Structural comparisons among the short-chain helical cytokines

Denise A Rozwarski[1], Angela M Gronenborn[2], G Marius Clore[2], J Fernando Bazan[3], Andrew Bohm[4], Alexander Wlodawer[5], Marcos Hatada[6] and P Andrew Karplus[1]*

[1]Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA, [2]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD 20892, USA, [3]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448, USA, [4]Calvin Lab, University of California, Berkeley, CA 94720, USA, [5]Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, MD 21702-1201, USA and [6]Ariad Pharmaceuticals, 26 Landsdowne Street, Cambridge, MA 02139, USA

Background: Cytokines and growth factors are soluble proteins that regulate the development and activities of many cell types. One group of these proteins have structures based on a four-helix bundle, though this similarity is not apparent from amino acid sequence comparisons. An understanding of how diverse sequences can adopt the same fold would be useful for recognizing and aligning distant homologs and for applying structural information gained from one protein to other sequences.

Results: We have approached this problem by comparing the five known structures which adopt a granulocyte-macrophage colony-stimulating factor (GM-CSF)-like, or short-chain fold: interleukin (IL)-4, GM-CSF, IL-2, IL-5, and macrophage colony-stimulating factor. The comparison reveals a common structural framework of five segments including 31 inner-core and 30 largely exposed residues. Buried polar interactions found in each protein illustrate how complementary substitutions maintain protein stability and may help specify unique core packing. A profile based on the known structures is not sufficient to guarantee accurate amino acid sequence alignments with other family members. Comparisons of the conserved short-chain framework with growth hormone define the optimal structural alignment.

Conclusions: Our results are useful for extrapolating functional results among the short-chain cytokines and growth hormone, and provide a foundation for similar characterization of other subfamilies. These results also show that the placement of polar residues at different buried positions in each protein complicates sequence comparisons, and they document a challenging test case for methods aimed at recognizing and aligning distant homologs.

Structure 15 March 1994, **2**:159–173

Key words: cytokine, helical bundle, homology modeling, protein evolution, protein folds

Introduction

Cytokines are a group of proteinaceous intercellular messengers involved in the development, activation, and regulation of cells of the circulatory system [1]. Structural studies of these molecules are progressing rapidly and have already revealed that the cytokines can be grouped into a number of different structural families. These include the interleukin (IL)-1-like β-trefoil family [2], the IL-8-like family [3,4], and the growth hormone-like helical bundle family [5,6]. As shown in Fig. 1, members of the helical bundle family contain helices arranged in an up-up-down-down topology, which does not exist in any other known protein structures.

Tertiary structures have been determined for nine members of the helical bundle family of cytokines, and they show that the family can be further divided into three apparent groups which we will refer to as the short-chain (five structures known), the long-chain (two structures known) and the interferon-like (two structures known) subfamilies. Although the helical bundle topology is the same within these three subfamilies, the folds are sufficiently different that successful homology model building applications would require that the folds be treated as distinct. As demonstrated in Fig. 1, key features of the long-chain subfamily are a bundle of four well-aligned helices 20 to 30 residues in length and a helix A to helix B crossover passing in front of helix D. The crossover connections include short helices, but these are not well-aligned with or integrated with the bundle core. For the short-chain subfamily members, the helices of the bundle are aligned less well and are only 10 to 20 residues long. Also, the crossover from helix A to helix B passes behind helix D and is part of a short two-stranded antiparallel β-sheet which contributes to the bundle core. The change in the crossover connection suggests there is some difference in the folding pathway of these two subfamilies. The interferons make up the third subfamily, and have features intermediate between the other two. The helix packing angles in these molecules are closer to those found in short-chain cytokines, but their

*Corresponding author.

 ISSN 0969-2126

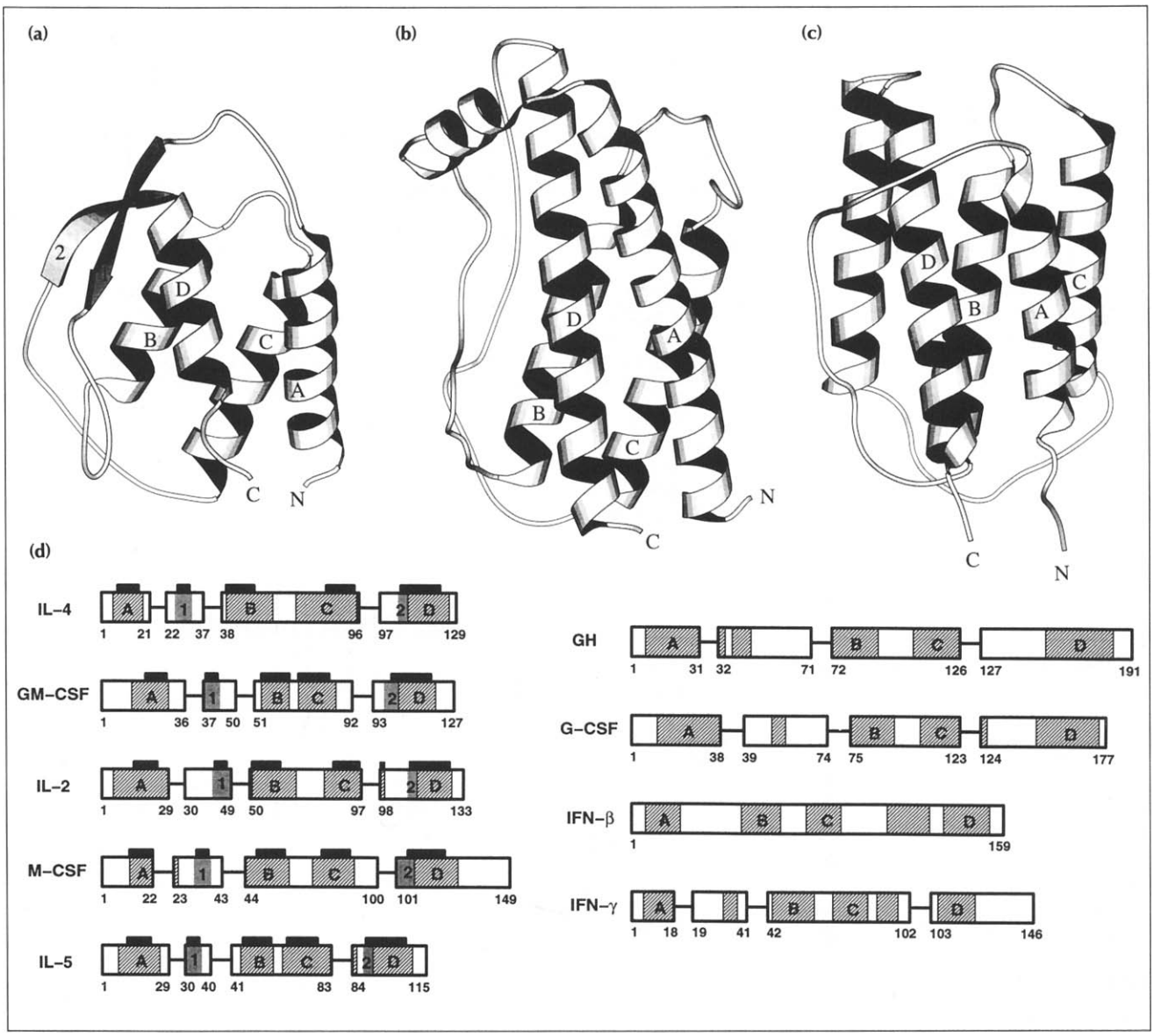

Fig. 1. Ribbon diagrams showing the structural features common to the four α-helix bundle cytokine family. The four helices are designated A, B, C and D. The up-up-down-down topology refers to the fact that helices A and B point up while helices C and D point down. This creates a situation in which each helix is antiparallel to both of its neighboring helices. This topology is also called a double crossover antiparallel helical bundle since there are two long segments after helices A and C which cross from one end of the bundle to the other. **(a)** The fold typical of the short-chain subfamily, whose members are listed in Table 1. Key elements include a two-stranded β-sheet and a crossover connection that passes behind helix D. Generally, these proteins have fewer than 150 residues. **(b)** The fold typical of the long-chain subfamily, whose structurally known members include growth hormone (GH) [19] and granulocyte colony-stimulating factor (G-CSF) [42]. Key elements include no β-strands and a crossover connection that passes in front of helix D. Also, a short helix after αA that is involved in receptor binding and is common to the two known structures is shown. Generally, these proteins are longer than 160 residues. **(c)** The fold typical of the interferon-like subfamily, whose structurally known members include interferon-β (IFN-β) [43] and interferon-γ (IFN-γ) [44,45]. Key elements include a helix in the C–D crossover and an A–B crossover connection that passes in front of helix D. **(d)** Exon structures of the structurally known members of each subfamily (references in the order of appearance: [46–55]). The major helices (striped) and β-strands (shaded) are indicated and labeled as in (a)–(c).

helix A to helix B crossover passes in front of helix D, as it does in the long-chain cytokines. One unique feature of the interferon fold is that the helix C to helix D crossover forms a fifth α-helix which contributes to the bundle core. It should be noted that macrophage colony-stimulating factor (M-CSF) is naturally dimeric, and that IL-5 and interferon-γ, in the short-chain and interferon-like subfamilies respectively, form unusual interdigitating homodimers so that in the prototype domain shown in Fig. 1 the helix C to helix D crossover and the D-helix come from the second chain.

A striking feature of this family is that there is very little amino acid sequence similarity between family members, and even between pairs within the same subfamily the sequences are not recognizably similar. Despite this low level of sequence similarity, a strong case can be made that the members of this family are all related by

Table 1. Members of the short-chain helical cytokine subfamily whose structures are known.

