doi:10.1016/j.jmb.2005.01.037 J. Mol. Biol. (2005) 347, 231–241

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

COMMUNICATION

A Survey of Left-handed Helices in Protein Structures

Marian Novotny[1,2] and Gerard J. Kleywegt[1]*

[1]*Department of Cell and Molecular Biology, Uppsala University, Box 596, SE-751 24 Uppsala, Sweden*

[2]*Linnaeus Centre for Bioinformatics, Biomedical Centre, Box 598, SE-751 24 Uppsala, Sweden*

**Corresponding author*

All naturally occurring amino acids with the exception of glycine contain one or more chiral carbon atoms and can therefore occur in two different configurations, L (*levo*, left-handed) and D (*dextro*, right-handed). Proteins are almost exclusively built from L-amino acids. The stereochemical bias of nature is further reflected at the secondary structure level where right-handed helices are strongly preferred over left-handed helices.

The handedness of helices has not received much attention in the past and is often overlooked during the analysis, description and deposition of experimentally solved protein structures. Therefore, an extensive survey of left-handed helices in the Protein Data Bank (PDB) was undertaken to analyse their frequency of occurrence, length, amino acid composition, conservation and possible structural or functional role.

All left-handed helices (of four or more residues) in a non-redundant subset of the PDB, were identified using hydrogen-bonding analysis, comparison of related structures, and experimental electron density assessment to filter out likely spurious and artefactual hits. This analysis yielded 31 verified left-handed helices in a set of 7284 proteins. The ϕ angles of the residues in the left-handed helices lie between 30°and 130° and the ψ angles lie between $-50°$ and 100°. Most of the helices are short (four residues) and for 87% of them, it was possible to determine that they are important for the stability of the protein, for ligand binding, or as part of the active site. This suggests that, even though left-handed helices are rare, when they do occur, they are structurally or functionally significant.

Four secondary structure assignment programs were tested for their ability to identify the handedness of the helices. Of these programs, only DSSP correctly assigns the handedness.

Keywords: left-handed helix; protein fold; protein structure; secondary structure; structure–function relations

Nature shows a profound right-left asymmetry.[1] DNA appears mainly in the right-handed *B*-conformation as do α-helices in proteins. This right-handed preference of biological macromolecules is a consequence of the selective incorporation of L-amino acids into proteins and of D-monosaccharides into DNA. The basis for this selectivity remains unknown, although a number of explanations have been proposed.[1]

α-Helices composed of L-amino acids are energetically more favourable in a right-handed conformation than in the left-handed mirror image of this arrangement due to steric hindrance between side-chain atoms and the main-chain carbonyl moiety.[2] Conversely, D-amino acids will form more stable α-helices with a left-handed than with a right-handed conformation. This phenomenon is perhaps most convincingly illustrated by the fact that the structures of all-D proteins, such as D-rubredoxin[3] and D-monellin,[4] are the mirror images of the corresponding all-L proteins.

α-Helices are characterised by a typical hydrogen bonding pattern in which the main-chain carbonyl oxygen of residue *i* forms a hydrogen bond with the main-chain amide hydrogen of residue $i+4$. Other types of helices can also occur in proteins, namely 3_{10} and π-helices with $(i, i+3)$ and $(i, i+5)$ hydrogen bonds, respectively. 3_{10}-Helices occur fairly often in proteins, although they are frequently shorter than α-helices.[5] π-Helices are relatively rare motifs, but when they do occur they can be of

Abbreviations used: PDB, Protein Data Bank; EDS, Electron Density Server.

E-mail address of the corresponding author: gerard@xray.bmc.uu.se

functional importance due to their unique structural parameters.[5] To our knowledge, the handedness of helices in proteins has not been studied systematically, and not much is known about left-handed helices other than that they are very rare.

