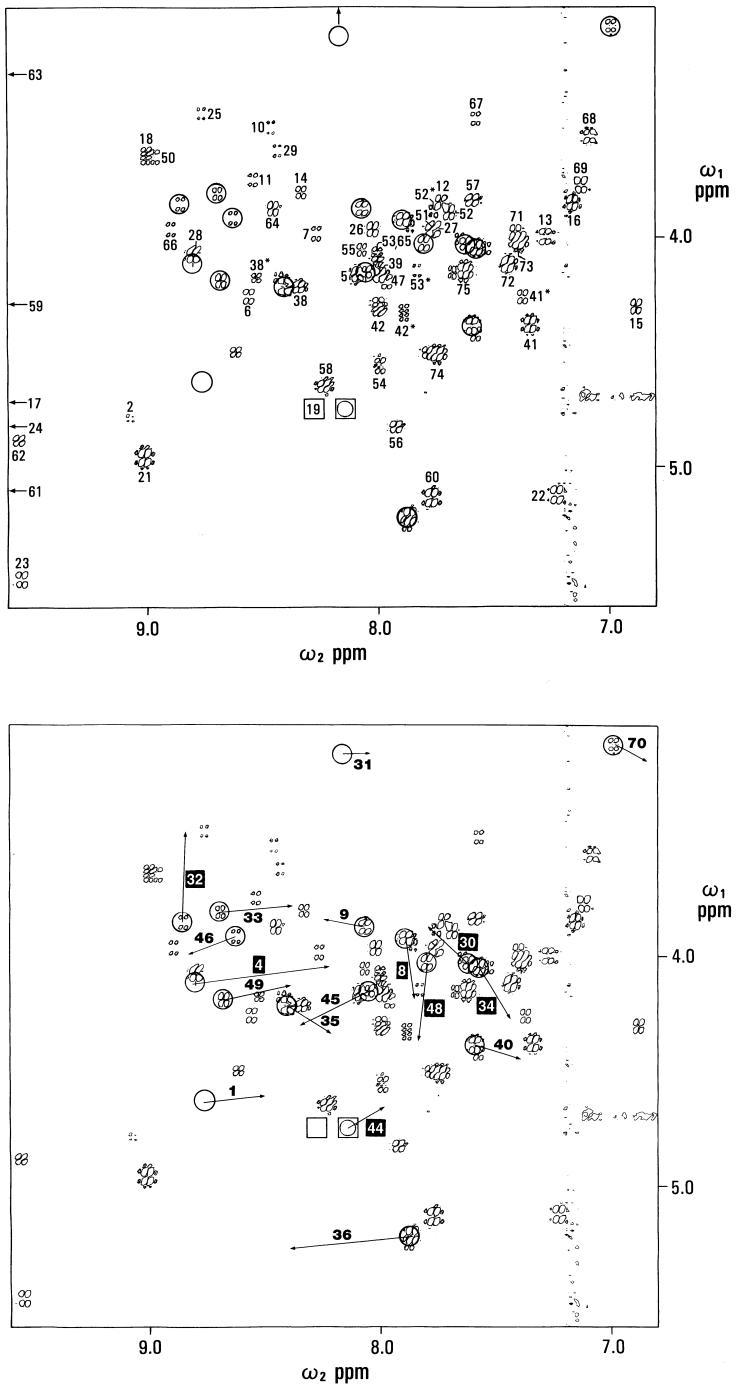

Figure 3. Comparison of backbone amide fingerprint of Ca^{2+}-loaded bovine and porcine calbindin D_{9k} at pH 6.0, 300 K. In both the upper and lower panels, the COSY spectrum of the bovine protein is plotted. In the upper panel, the residue numbers are marked for each porcine cross peak whose backbone amide and C^{α} resonances fall within the 0.10 ppm thresholds. Circles are drawn around the cross peaks for the 18 residues in the porcine protein that do not fall within the thresholds, and their location in the porcine protein is shown in the lower panel. These include residues Lys^{1}, $\underline{Glu}^{4}$, $\underline{Gly}^{8}$, Ile^{9}, $\underline{Leu}^{30}$, Leu^{31}, $\underline{Leu}^{32}$, Gln^{33}, $\underline{Thr}^{34}$, Glu^{35}, Phe^{36}, Leu^{40}, $\underline{Ser}^{44}$, Thr^{45}, Leu^{46}, $\underline{Glu}^{48}$, Leu^{49}, and Val^{70}. The seven amino acid differences between the bovine and porcine proteins are underlined and highlighted in the lower panel by white on black labeling. Arrows pointing to the edges of the plot indicate that the cross peak is found outside the region shown. Boxes are drawn around the location of two cross peaks that are not observed in this spectrum due to presaturation at the solvent frequency. The peaks marked with $*$ arise from the minor *cis*-Pro ^{43} isoform, and are observed for both the bovine and the porcine proteins.