Name (abbreviation)	Chain length	Structure determination[a]	PDB code[b]	Reference
Interleukin-4 (IL-4)	129	NMR (8.3, 1.0 Å)	1bbn	[8]
		NMR (14.1, 1.0 Å)	1itl	[9]
		X-ray (2.35 Å, 23 %)		[10]
		X-ray (2.25 Å, 22 %)	1rcb	[11]
Granulocyte-macrophage colony-stimulating factor (GM-CSF)	127	X-ray (2.4 Å, 20 %)	1gmf	[12]
		X-ray (2.8 Å, 25 %)	1rgm	[13]
Interleukin-2 (IL-2)	133	X-ray (2.5 Å, 20 %)	3ink	[14]
		NMR		[15]
		X-ray (2.0 Å, 19 %)		MH, unpublished data
Macrophage colony-stimulating factor (M-CSF)[c]	158 × 2	X-ray (2.5 Å, 20 %)	1hmc	[16]
Interleukin-5 (IL-5)	115 × 2	X-ray (2.4 Å, 21 %)		[17]

[a]Indicates the method by which the structures were solved. For structures determined by X-ray crystallography (X-ray), the resolution of the analysis (in Å) and R-factor (%) is given. For the structures determined by multidimensional NMR spectroscopy (NMR), the number of constraints per residue and the estimated precision of the main chain atoms (Å) is given. [b]Refers to the access code to obtain the coordinates from the Brookhaven protein databank [57]. [c]M-CSF here refers to M-CSFα, which is a shorter version of M-CSFβ and M-CSFγ.

divergent evolution: they are functionally similar, all being extracellular signaling molecules and, for the most part, binding to homologous receptors [5,7]; they are structurally similar, all having a helical bundle of unique topology; and they are genetically similar [7], most having separate exons encoding the following structural segments: helix A; the A–B crossover; helices B and C and part of the C–D crossover; and the rest of the C–D crossover and helix D (Fig. 1d). The striking similarity in exon structure is consistent with the hypothesis that all of the members of this family have diverged from an ancestral protein in which this exon structure was established; differences in individual family members would then be due to insertions or deletions occurring in the individual exons or, in the case of interferon-β, intron loss.

The short-chain subfamily is thought to include at least nine members (Table 1). The structures of five of them have been determined; these are IL-4 [8–11], granulocyte-macrophage colony-stimulating factor (GM-CSF) [12,13], IL-2 ([14,15] and Marcos Hatada, unpublished data), M-CSF [16] and recently IL-5 [17]. Their complete amino acid sequences and secondary structures are shown in Fig. 2. Understanding how these molecules bind to their receptors and facilitate signal transmission has been a primary motivation for the structural studies, and valuable insights have been obtained [18]. Especially important in this regard is the structure of human growth hormone in complex with its receptor which has provided a paradigm for cytokine–receptor interactions in the helical bundle family [19,20]. Comparisons of these structures are important to assess how structural and functional information from these family members can be extrapolated to other, unknown, structures. In addition, because of the extensive differences in the amino acid sequences, this family is a fertile subject for the study of protein folding and evolution as well as a testing ground for sequence alignment and homology model building methods. We present here a comparison of the members of the short-chain subfamily whose structures are known, to assess which features of the fold are conserved and to build a foundation on which comparisons between all helical bundle cytokines can be made.

Results

Family consensus framework

The short chain helical cytokines contain six common elements of secondary structure: four α-helices and two strands of a β-sheet (Figs 1 and 2). Since β-strand 2 is contiguous with α-helix D, it was here treated as a single segment. The conserved juncture between these two structural elements provided an important tether point for the overlays by allowing every α-helix D to be aligned unambiguously. Starting with these five segments, a consensus framework for the fold was determined by carrying out all pairwise comparisons of the five cytokine structures and retaining α-carbon positions common to all of the pairs. The results of the pairwise comparisons are summarized in Table 2 and Fig. 3.

The pairwise comparisons yield from as few as 68 equivalent residues in the GM-CSF/M-CSF pair to as many as 81 equivalent residues in the nIL-4/IL-2, GM-CSF/IL-5 and IL-2/IL-5 pairs. (nIL-4 refers to the

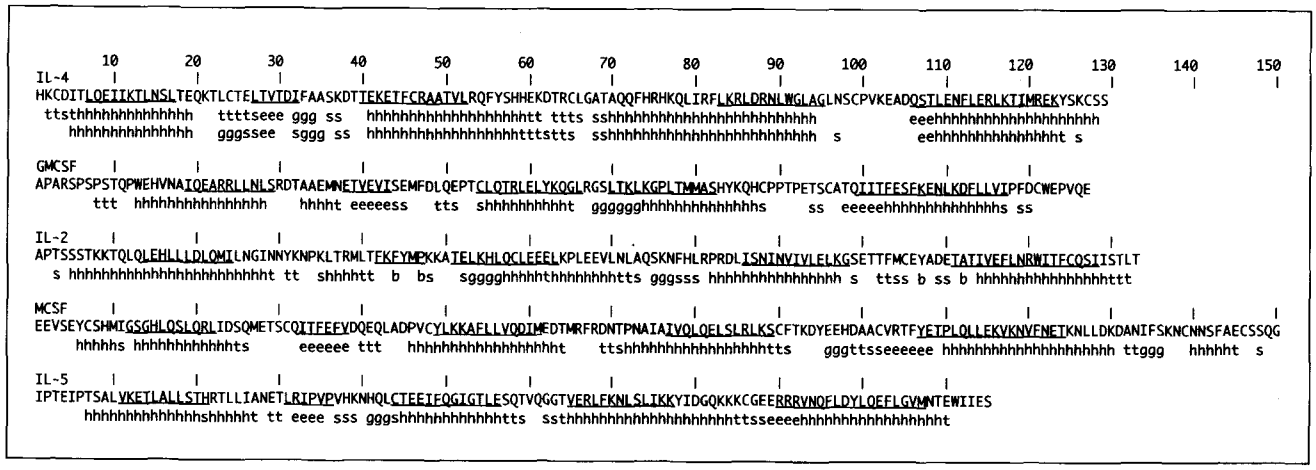

Fig. 2. Sequences and secondary structures of the five cytokines compared in this study. For each protein, the first line gives the amino acid sequence in the one-letter code (upper-case) and the second line gives the secondary structures according to the program DSSP [39]; e = β-sheet, h = α-helix, g = 3_{10} helix, b = β-bridge, t = hydrogen-bonded turn and s = bend. For IL-4, the structure assignment for the X-ray diffraction-derived model (xIL-4) is given first and the NMR-derived model (nIL-4) is on the following line. Regions which are part of the consensus framework (Fig. 3) are underlined.

NMR-derived structure.) The equivalent α-carbon positions common to all of the combinations constitute a common framework for the fold, which consists of 11 residues from α-helix A, 6 residues from β-strand 1, 13 residues from α-helix B, 13 residues from α-helix C and 18 residues from β-strand 2/α-helix D for a total of 61 residues. These five conserved segments will be referred to as αA, β1, αB, αC, and β2/αD, respectively. The consensus framework as defined includes 41–48% of the residues of each cytokine. As can be seen in the superimposed structures (Fig. 4) and has been discussed in many of the original publications, the remainder of the residues are accounted for by the varying lengths of the six main secondary structural elements plus variations in the connecting loops and termini which can include additional short secondary structural elements.

Table 2 lists the root mean square (rms) deviations of the final overlays for the α-carbons in each of the structural elements based on all pairwise equivalent residues and the subset included in the consensus framework. The overall deviations in the pairwise combinations range from 1.7 Å in the IL-4/IL-2 and IL-4/GM-CSF pairs to 2.9 Å in the IL-4/M-CSF and IL-2/IL-5 pairs. Comparisons of the individual segments of the framework show that the αA and αD elements have the lowest average rms deviations (1.6 Å and 1.8 Å), while the β1 and αB elements have the highest average rms deviations (2.8 Å and 2.9 Å). The very high deviations of αB in all IL-5 comparisons are due to a rotation of this helix in IL-5. These results reflect that the A, C and D helices are most consistently packed relative to one another. However, this trend is not completely uniform. For instance, αA and αB are the most similar segments in the IL-4/M-CSF pair.

For both the pairwise combinations and the common framework, the overall deviations show that IL-4, GM-CSF, and IL-2 are more similar to each other than to M-CSF or IL-5. It is interesting that the various levels of structural similarity are not reflected in the level of amino acid sequence identity. In fact, within the common core, the xIL-4/M-CSF pair has both a greater sequence identity (20%) and higher structural deviation (2.8 Å), while the GM-CSF/IL-2 pair has a much lower sequence identity (11%) and a lower structural deviation (2.1 Å). This observation suggests that divergence in sequence among this family of proteins is extensive enough that, for the most part, residue identities as opposed to similarities are likely to be coincidental rather than the result of absolute conservation of a residue from a common ancestor. This is supported by the observation that many identical residues adopt different side chain conformations in the various structures (data not shown). Unfortunately, this phenomenon makes sequence alignment and homology building rather difficult.

Common inner core

The 61 residues that make up the family framework include most of the residues that contribute to the buried inner core of the fold. To identify these residues and to further assess how similar the frameworks of the individual structures are, we calculated the accessible surface area as a function of residue number (Fig. 5). There is general agreement between the plots, as the α-helices and β-strands tend to show the expected periodicity of three to four residues and two residues, respectively. These trends are particularly visible in the average profile (Fig. 5f).

Among the five structures, 31 residues within the framework average $<20\%$ surface accessibility. As anticipated, these residues correlate well with those which can be seen to contribute to the packing of the inner core of each protein. The major exceptions in the individual profiles are located in αB, where some non-core residues have lower than expected accessibilities because they are covered by the αC to αD connect-

Table 2. Structural deviations and sequence identities for pairwise overlays.