Stretches of amino acids with unusual backbone conformations (e.g. left-handed helices) often appear at ligand-binding sites, protein–protein interfaces or other functional sites. It has been suggested that proteins may sacrifice a part of their stability to form an effective functional site.[6,7] It has further been suggested that searches for regions with unusual backbone conformations could be used for the annotation of novel protein structures, since such regions are candidates for being functional sites.[7]

Little is known about how residues with left-handed helical conformations affect the stability of proteins. It is often assumed that such residues (with the exception of glycine residues) would suffer steric clashes between main-chain and C^{β} atoms, thereby reducing the stability of the protein.[2] However, Takano *et al.* found that five of six non-glycine left-handed residues in lysozyme do not significantly impair the stability of lysozyme.[8] On the other hand, these residues are scattered throughout the lysozyme sequence and do not form a left-handed helix. Scattered amino acids with a left-handed helical conformation also occur in type I″ turns, where two amino acids have this conformation, and in helix stop signal[9] and Schellman motifs.[10]

To our knowledge, there are currently only three protein structure entries in the PDB that have a left-handed helix assigned. Two of these are four residues long and they are found in thermolysin (PDB code 8tln)[11] and neutral protease (1npc).[12] The third one is a three-residue long helix in granulocyte-colony-stimulating factor (1rhg).[13] A PubMed search† yielded one more left-handed helix, namely in spinach glycolate oxidase (1gox).[14] Interestingly, this left-handed helix is a part of the active site of the enzyme. The left-handed helices in glycolate oxidase and granulocyte-colony-stimulating factor both have a hydrogen-bonding pattern consistent with a 3_{10}-helix.

A few years ago we discovered a left-handed helix that plays an important role in alanine racemase[15] and that had previously been classified incorrectly as right-handed.[16] We also noted at the time that several secondary structure assignment programs failed to annotate the handedness of this helix. These findings prompted us to undertake a survey of left-handed helices in the PDB with respect to their frequency of occurrence, their length, their sequence characteristics (if any), their conservation and, in particular, their possible functional or structural roles. We also investigated in more detail if (and how successfully) four secondary structure assignment programs assign the handedness of left-handed helices.

† http://pubmed.gov/.

Detection and verification of left-handed helices

Initially, ideal left-handed helices of different lengths (four to ten residues) were constructed and used as templates in SPASM searches[15] against subsets of the PDB.[17] The hits obtained in these searches were used to define putative left-handed helices as being continuous stretches of at least four residues whose ϕ angles are all between 30° and 130° and whose ψ angles are all between −50° and 100°. Subsequently, a non-redundant subset of the PDB (version of September, 2003) was generated with the PISCES server.[18] To produce a large enough subset, relaxed criteria were used in the generation of the subset: none of the pairs of protein chains had more than 90% sequence identity, and all crystal structures with a resolution better than 3.5 Å and an *R*-value less than 0.4 were included, and so were all proteins whose structure had been solved by nuclear magnetic resonance (NMR) spectroscopy. The minimum chain length was set to ten amino acid residues. The resulting subset contained 7284 protein chains from 6535 PDB entries and included 1,687,315 amino acids. A Perl program was written to find all instances of left-handed helices (according to the definition above) in this subset. Initially, wider ranges for the torsion angles were used (ϕ between −20°and 145° and ψ between −70° and 145°), and the hits that were obtained were examined to check if their hydrogen-bonding pattern was compatible with left-handed α or 3_{10}-helices. The final ranges yielded no false negatives and relatively few false positives.

If any part of a structure satisfied our definition of a left-handed helix it was designated a putative hit. The hits were visualised with Deep View,[19] and hydrogen-bonding patterns were analysed with Deep View, DSSP[20] and HBPLUS.[21] Putative left-handed helices were subjected to additional checks to validate them or to dismiss them as probable artefacts. Any hits in NMR structures were only accepted if the left-handed helix occurred in at least 50% of the models of the ensemble. For those crystal structures for which electron-density maps were available from the Uppsala Electron Density Server (EDS),[22] the density for the putative helices was inspected. In cases where no map was available (including all hits found in NMR structures), the PDB was checked for structures of the same or related proteins and, if any were found, it was checked if the corresponding residues in the related structures had similar ϕ and ψ angles as the residues in the putative left-handed helix. If no related structures could be found in the PDB, we trusted the authors and accepted the hit as a true left-handed helix.