calbindin D_{9k} proteins (comparison #5) have 88% sequence identity and exhibit 98% and 91% chemical shift similarities at the 0.3 and 0.1 ppm thresholds, respectively. The N-terminal domains of calmodulin and troponin C (#39) have only 65% sequence identity but exhibit almost the same chemical shift similarity (94%) as the more homologous pair at the 0.3 ppm threshold, but only 68% at the 0.1 ppm threshold. When the sequence identity drops to 38% as in the comparison of calbindin D_{9k} and calcyclin (#40), the chemical shift similarities remain relatively high (75%) at the 0.3 ppm threshold, but drop to 36% at the 0.1 ppm threshold. These trends have been observed not only for EF-hand calcium-binding proteins but also within other families of homologous proteins (see Table 1). Our results suggest that only protein pairs with very high sequence homology give rise to a sufficient percentage of secondary chemical shifts within the 0.1 ppm threshold to permit assignment by comparison.

In practical terms, assignment by comparison cannot rely on a single chemical shift parameter, and would ordinarily be carried out by examining correlations of two (or more) resonances directly on multidimensional spectra. An example of this approach is provided in Figure 3, which shows the backbone fingerprint region of the COSY spectrum of bovine calbindin D_{9k}, and indicates the location of the corresponding cross peaks for the porcine protein. In this case of very high sequence homology, the approach is seen to be quite successful. Of the 75 cross peaks expected for calbindin D_{9k}, 57 of the porcine cross peaks could be accurately assigned by comparison because they coincide with the bovine cross peak to within 0.1 ppm in both NH and $C^{\alpha}H$ frequency (Figure 3, upper panel). The lower panel in Figure 3 shows the changes in chemical shifts for the 18 other residues, seven of which correspond to the seven amino acid substitutions between the two proteins. Since the amide frequencies are so sensitive to experimental conditions and C^{α} protons are notorious for their spectral degeneracy, this example constitutes a stringent test of the ability to transfer assignments from one protein to the other. An attempt at a similar analysis of the backbone fingerprint of the COSY spectra of calbindin D_{9k} and calcyclin (38% sequence identity) was unsuccessful, illustrating the difficulties associated with assignment by comparison for protein pairs with lower sequence homology (data not shown). One major factor contributing to the lack of success is that assignment by comparison of spectra requires the use of the primary, not the secondary, chemical shift. Consequently, the greatly increased number of amino acid substitutions makes direct spectral comparisons impractical. Large differences in the backbone amide proton chemical shifts were particularly problematical. Clearly, many more examples need to be investigated in order to establish the lower limit on sequence homology for which such approaches will be valid. It is also important to note that these assignments serve to facilitate, but not substitute for, the sequential resonance assignment procedures. We expect that inclusion of more than one nucleus will facilitate and extend the range of assignment by comparison.

Conclusions

Our empirical, comparative analysis is intended to serve as a practical guideline for estimating the chemical shift similarities for structurally homologous proteins. Clearly, proteins with very high sequence homologies will have very similar structures, which implies similar chemical shifts. In our database, homologous proteins with $>80\%$ sequence identity have $>80\%$ of their corresponding $C^{\alpha}H$ secondary chemical shifts resonating within 0.3 ppm of each other and $>65\%$ within a 0.1 ppm threshold. For proteins known to have very similar structures such as the EF-hand calcium-binding proteins, the chemical shift homology is even greater.

The correlation between chemical shift similarity and amino acid sequence identity demonstrated in this study verifies the existence of a relationship that most biomolecular NMR practitioners have come to expect based on their own experiences and published reports. This relationship might be viewed by treating the protein backbone as a scaffold for the amino acid side chains. As the side chains are varied via amino acid substitution, the resulting change in the electronic environment perturbs the chemical shift parameter; the greater the number of amino acid substitutions, the greater the observed chemical shift differences. In our analysis, we have corrected for amino acid substitutions by employing the secondary chemical shift, as this option would certainly be available given that the amino acid sequence is known and can be aligned with that of the homologous protein. No prior knowledge of the structure is necessary. Of course, additional factors come into play. Ring currents are one obvious source of chemical shift differences, and we have described specific examples of dramatic ring current

effects for isolated amino acids. We expect that correcting for ring current effects would decrease the slope of the curves shown in Figure 2, i.e., make the chemical shifts more similar over a range of sequence identities. However, our empirical study has been designed from the viewpoint of the structural biologist confronted with a new protein, so ring current calculations have not been included in this analysis because these require accurate 3D structures, which would rarely be available in the context of obtaining resonance assignments for a homologue.