		αA	β1	αB	αC	β2/αD	Overall
xIL-4/GM-CSF	equiv	14 (11)	10 (6)	13 (13)	14 (13)	18 (18)	69 (61)
	rmsd	1.2 (1.2)	2.0 (2.0)	2.3 (2.3)	1.9 (1.8)	1.1 (1.1)	1.7 (1.7)
nIL-4/GM-CSF	equiv	14 (11)	10 (6)	13 (13)	14 (13)	18 (18)	69 (61)
	rmsd	0.9 (0.9)	1.7 (1.9)	2.1 (2.1)	2.1 (2.1)	1.2 (1.2)	1.7 (1.7)
	ident	29 % (36 %)	30 % (50 %)	8 % (8 %)	29 % (31 %)	33 % (33 %)	26 % (30 %)
xIL-4/IL-2	equiv	18 (11)	6 (6)	13 (13)	19 (13)	22 (18)	78 (61)
	rmsd	1.2 (1.1)	1.7 (1.7)	2.4 (2.4)	1.7 (1.5)	1.3 (1.3)	1.7 (1.6)
nIL-4/IL-2	equiv	18 (11)	10 (6)	13 (13)	19 (13)	21 (18)	81 (61)
	rmsd	1.3 (1.3)	2.5 (1.0)	2.3 (2.3)	1.9 (1.7)	1.2 (1.2)	1.8 (1.6)
	indent	11 % (18 %)	0 % (0 %)	23 % (23 %)	11 % (15 %)	27 % (28 %)	17 % (20 %)
xIL-4/M-CSF	equiv	15 (11)	7 (6)	15 (13)	19 (13)	24 (18)	80 (61)
	rmsd	2.3 (2.6)	4.2 (3.8)	2.8 (2.9)	2.6 (2.5)	3.0 (2.5)	2.9 (2.8)
nIL-4/M-CSF	equiv	15 (11)	8 (6)	15 (13)	19 (13)	21 (18)	78 (61)
	rmsd	1.9 (1.9)	3.9 (3.5)	2.2 (2.4)	2.4 (2.1)	3.0 (3.0)	2.7 (2.6)
	ident	13 % (18 %)	14 % (17 %)	13 % (15 %)	11 % (15 %)	21 % (28 %)	15 % (20 %)
xIL-4/IL-5	equiv	15 (11)	6 (6)	16 (13)	15 (13)	22 (18)	74 (61)
	rmsd	1.7 (1.2)	4.5 (4.5)	4.1 (4.2)	2.5 (2.1)	1.8 (1.7)	2.8 (2.8)
nIL-4/IL-5	equiv	15 (11)	7 (6)	16 (13)	15 (13)	19 (18)	72 (61)
	rmsd	1.3 (0.9)	4.8 (4.6)	3.9 (4.0)	2.3 (2.1)	1.7 (1.7)	2.8 (2.7)
	ident	13 % (18 %)	17 % (17 %)	25 % (23 %)	27 % (31 %)	14 % (17 %)	19 % (21 %)
GM-CSF/IL-2	equiv	18 (11)	13 (6)	14 (13)	14 (13)	19 (18)	78 (61)
	rmsd	1.7 (1.4)	2.9 (1.6)	3.3 (3.1)	2.7 (2.5)	1.2 (0.9)	2.4 (2.1)
	ident	6 % (9 %)	0 % (0 %)	14 % (15 %)	0 % (0 %)	21 % (22 %)	9 % (11 %)
GM-CSF/M-CSF	equiv	11 (11)	10 (6)	13 (13)	15 (13)	19 (18)	68 (61)
	rmsd	2.3 (2.3)	3.1 (2.2)	2.2 (2.2)	2.7 (2.8)	2.4 (2.4)	2.6 (2.4)
	ident	9 % (9 %)	20 % (33 %)	15 % (15 %)	13 % (15 %)	16 % (17 %)	15 % (16 %)
GM-CSF/IL-5	equiv	16 (11)	11 (6)	15 (13)	20 (13)	19 (18)	81 (61)
	rmsd	1.6 (1.3)	3.5 (3.0)	2.7 (2.7)	2.9 (2.6)	1.5 (1.4)	2.5 (2.2)
	ident	19 % (27 %)	9 % (17 %)	7 % (8 %)	10 % (15 %)	26 % (28 %)	15 % (20 %)
IL-2/M-CSF	equiv	15 (11)	8 (6)	14 (13)	18 (13)	21 (18)	76 (61)
	rmsd	2.2 (2.6)	2.8 (2.5)	2.7 (2.6)	1.6 (1.6)	2.2 (2.1)	2.3 (2.3)
	ident	20 % (27 %)	13 % (17 %)	7 % (8 %)	22 % (31 %)	10 % (11 %)	14 % (18 %)
IL-2/IL-5	equiv	17 (11)	12 (6)	14 (13)	13 (13)	25 (18)	81 (61)
	rmsd	1.3 (1.2)	4.6 (3.2)	4.4 (4.6)	1.9 (1.9)	1.5 (1.5)	2.9 (2.7)
	ident	18 % (18 %)	17 % (17 %)	7 % (8 %)	8 % (8 %)	16 % (17 %)	14 % (13 %)
M-CSF/IL-5	equiv	12 (11)	7 (6)	15 (13)	17 (13)	22 (18)	73 (61)
	rmsd	2.1 (2.0)	3.4 (3.2)	3.1 (3.1)	2.3 (2.0)	3.0 (2.9)	2.8 (2.7)
	ident	17 % (18 %)	0 % (0 %)	7 % (8 %)	12 % (15 %)	9 % (11 %)	10 % (11 %)
Average	equiv	15 (11)	9 (6)	14 (13)	16 (13)	21 (18)	76 (61)
	rmsd	1.6 (1.6)	3.3 (2.8)	2.9 (2.9)	2.3 (2.1)	1.9 (1.8)	2.4 (2.3)
	ident	16 % (20 %)	12 % (17 %)	13 % (13 %)	14 % (18 %)	19 % (21 %)	15 % (18 %)

For each pairwise comparison, statistics are given for all equivalent residues in the structural element and, in parentheses, for the subset of residues found in the family consensus framework. The statistics reported are as follows: equiv = the number of equivalent residues in the segment; rmsd = the root mean square deviations of the equivalent C_α atoms; ident = the percent amino acid sequence identity based on the equivalent residues. A comparison of the xIL-4 and nIL-4 models assigned 103 equivalent residues with an overall rmsd of 1.1 Å and a core rmsd of 0.9 Å.

ing segment which passes around the outside of αB. This chain crossing takes place in different positions for the five structures and causes major differences in the individual accessibility patterns of αB. In GM-CSF and M-CSF the crossover is very low and mostly covers residues 1, 3 and 4 of αB, while for IL-2 it passes higher mainly covering residues 4 and 7 of αB. In IL-4 it passes still higher to cover residues 7 and 11 of αB (Figs 4 and 5). In IL-5 this is where the crossover between the subunits occurs, so that the αC to αD connecting

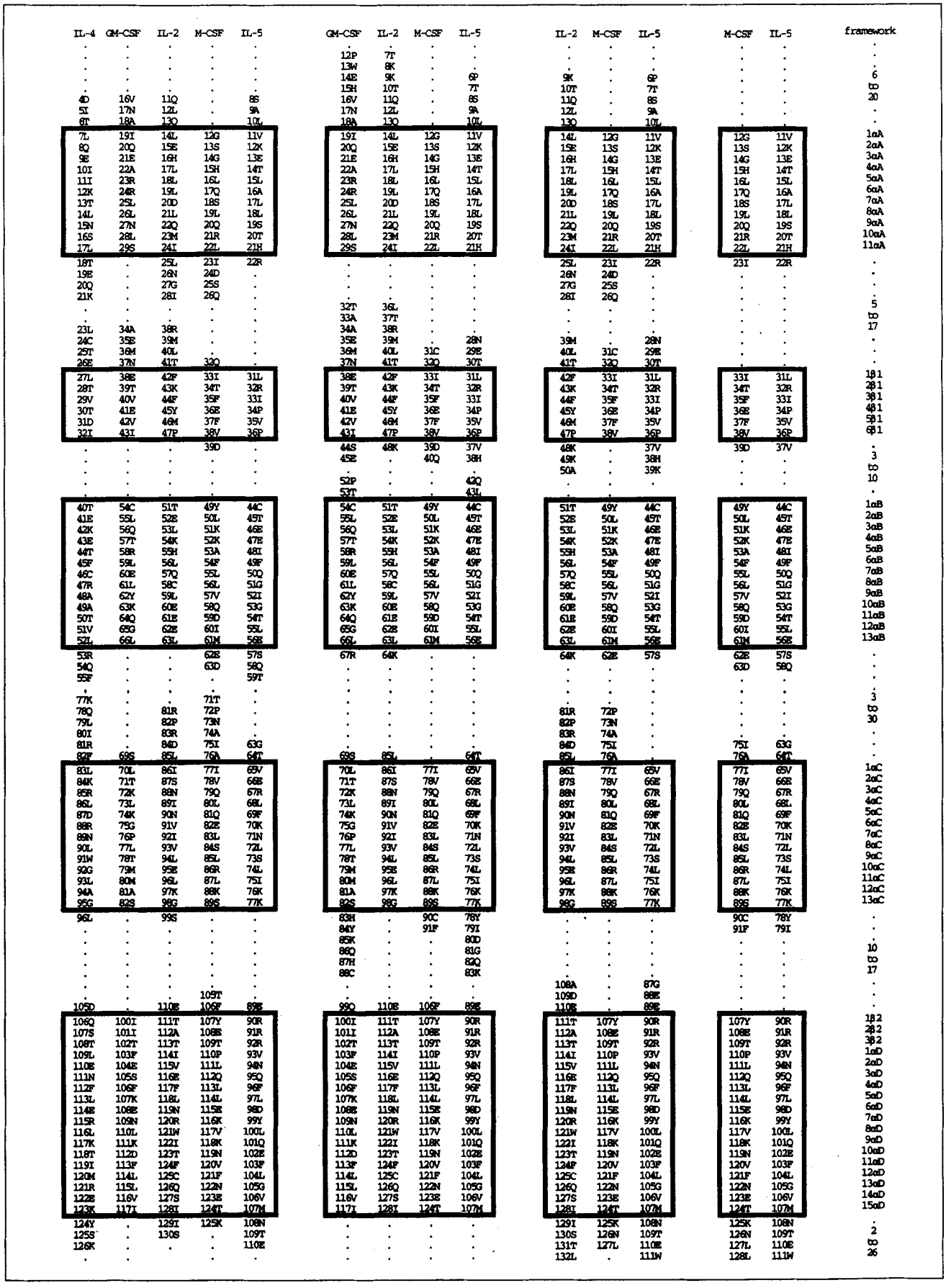
IL-4 GM-CSF IL-2 M-CSF IL-5 framework
GM-CSF IL-2 M-CSF IL-5
IL-2 M-CSF IL-5
M-CSF IL-5

Fig. 3 (opposite). Structurally equivalent residues for each pairwise comparison. The four groups of columns show pairwise comparisons of IL-4 with the other four structures, then GM-CSF with the remaining three structures, then IL-2 with the remaining two and lastly M-CSF with IL-5. The family consensus framework produced by all the comparisons is boxed. In the framework column, each residue position is given a designation according to the segment in which it belongs. In this same column, the range of the number of residues observed between each segment in all of the protein sequences is included.

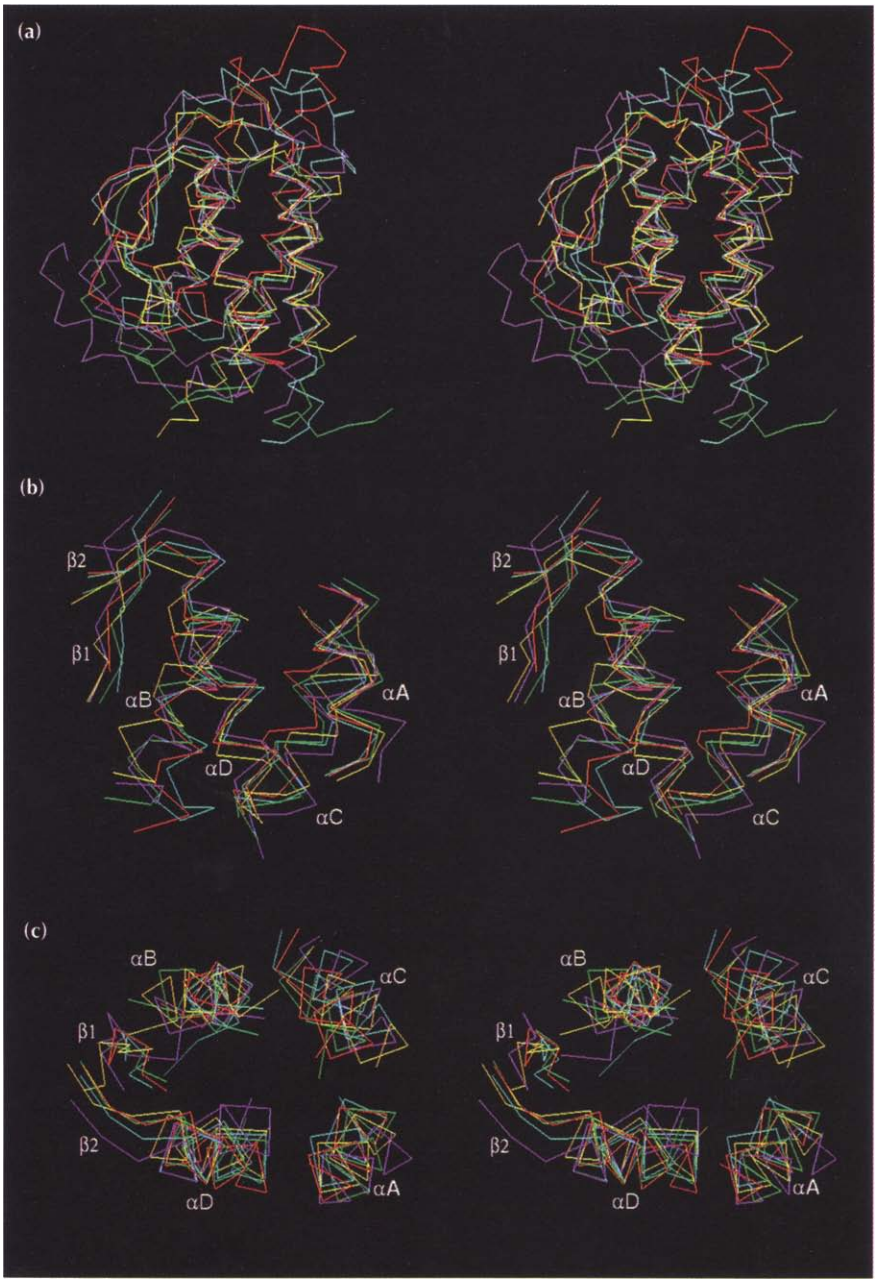

Fig. 4. Stereoviews of the superposition of the α-carbon backbones of the five short-chain cytokine structures: IL-4 (red), GM-CSF (green), IL-2 (cyan), M-CSF (violet) and IL-5 (yellow). **(a)** Complete C_α-backbone trace shown in a similar orientation to Fig. 1a. The central portions of the structures, which constitute the consensus framework, superimpose well, even though the individual helix lengths are quite diverse. Of note is the variation in the path of the chain coming around the left side of the structure from the bottom of helix C to the top of helix D. In GM-CSF (green) the chain passes near the bottom of helix B, in M-CSF (violet) and IL-2 (cyan) it passes somewhat higher and in IL-4 (red) it passes higher still. The structure shown for IL-5 is a composite monomer comprising residues 5–87 from one chain and 88–112 from the other chain. **(b)** C_α-backbone trace showing just the consensus framework in the same orientation as (a). **(c)** C_α-backbone trace of the consensus framework rotated through 90° relative to (a) and (b) such that the view looks down the axis of the bundle.

segment does not cover αB at all. In IL-4, GM-CSF and IL-2 this chain crossing is stabilized by a disulfide bond to αB.

Even among the 31 inner-core residues, there are variations in the individual accessibility patterns. These are partly correlated with small rotations in the helix axes, but mostly are due to the occurrence of residues with different polarity at sites that are near the surface of the proteins. There are 13 cases (out of $31 \times 5 = 155$ total positions) for which the individual accessibility value of an inner-core residues exceeds 20% (Fig. 5). The composition of the amino acids found at these positions is noteworthy: four are lysine, four glutamic acid, two arginine, two glycine and there is one each of threonine, proline and isoleucine. Ten of these are polar residues with long apolar stems which can fold to bury their methylene groups and turn outward to place the polar end in the solvent. A good example of this is seen at position 5αA, where four structures have apolar residues, while human GM-CSF has an arginine (Figs 3 and 6a).

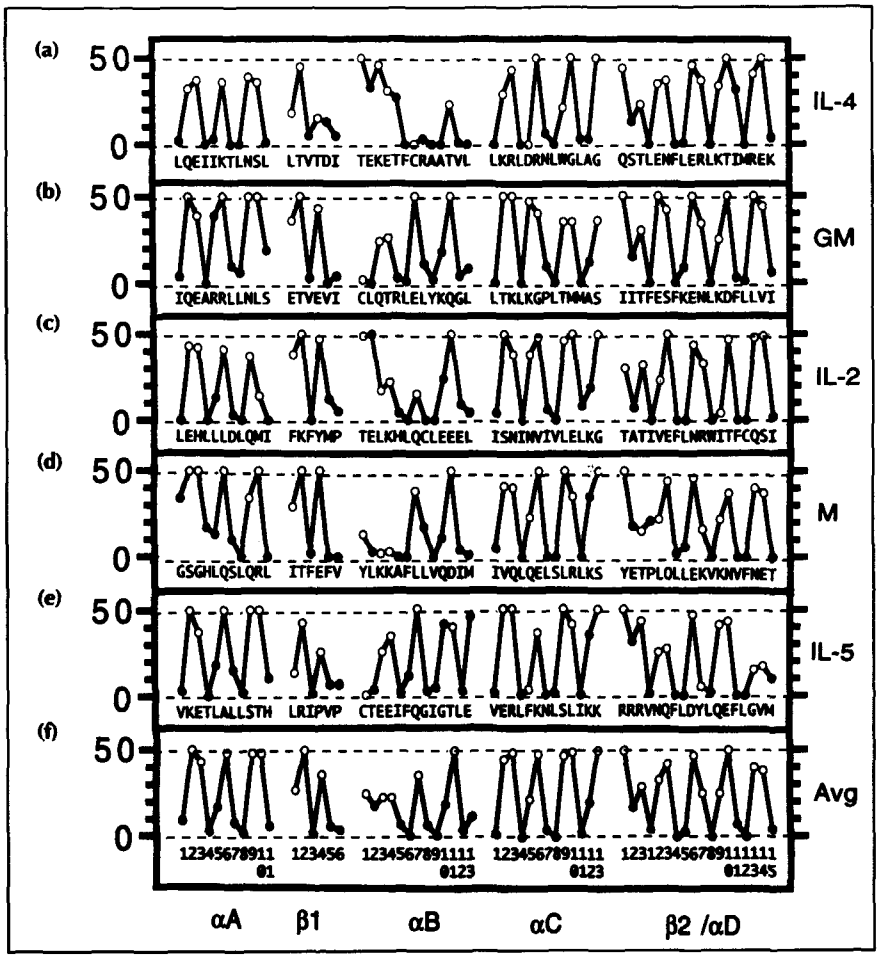

Fig. 5. Solvent accessibility patterns of the framework residues. **(a)** IL-4 (X-ray derived), **(b)** GM-CSF, **(c)** IL-2, **(d)** M-CSF, **(e)** IL-5, and **(f)** average. The accessible surface area for each residue was calculated with the program DSSP [39] using all atoms of the structures. For M-CSF the monomer was used. To convert to percent accessibilities, the surface areas were divided by those calculated for amino acids in a Gly-X-Gly tripeptide [56]. Values $>50\,\%$ were set to 50 % for plotting purposes but not for the averaging. The positions which average $<20\,\%$ accessibility and are defined as the inner-core positions are shown by filled circles. The fact that some residues which have average accessibilities of between 20 and 25 % are excluded from the inner-core while some residues with exactly 20 % are included emphasizes that such a 20 % cutoff, although useful for analysis, is somewhat artificial.