Considering the results presented herein, it is conceivable that one could even push the process further by combining the chemical shift data obtained from two homologous proteins with a structure 'back-calculation'. If amino acid sequence comparison suggests that two proteins will adopt a similar global folding pattern, one might build a homology model of the previously studied protein, then start with the chemical shift values of the previously assigned protein, correct for secondary chemical shift, and subsequently refine against the ring currents of the new protein. Clearly, as the number of proteins with complete resonance assignments increases, and advances in *ab initio* calculations of chemical shift terms are made, we can look forward to significant advances in the use of chemical shift data to expedite the resonance assignment process for new proteins and to obtain direct structural information about proteins in solution.

Acknowledgements

We are indebted to many individuals for their generous contributions of chemical shift data, including Drs. Ad Bax, Alex Drohat, Julie Forman-Kay, Stephane Gagne, Garry Gippert, Mitsu Ikura, Fuyuhiko Inagaki, Shohei Koide, Daisuke Kohda, Yoshifumi Nishimura, Steve Pascal, Nick Skelton, Alex Singer, Brian Sykes, Toshiyuki Tanaka, and David Weber. We also thank Dr. Garry Gippert for helpful discussions. This work was supported by operating grants from the National Institutes of Health (GM PO1 48495 and RO1 40120) and fellowships from the American Cancer Society (PF-3934 to B.C.M.P. and FRA-436 to W.J.C.).

References

Adman, E.T. (1991) *Adv. Protein Chem.*, **42**, 145–197.
Carlström, G. and Chazin, W.J. (1993) *J. Mol. Biol.*, **227**, 415–430.
Eklund, H., Gleason, F.K. and Holmgren, A. (1991) *Proteins Struct. Funct. Genet.*, **11**, 13–28.
Gippert, G.P. (1995) Ph.D. Thesis, The Scripps Research Institute, La Jolla, CA, U.S.A.
Gronenborn, A.M. and Clore, G.M. (1994) *J. Biomol. NMR*, **4**, 455–458.
Jamin, N., Gabrielsen, O.S., Gilles, N., Lirsac, P. and Toma, F. (1993) *Eur. J. Biochem.*, **216**, 147–154.
Kofman, V., Farver, O., Pecht, I. and Goldfarb, D. (1996) *J. Am. Chem. Soc.*, **118**, 1201–1206.
Kretsinger, R.H., Tolbert, D., Nakayama, S. and Pearson, W. (1991) In *Novel Calcium-Binding Proteins* (Ed., Heizmann, C.W.), Springer, New York, NY, U.S.A., pp. 17–37.
Kuriyan, J. and Cowburn, D. (1993) *Curr. Opin. Struct. Biol.*, **3**, 828–837.
Kuszewski, J., Qin, J., Gronenborn, A.M. and Clore, G.M. (1995) *J. Magn. Reson.*, **106**, 92–96.
Lee, M.S., Mortishire-Smith, R.J. and Wright, P.E. (1992) *FEBS Lett.*, **309**, 29–32.
Mortishire-Smith, R.J., Lee, M.S., Bolinger, L. and Wright, P.E. (1992) *FEBS Lett.*, **296**, 11–15.
Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S. and Nishimura, Y. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6428–6432.
Oldfield, E. (1995) *J. Biomol. NMR*, **5**, 217–225.
Ösapay, K., Theriault, Y., Wright, P.E. and Case, D.A. (1994) *J. Mol. Biol.*, **244**, 183–197.
Potts, B.C., Carlström, G., Okazaki, K., Hidaka, H. and Chazin, W.J. (1996) *Protein Sci.*, **5**, 2162–2174.
Press, W., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986) *Numerical Recipes*, Cambridge University Press, New York, NY.
Redfield, C. and Dobson, C.M. (1990) *Biochemistry*, **29**, 7201–7214.
Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J.L. (1991) *J. Biomol. NMR*, **1**, 217–236.
Skelton, N.J., Aspiras, F., Ogez, J. and Schall, T.J. (1995) *Biochemistry*, **34**, 5329–5342.
Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.
Sykes, A.G. (1991) *Adv. Inorg. Chem.*, **36**, 377–408.
Szilagyi, L. (1995) *Prog. NMR Spectrosc.*, **27**, 325–443.
Williamson, M.P., Kikuchi, J. and Tetsuo, A. (1995) *J. Mol. Biol.*, **247**, 541–546.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.
Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.
Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.
Wittekind, M., Reizer, J., Deutscher, J., Saier, M.H. and Klevit, R.E. (1989) *Biochemistry*, **28**, 9908–9912.