The overall amino acid compositions of the inner core and the exterior framework are given in Table 3. The large majority of the inner core residues are hydrophobic, but still 23 % of them have a side chain with hydrogen-bonding potential. Most of these side chains are folded so that the hydrogen-bonding group is at the surface of the protein, but some of them are involved in buried hydrogen-bonding interactions (see below). If the inner core were to be defined more restrictively, including only residues having an average accessibility of $\leq 5\,\%$, 18 positions would qualify, and hydrogen bonding side chains would only account for about 12 % of the residues in each of the protein cores. The polar side chains are fairly well distributed among the inner-core residues so that only 12 of the 31 inner-core positions remain purely apolar. If all known sequences of these proteins are included in the analysis (Fig. 7) then only nine of the inner-core residues remain consistently apolar, and only a single inner-core residue, the leucine at 8αA, is perfectly conserved. (The reported DNA sequence for sheep IL-4 has a proline at this position, but as only a single base change converts the codon to a leucine, this may be a sequencing error.) Even for this leucine, however, the side chain torsion angles are not conserved, suggesting there may not be an absolute structural requirement for leucine at this position. It is also of interest that none of the disulfide bonds are within the helical bundle core. The cysteine in the inner core at 15αD of IL-2 is present as a sulfhydryl and that at 8αB is in a disulfide bond with the C to D crossover.

An analysis of the major packing interactions in each protein shows that despite the diversity in sequences, the inner core contact patterns are reasonably well conserved. For inner-core contacts within 4.5 Å, IL-4 has 52 interactions, GM-CSF has 57 interactions, and IL-2 has 54 interactions. Twenty-four of these interactions occur at common residue positions in these three proteins. An interesting feature of this is the presence of hydrophilic side chains that have been substituted for hydrophobic side chains at buried positions. These substitutions are usually tolerated through differences in pairs (or triplets) of residues and may also involve changes in main chain conformation. An example of this occurs at positions 11αA and 5αD. In GM-CSF a serine–lysine hydrogen bond pair exists, while in the other structures, and even in the mouse GM-CSF sequence [21], hydrophobic side chains occupy this location (Fig. 6b). Other examples include the Asp31–Ser107 pair in IL-4 (Fig. 6c), the Tyr62–Asn109 pair in GM-CSF, the Gln58–Ser84 pair in

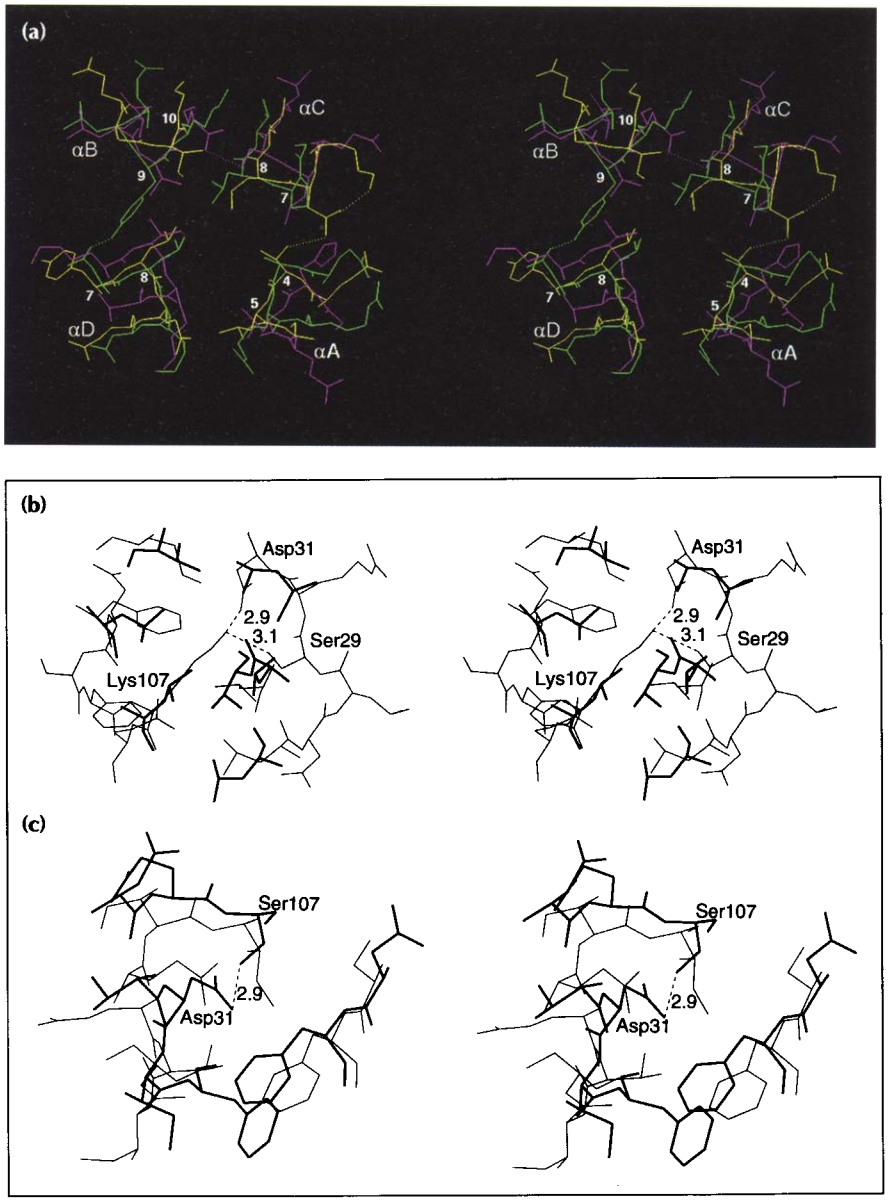

Fig. 6. Examples of polar residues in inner-core positions. **(a)** A slice through the four-helix bundle showing residues 3αA–6αA, 7αB–10αB, 6αC–9αC and 7αD–10αD. GM-CSF (green), IL-5 (yellow), and M-CSF (violet) are shown. Arg23 of GM-CSF (at position 5αA) and His15 of M-CSF (at position 4αA) exemplify residues at inner-core positions which adopt conformations that expose their polar moieties to solvent. Dotted lines highlight buried hydrogen bonds involving Tyr62...Asn109 in GM-CSF (green), Thr14...Asn71...Lys70 of IL-5 (yellow), and Gln58...Ser84 in M-CSF (violet). The labels give the consensus framework numbering. **(b)** The Ser29 (11αA), Asp31, Lys107 (5αD) triplet of human GM-CSF. Asp31 is not a framework residue, but is included in the interaction. It should be noted that even in mouse GM-CSF, these three residues are changed to a hydrophobic triple of methionine, valine and isoleucine [21]. **(c)** The Asp31 (5β1), Ser107 (2β2) pair in IL-4. In (b) and (c) GM-CSF (thick lines) and IL-4 (thin lines) are both shown to contrast the polar and apolar clusters and lengths of hydrogen bonds are given in Å.

M-CSF, and the Thr14–Asn71 pair of IL-5 (all shown in Fig. 6a).

Alignment of unknown structures

In order to predict the secondary structure locations of other members of the short-chain subfamily, an accurate amino acid sequence alignment is necessary. Unfortunately, given the high divergence of the sequences in this family, such alignments have been difficult, as has been noted previously [22–25]. However, the analysis presented above shows that the short-chain helical cytokine family has a well conserved structural framework whose information content could be extended to other members of the family. Fig. 7 shows a sequence alignment profile for the short-chain subfamily based on the framework of the five structures analyzed here. The profile was improved by including the sequences of other species for each cytokine. It consists of the five segments of the consensus framework in which no gaps are allowed, a minimum spacing between each segment, and a requirement that each framework element map largely onto its appropriate exon (see Fig. 7 legend).

The power of the profile was tested by cross-validation analysis, where the five cytokines were each individually removed from the profile and the remaining four were used to find the optimal alignment of the one excluded. The cross-validation tests showed that αA and αC could be correctly aligned in all five cases, while β2/αD was correctly aligned in four cases, and β1 and αB were correctly aligned in only three cases. The lower performance for β1 may relate to its short length, and for αB it may relate to the variation in its interaction with the αC to αD connection which causes the hydrophobicity pattern to be less conserved. For the helices that are misaligned, the alignment is incorrect by either 3 or 7 residues (one or two turns) so that the hydrophobicity

Table 3. Amino acid composition of inner-core and external framework residues.

	Residue type	Inner-core	External
Non-hydrogen bonding:	Gly	4	6
	Ala	7	1
	Cys	2	3
	Pro	4	1
	Val	13	5
	Ile	20	3
	Leu	50	11
	Met	5	2
	Phe	15	2
	Subtotal	120 (77 %)	34 (23 %)
Hydrogen bonding:	Ser	4	9
	Thr	5	19
	Asn	2	10
	Asp	2	4
	His	3	1
	Gln	1	16
	Glu	6	24
	Lys	6	16
	Arg	4	12
	Tyr	1	4
	Trp	1	1
	Subtotal	35 (23 %)	116 (77 %)
	Totals	**155**	**150**

The number of residues is summed over the five structures used in the comparison. The 31 residues of the inner-core and the 30 residues which are external are defined in Fig. 5. Although the sulfur atoms of cysteine and methionine are technically polar and can be involved in hydrogen bonding, such interactions are rarely seen [58] and we have classified the residues as non-hydrogen bonding.

pattern is in register. Fig. 7 also reports the alignments yielded by our profile for the putative members of the short-chain subfamily IL-3, IL-7 and stem cell factor (SCF).

Comparison with growth hormone

Human growth hormone is a member of the long-chain subfamily, which clearly has its own unique structural framework. However, since much is known about the way in which growth hormone binds to its receptor [19,20], a comparison of the short-chain framework with growth hormone could give insight into how the short-chain helical cytokines bind their receptors. Because the helices are much longer in growth hormone and the helix packing angles are somewhat different, six different alignments are possible which allow a reasonable correspondence of all helical framework residues from the short-chain fold with helical residues from growth hormone (Fig. 8). Overlay statistics were calculated for each of the six possible overlays (rasters) using each of the five short chain structures (Table 4). No single raster is best for all five structures, with lowest deviations occurring for rasters 1a (for IL-2), 3a (for IL-4, GM-CSF and M-CSF), and 3b (for IL-5). However, if it is assumed that a single overlay is valid for all of the structures then raster 3a clearly stands out as the best choice. The alignment for this raster is included in Fig. 7 (as human GH).

Discussion

A comparison of the five known short-chain helical cytokine structures has revealed that there is a structurally conserved framework which includes five segments and nearly half of the residues in each protein. However, the level of structural difference (as great as 2.9 Å rms deviation) highlights the malleability of the short-chain cytokine fold. In a study of protein structural evolution by Chothia and Lesk [26], the pairs with the greatest divergence (hemoglobin α-chain *versus* erythrocruorin and plastocyanin *versus* azurin) have an rms deviation of only 2.3 Å. We speculate that there are three main factors contributing to the high plasticity of the cytokine fold: firstly, there is no active site which demands extremely precise positioning of reactive groups; secondly, the individual cytokine structures can co-evolve with their cognate receptor so that even those residues which are important for recognition may have more freedom to change and shift, and thirdly, the fold is based on a helical bundle so that, unlike the β-trefoil family of factors [2], there are no main chain–main chain interactions in the core which must be conserved and thus limit structural divergence.

As in all comparative studies, the exact composition of the framework segments defined here is not absolute: it is a function of the set of cytokine structures used, the fact that equivalencies were based on local rather than global overlays, and the 3 Å cutoff value used. Examining the residues included in the β1 segment illustrates this point. If only nIL-4, GM-CSF and IL-2 were used for comparison, four residues which form a short α-helix between αA and β1 would have been included in the framework (Fig. 3). In addition, although nIL-4 and xIL-4 (the NMR-derived and X-ray-derived structures, respectively) are very similar, these same extra four residues are not scored as equivalent in the xIL-4/IL-2 pair. This is due to a small difference between xIL-4 and nIL-4 such that the residue preceding β1 is 3.1 Å away from its equivalent in IL-2 and is barely excluded by the 3 Å cutoff.

In addition to the defined structural elements of the framework, other common features of the fold exist. Four of the five proteins have a disulfide bond connecting αB to the loop between αC and β2/αD. However, the cysteine in αB occurs at a unique framework position in each protein (7αB in IL-4, 1αB in GM-CSF and IL-5, and 8αB in IL-2; M-CSF has a disulfide involving the residue preceding 1αB, but connects to a residue after β2/αD). The disulfide bond and conserved β-sheet in the short-chain subfamily may be involved in stabilizing the chain crossings which differ from those of the other subfamilies (Fig. 1).

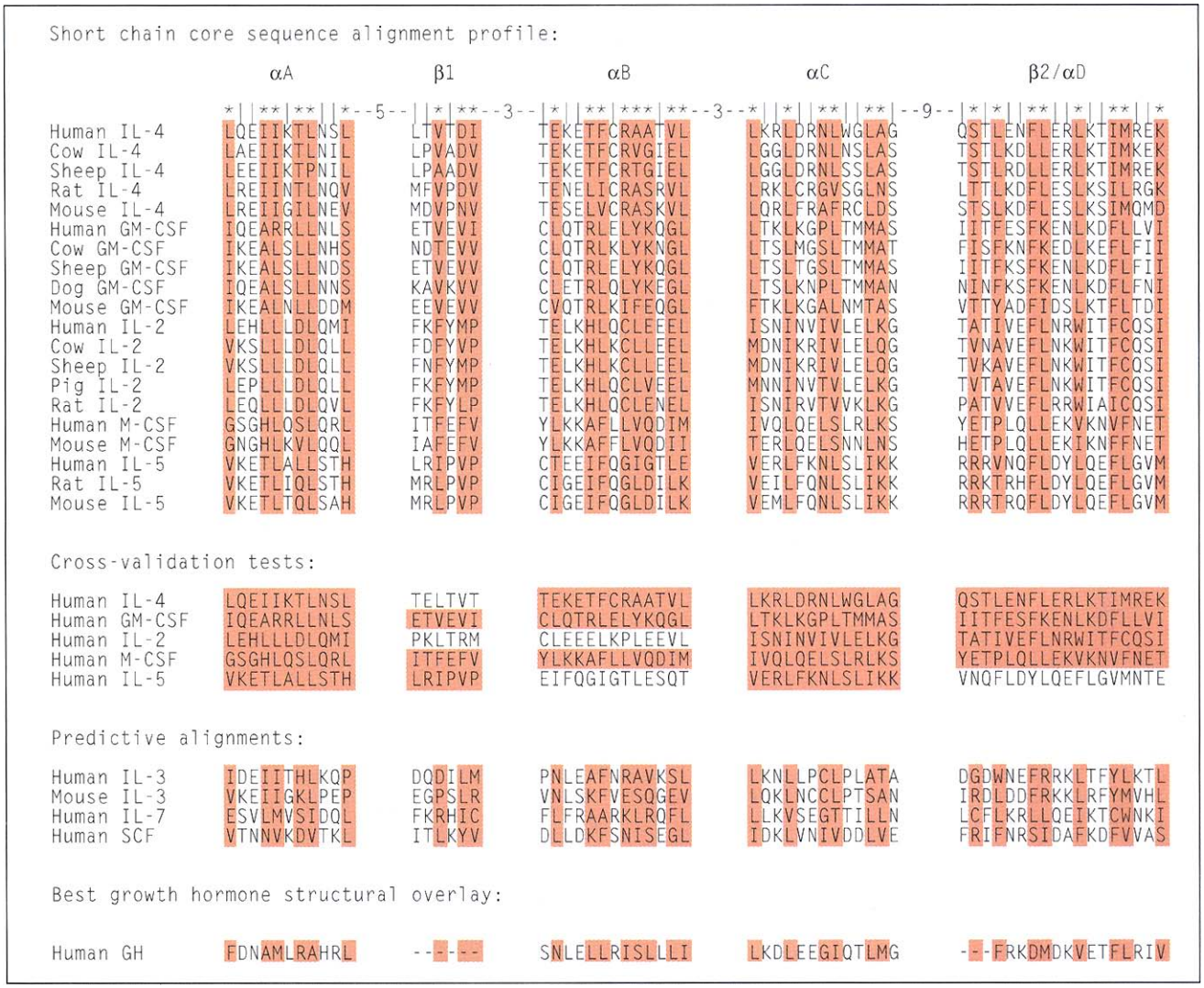

Fig. 7. Amino acid sequence alignments based on framework residues. The first section shows the aligned sequences of the five known structures and their cognates from different species. The vertical shading shows the residues that constitute the inner core. The numbers between the framework segments are the minimum number of residues required to bridge the adjacent segments. The five lines in the second section are the best alignments resulting from the cross-validation tests. The best alignments which correctly match the true structural alignment are shaded. The four lines in the third section are the optimal alignments of the structurally unknown family members IL-3, IL-7, and stem cell factor (SCF). The vertical shading again highlights the inner core. The last line gives the optimal structural alignment of growth hormone (raster 3a of Fig. 8, Table 4).

Implications for protein design and modeling

In *de novo* protein design studies, the four helix bundle has been a common target because of its simplicity [27]. It has been shown that designs employing amphipathic helices, with leucines uniformly on the inside and hydrophilic residues on the outside, form extremely stable bundles. However, NMR studies reported that the designs had more of a molten globule-like nature instead of a unique mature structure [28]. The short-chain helical cytokines, each forming a uniquely structured inner core, provide an interesting contrast. Their inner cores contain primarily leucine residues, but also include a significant number of isoleucine, phenylalanine and valine residues, as well as 23% polar side chains (especially the longer ones), of which about one third are truly fully buried.

Although isoleucine and valine are not strong helix-forming residues [29], when they do occur in a helix, their β-branched side chains can only adopt a single conformation at χ_1 [30], providing a significant ordering influence on the core. Similarly, the phenylalanine residues contribute to ordering as they are more conformationally restrained than leucines and are also found primarily as a single rotomer. In addition, buried polar side chains will tend to interact with other polar groups (either on the surface or inside) in a unique fashion, thus further contributing to a specific architecture of the inner core. A detailed analysis of the non-framework residues in each protein is beyond the scope of this paper, but a cursory inspection reveals a number of well buried polar side chains which hydrogen bond to distant portions of the chain to stabilize

Table 4. Statistics for overlays on human growth hormone.

Raster	IL-4	GM-CSF	IL-2	M-CSF	IL-5	Average
1a	1.9	2.4	**1.8**	2.2	3.0	2.3
1b	2.5	2.5	2.2	2.9	2.5	2.5
2a	2.5	2.8	2.6*	2.9	3.5	2.9
2b	2.4	2.6	2.5*	3.1	2.9	2.7
3a	**1.8**	**2.0**	1.9*	**1.9**	2.7	**2.1**
3b	2.5	2.3	2.4*	2.9	**2.2**	2.5

The root mean square deviations (in Å) are given for each of the five short-chain structures for each of the six possible ways they can be overlayed on the structure of human growth hormone (see Fig. 8). The best raster for each structure is highlighted in bold. All 53 C_α positions (from the four helices of the bundle and residue 3β2) are found to be equivalent, except for four cases of IL-2 noted by asterisks in which only 52 residues were defined as equivalent.

the tertiary fold. This is reminiscent of protein–protein interfaces where it is thought that apolar contacts provide the majority of the binding strength and polar contacts largely provide specificity [31].

The cross-validation test shows that alignments based on a four structure profile are imperfect, but are approaching a reasonable level of reliability. It should be noted that since some of the criteria used in the alignment (the minimum connecting lengths and the exon requirements) were derived from the five known structures, some bias is present even in the cross-validation tests. For instance, if the exon requirement is left out, the number of incorrectly aligned segments jumps from five to eight (data not shown). For predictive purposes, these additional requirements are only helpful if they are obeyed by the structures to be aligned. Indeed, if they are not obeyed then their inclusion guarantees that the profile will deliver an incorrect answer! We purposely have used no other information to generate the predictive alignments of IL-3, IL-7 and SCF, although such information is available in some cases. For instance, we suspect that αB of SCF is not correctly aligned in Fig. 7, but that the alignment with M-CSF reported by Bazan [32] which conserves a cysteine preceding residue 1αB will be correct.

One of the major contributing factors to the difficulty of aligning members of this subfamily is the sporadic presence of polar residues at inner-core positions. There are nine inner-core positions for which four structures have apolar residues and only one has a polar side chain. It is striking that every one of the five structures contributes such a unique residue: IL-4 has aspartic acid at 5β1 and arginine at 8αB; GM-CSF has arginine at 5αA, tyrosine at 9αB, and lysine at 5αD; IL-2 has glutamic acid at 12αB and tryptophan at 8αD; M-CSF has serine at 8αC; and IL-5 has glutamic acid at 13αB. This suggests that even the nine inner-core positions that are seen to be apolar in all of the sequences of the profile need not be apolar in other short-chain cytokines. Since many occurrences of polar residues in the inner core involve complementary substitutions at two or three interacting positions (Fig. 6), sequence structure alignment methods using two-dimensional residue contact information [33,34] may perform better than one-dimensional methods [35,36]. However, even these methods will have difficulty with residues at the interfaces of the inner core and surface which can tuck into the core or point out to the surface depending on the nature of the side chain (Fig. 6a).

Structure/function predictions

Biochemical studies to identify receptor recognition residues in the short-chain helical cytokines have uni-

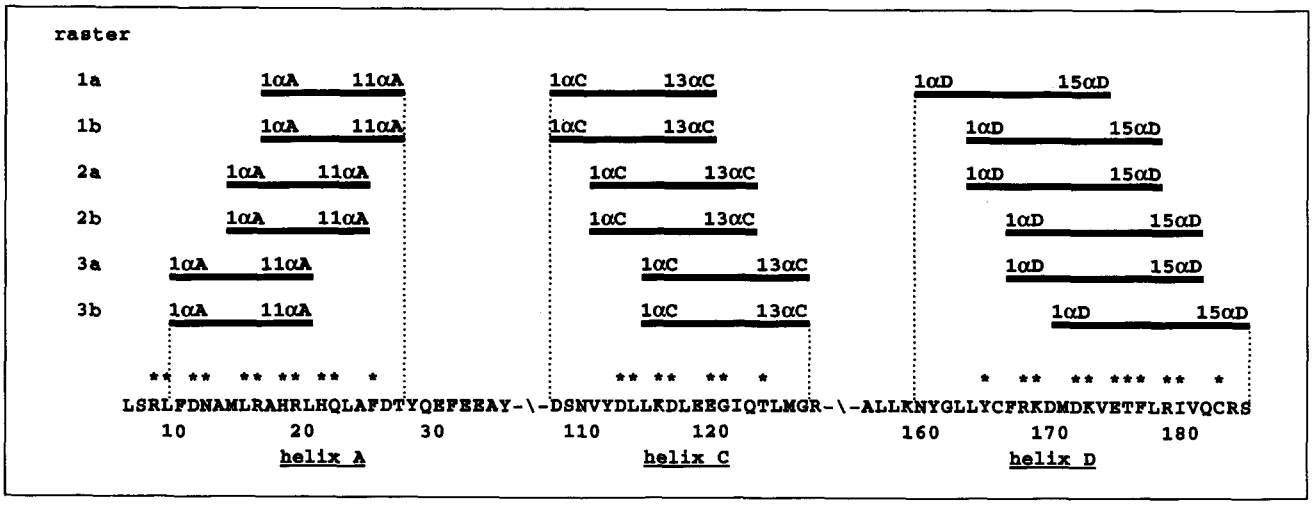

Fig. 8. The six tested alignments of the short chain consensus core with the α-helices of human growth hormone. The amino acid sequence of human growth hormone is given with asterisks indicating those residues directly involved in receptor binding [19,20]. The black bars above the sequence correspond to the aligned positions of the consensus framework helices A, C, and D. Rasters 1, 2 and 3 allow for the sliding of the shorter short-chain consensus helices along the longer growth hormone helices and rasters a and b allow for different tilts of the short-chain framework to match the different packing angles of the growth hormone helices.

formly identified residues which occur on the A, C and/or D helices [5,18]. One residue implicated to be important in IL-4, GM-CSF, IL-2 and IL-5 is an acidic residue on helix A. Within the consensus framework defined here, these residues are structurally equivalent in IL-4, GM-CSF and IL-5 (position 3αA), but in IL-2 the residue is one turn further up the helix (position 7αA). However, the side chain conformations are such that the carboxylates superimpose within about 3 Å, and could conceivably interact with an equivalent residue on a receptor. This example highlights the danger of assigning structural equivalencies on the basis of functional similarities.

Helices A, C and D are also the three main regions of the growth hormone structure which are recognized by its receptor [19,20]. This suggests that if an accurate structural alignment can be obtained between growth hormone and the short-chain subfamily, then it may be possible to predict receptor-binding residues within the short-chain subfamily. Our comparisons of the short-chain framework with growth hormone suggest that one of the six conceivable alignments is notably better than the others (raster 3a in Fig. 8 and Table 4). This alignment (shown in Fig. 7) is identical to those previously determined by considering all residues (not just the framework residues) of some individual short-chain structures [18,25,37]. We will not repeat the discussion of those papers, but note that this alignment does indeed align many known functional residues of the short-chain family members with growth hormone residues which contact the receptor (for example, Asn12 of growth hormone aligns with the above mentioned acidic residues at position 3αA). From a structural perspective, it is interesting to note that Asp169 of growth hormone aligns with 4αD, a position which is fully buried and occupied by phenylalanine or leucine in all the short-chain structures. As noted by de Vos *et al.* [19], Asp169 makes buried polar interactions with the side chains of Ser55 and Trp86.

Despite the appeal of the comparison with growth hormone, we note that there may not be a single choice for the correct alignment, as the individual helical cytokines may not all bind their receptors with equivalent residues. The growth hormone–receptor complex suggests that variation will occur because many residues outside of the four-helix bundle are also involved in receptor binding and these residues have no obvious equivalents in the short-chain subfamily. The possibility of the binding site sliding up or down the helical framework by a turn or two along the four-helix bundle would bring another level of flexibility to the evolution of binding specificity for this family fold. Additional cytokine–receptor complex structures are clearly required to provide answers to some of these issues.

Biological implications

Studies of the structure/function relations for cytokines are part of understanding the mechanisms of cell–cell communication at the molecular level. Structural studies have shown that cytokines and growth factors can be grouped into structural classes which transcend the classical nomenclature and groupings originally arising from the different disciplines of immunology, virology, hematology, and physiology. For instance, many proteinaceous messengers which are historically and biochemically quite distinct, including growth hormone, colony-stimulating factors, interferons and some interleukins, are found to share a common structural framework (a helical bundle) and a common receptor type. The members of this structural family can be divided into three subfamilies that are represented by the structures of granulocyte-macrophage colony-stimulating factor (GM-CSF), growth hormone and interferon-β, and which we refer to as the short-chain, the long-chain, and the interferon-like fold, respectively. An important goal is to determine how these molecules compare with each other and how information about structure and structure/function relationships can be generalized among these molecules.

In this report, we define a number of structural features shared by the five cytokines that are known to adopt a short-chain fold: IL-2, IL-4, IL-5, GM-CSF and macrophage colony-stimulating factor. However, we also show that even this large amount of information is insufficient to allow us to predict which other proteins will adopt the same fold, because of the high level of divergence in amino acid sequence and in structure seen in this family. We also identify the best structural alignment of the conserved framework of the short-chain fold with the growth hormone structure. This yields potential insights into the mode of receptor binding of the short-chain cytokines based on the structure of the growth hormone–receptor complex.

Materials and methods

Coordinates of the published structures of the cytokines were used for the comparisons (Table 1). For IL-4, only one NMR

structure [8,37] and one X-ray structure [11] were utilized, although more structures have been determined by both methods [9,10]. For IL-2, the high resolution structure from Hatada (unpublished data), for which we obtained permission to use the C_α coordinates, was used for structural comparisons, but it is quite similar to the structure reported by McKay [14].

Structural superpositions were performed based on α-carbon atoms alone. In order to overlay the structures and assign equivalent residues, the strategy developed by Chothia and Lesk [26] was adopted. In this procedure, equivalent secondary structural elements are chosen manually and used to define an initial conservative set of structurally equivalent segments. One segment at a time, the structures are optimally superimposed [38] and additional residues, adjacent to the original input residues, are identified as equivalent based on a 3 Å cutoff distance. The residues of the extended segments are then combined to define a complete set of equivalent residues which can be used to calculate the final overlay. This method has the advantage of being more generous in defining equivalent residues than methods based only on separation of residues in a global overlay. In the cases studied here, α-carbons shifted as far as 6.8 Å were included in the final overlay. However, we have found that the set of equivalent atoms defined is dependent on which atoms are used in the starting set, and if the output of one round is used as input to the next round the method does not always lead to a self-consistent and reasonable set of equivalent atoms. In this study, a self-consistent result was obtained by a cyclic procedure which used the set of equivalent atoms common to all pairwise comparisons as the input to the next round.

All accessibility calculations and secondary structure definitions were accomplished with the program DSSP [39]. Amino acid sequence alignments were carried out with a modified version of the package developed at Kansas State University [40] using the McLachlan similarity matrix [41]. The modification allows one to specify segments in one of the sequences into which gaps cannot be placed and which we have used to disallow gaps in the segments containing the core fold. Other scoring matrices were also used, but did not give improved results in the cross-validation tests.

Acknowledgements: We thank Steve Jordan and Mike Milburn, Glaxo Research Institute, Research Triangle Park, NC 27709, for allowing us access to the full coordinates of IL-5. We thank Brad Graves, Hoffman-LaRoche Inc., Nutley, NJ 07110, for allowing us early access to the α-carbon coordinates of IL-2 and running DSSP on the full coordinate set. This work was supported by grant GM43566-03 from the National Institutes of Health (NIGMS) to PAK, contract no. NO1-CO-74101 from the National Cancer Institute, DHHS with ABL to AW, and a grant from the AIDS Directed Antiviral Program of the Office of the Director of the National Institute of Health to AMG and GMC.

References

1. Thomson, A.W. (1991). *The Cytokine Handbook.* Academic Press, London and San Diego.
2. Murzin, A.G., Lesk, A.M. & Chothia, C. (1992). Beta-trefoil fold patterns of structure and sequence in the Kunitz inhibitors, interleukins 1-beta and 1-alpha, and fibroblast growth factors. *J. Mol. Biol.* **223**, 531–544.
3. Clore, G.M., Appella, E., Yamada, M., Matsushima, K. & Gronenborn, A.M. (1990). The three-dimensional structure of interleukin-8 in solution. *Biochemistry* **29**, 1689–1696.
4. Baldwin, E.T., *et al.*, & Wlodawer, A. (1991). Crystal structure of interleukin-8: symbiosis of NMR and crystallography. *Proc. Natl. Acad. Sci. USA* **88**, 502–506.
5. Sprang, S.R. & Bazan, J.F. (1993). Cytokine structural taxonomy and mechanisms of receptor engagement. *Curr. Opin. Struct. Biol.* **3**, 815–827.
6. Boulay, J.L. & Paul, W.E. (1992). The interleukin-4 family of lymphokines. *Curr. Opin. Immunol.* **4**, 294–298.
7. Bazan, J.F. (1990). Structural design and molecular evolution of a cytokine receptor superfamily. *Proc. Natl. Acad. Sci. USA* **87**, 6934–6938.
8. Powers, R., Garret, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. & Clore, G.M. (1992). Three-dimensional solution structure of human interleukin-4 by multidimensional heteronuclear magnetic resonance spectroscopy. *Science* **256**, 1673–1677.
9. Smith, L.J., *et al.*, & Dobson, C.M. (1992). Human interleukin-4. *J. Mol. Biol.* **224**, 899–904.
10. Walter, M.R., *et al.*, & Bugg, C.E. (1992). Crystal structure of recombinant human interleukin-4. *J. Biol. Chem.* **267**, 20371–20376.
11. Wlodawer, A., Pavlovsky, A. & Gustchina, A. (1992). Crystal structure of human recombinant interleukin-4 at 2.25 Å resolution. *FEBS Lett.* **309**, 59–64.
12. Diederichs, K., Boone, T. & Karplus, P.A. (1991). Novel fold and putative receptor binding site of granulocyte-macrophage colony stimulating factor. *Science* **254**, 1779–1782.
13. Walter, M.R., Cook, W.J., Ealick, S.E., Nagabhusan, T.L., Trotta, P.P. & Bugg, C.E. (1992). Three-dimensional structure of recombinant human granulocyte-macrophage colony stimulating factor. *J. Mol. Biol.* **224**, 1075–1085.
14. McKay, D.B. (1992). Unraveling the structure of IL-2 [Response]. *Science* **257**, 412–413.
15. Mott, H.R., Driscoll, P.C., Boyd, J., Cooke, R.M., Weir, M.P. & Campbell, I.D. (1992). Secondary structure of human interleukin-2 from 3D heteronuclear NMR experiments. *Biochemistry* **31**, 7741–7744.
16. Pandit, J., Bohm, A., Jancarik, J., Halenbeck, R., Koths, K. & Kim, S. (1992). Three-dimensional structure of dimeric human recombinant macrophage colony-stimulating factor *Science* **258**, 1358–1362.
17. Milburn, M.V., *et al.* & Wells, T.N.C. (1993). A novel dimer configuration revealed by the crystal structure at 2.4 Å resolution of human interleukin-5. *Nature* **363**, 172–176.
18. Kaushansky, K. & Karplus, P.A. (1993). Hematopoietic growth factors: understanding functional diversity in structural terms. *Blood* **82**, 3229–3240.
19. de Vos, A.M., Ultsch, M. & Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312.
20. Wells, J.A. & de Vos, A.M. (1993). Structure and function of human growth hormone: implications for the hematopoietins. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 329–351.
21. DeLamarter, J.F., Mermod, J.J., Liang, C.M., Eliason, J.F. & Thatcher, D.R. (1985). Recombinant murine GM-CSF from *E. coli* has biological activity and is neutralized by a specific serum. *EMBO J.* **4**, 2575–2582.
22. Schrader, J.W., Ziltner, H.J. & Leslie, K.B. (1986). Structural homologies among the hematopoietins. *Proc. Natl. Acad. Sci. USA* **83**, 2458–2463.
23. Parry, D.A.D., Minasian, E. & Leach S.J. (1991). Cytokine conformations: predictive studies. *J. Mol. Recogn.* **4**, 63–75.
24. Bazan, J.F. (1992). Unraveling the structure of IL-2. *Science* **257**, 410–412.
25. Wlodawer, A., Pavlovsky, A. & Gustchina, A. (1993). Hematopoietic cytokines: similarities and differences in the structures, with implications for receptor binding. *Protein Sci.* **2**, 1373–1382.
26. Chothia, C. & Lesk, A. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.* **4**, 823–826.
27. DeGrado, W.F., Wasserman, Z.R. & Lear, J.D. (1989). Protein design, a minimalist approach. *Science* **243**, 622–628.
28. Osterhout, J.J., *et al.*, & DeGrado, W.F. (1992). Characterization of the structural properties of α_1B, a peptide designed to form a four-helix bundle. *J. Am. Chem. Soc.* **114**, 331–337.
29. Chou, P.Y. & Fasman, G.D. (1978). Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**, 251–276.
30. McGregor, M.J., Islam, S.A. & Sternberg, M.J.E. (1987). Analysis of the relationship between side chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* **198**, 295–310.
31. Janin, J. & Chothia, C. (1990). The structure of protein-protein recognition sites. *J. Biol. Chem.* **265**, 16027–16030.

32. Bazan, J.F. (1991). Genetic and structural homology of stem cell factor and macrophage colony stimulating factor. *Cell* **65**, 9–10.
33. Godzik, A. & Skolnick, J. (1992). Sequence–structure matching in globular proteins: applications to supersecondary and tertiary structure determination. *Proc. Natl. Acad. Sci. USA* **89**, 12098–12102.
34. Jones, D.T., Taylor, R.W. & Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature* **358**, 86–89.
35. Bowie, J.U., Luthy, R. & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164–170.
36. Ouzounis, C., Sander, C., Scharf, M. & Schneider, R. (1993). Prediction of protein structure by evaluation of sequence-structure fitness. *J. Mol. Biol.* **232**, 805–825.
37. Powers, R., Garret, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. & Clore, G.M. (1993). The high resolution three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry* **32**, 6744-6762.
38. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922–923.
39. Kabsch, W. & Sander, C. (1983). Dictionary of protein structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
40. Reeck, G.R., Jackson, P.J. & Teller, D.C. (1982). Domain structure in high molecular weight high mobility group nonhistone chromatin proteins. *Nature* **300**, 76–78.
41. McLachlan, A.D. (1971). Tests for comparing related amino-acid sequences: cytochrome c and cytochrome c551. *J. Mol. Biol.* **61**, 409–424.
42. Hill, C.P., Osslund, T.D. & Eisenberg, D. (1993). The structure of granulocyte-colony stimulating factor and its relationship to other growth factors. *Proc. Natl. Acad. Sci. USA.* **90**, 5167–5171.
43. Senda, T., *et al.*, & Mitsui, Y. (1992). Three-dimensional crystal structure of recombinant murine interferon-β. *EMBO J.* **11**, 3193–3201.
44. Samudzi, C.T., Burton, L.E. & Rubin, J.R. (1991). Crystal structure of recombinant rabbit interferon-γ at 2.7 Å resolution. *J. Biol. Chem.* **266**, 21791–21797.
45. Ealick, S.E., *et al.*, & Bugg, C.E. (1991). Three-dimensional structure of recombinant human interferon-γ. *Science* **252**, 698–702.
46. Arai, N., *et al.*, & Arai, K. (1989). Complete nucleotide sequence of the chromosomal gene for human IL-4 and its expression. *J. Immunol.* **142**, 274–282.
47. Miyatake, S., *et al.*, & Arai, K. (1985). Structure of the chromosomal gene for granulocyte-macrophage colony stimulating factor: comparison of the mouse and human genes. *EMBO J.* **4**, 2561–2568.
48. Degrave, W., *et al.*, & Fiers, W. (1983). Cloning and structure of the human interleukin 2 chromosomal gene. *EMBO J.* **2**, 2349–2353.
49. Kawasaki, E.S., *et al.*, & Mark, D.F. (1985). Molecular cloning of a complementary DNA encoding human macrophage-specific colony-stimulating factor (CSF-1). *Science* **230**, 291–296.
50. Tanabe, T., *et al.*, & Honjo, T. (1987). Molecular cloning and structure of the human interleukin 5 gene. *J. Biol. Chem.* **262**, 16580–16584.
51. Roskam, W.G. & Rougeon, F. (1979). Molecular cloning and nucleotide sequence of the human growth hormone structural gene. *Nucleic Acids Res.* **7**, 305–320.
52. Nagata, S., *et al.*, & Yamazaki, T. (1986). The chromosomal gene structure and two mRNAs for human granulocyte colony-stimulating factor. *EMBO J.* **5**, 575–581.
53. Vodjdani, G., Coulombel, C. & Doly, J. (1988). Structure and characterization of a murine chromosomal fragment containing the interferon beta gene. *J. Mol. Biol.* **204**, 221–231.
54. Lawn, R.M., *et al.*, & Goeddel, D.V. (1981). Human fibroblast interferon gene lacks introns. *Nucleic Acids Res.* **9**, 1045–1052.
55. Taya, Y., Devos, R., Tavernier, J., Cheroutre, H., Engler, G. & Fiers, W. (1982). Cloning and structure of the human immune interferon-gamma chromosomal gene. *EMBO J.* **1**, 953–958.
56. Miller, S., Janin, J., Lesk, A.M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656.
57. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
58. Baker, E.N. & Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**, 97–177.

Received: **29 Dec 1993**; revisions requested: **Jan 17 1994**; revisions received: **26 Jan 1994**. Accepted: **27 Jan 1994**.