For each of the accepted left-handed helices, many sources of information (literature, contacts

with authors, and bioinformatics resources such as SWISS-PROT,[23] ProSite,[24] PDBsum[25] and Omim)† were consulted to find information about the possible functional or structural significance of the residues in the helix.

We also investigated if the left-handed helices are recognised as such by the secondary structure assignment tools DSSP,[20] STRIDE,[26] Promotif[27] and SecStruct,[5] and in the annotation in the original PDB files. The frequencies of occurrence of the 20 natural amino acids in the left-handed helices were calculated for comparison with their frequencies in the entire subset of the PDB. The sequences of the left-handed helices were used to search a 95% sequence identity subset of the PDB with the PATINPROT server[28] to find out if identical sequences occur in any other proteins whose structure is known. Any hits were then analysed to determine if the sequence displayed strong secondary structure preferences.

We also generated a non-redundant subset of the PDB with PISCES[18] using a 25% cut-off for the sequence identity (all other criteria were the same as in the generation of the 90% subset) and located all left-handed helices in that subset. For comparison purposes, the set of proteins from the 90% subset that contained a left-handed helix was also pruned with PISCES to yield a set of left-handed-helix containing proteins with no more than 25% sequence identity.

Analysis of the left-handed helices in the PDB

Ideal left-handed helices of four to ten residues were used in SPASM searches for similar motifs in the PDB. This yielded 21 left-handed helices that were used to define the initial ranges of ϕ and ψ angle values for residues in left-handed helices, namely 0° to 125° for ϕ and −50° to 75° for ψ. Later, these ranges were extended (−20° to 145° for ϕ and −70° to 145° for ψ) to make sure that no true left-handed helices would be missed (false negatives). Eventually, the ranges were narrowed down to the final values (30° to 130° for ϕ and −50° to 100° for ψ), which ensured that there were no false negatives, and not too many false positives. A minimum length of four residues was imposed on potential left-handed helices to ensure that any such helices would contain at least one full turn.

The criteria for detecting left-handed helices were implemented in a Perl program that was run on all the members of the non-redundant subset of the PDB described earlier, an exercise that yielded 56 putative hits. All hits in structures determined by NMR were required to have a left-handed helical conformation in more than half of the models in the ensemble (to filter out spurious hits with little support in the experimental NMR data), which reduced the number of hits to 38. In six of the 38 cases, one or more closely related structures were available in which no support could be found for any left-handed helical conformation, and in one case there was no support in the electron density (although the structure had been determined at 1.9 Å resolution and refined to a crystallographic *R*-value of 0.2). Finally, one entry was accepted in good faith. After these validation steps, there remained 31 hits that were deemed to be genuine left-handed helices (Table 1). Based on hydrogen-bonding analysis, the left-handed helices were further divided into α-helices (11 cases) and 3_{10}-helices (20 cases). Two of the hits occurred in NMR structures.

The left-handed helices were short; the longest helix was six residues long, but the majority were just four residues long (Table 1). The distribution of the ϕ and ψ torsion angle values of all the left-handed helices is shown in Figure 1. For α-helices, the average values of ϕ and ψ were 59° (σ=12°) and 42° (σ=13°), respectively; for 3_{10}-helices, these values were 67° (σ=21°) and 23° (σ=25°), respectively. The 3_{10}-helices thus had a wider distribution of ϕ and ψ angles than the α-helices. The observed torsion angle ranges for the left-handed helices are similar to those defined by Gunasekaran *et al.* (20° to 125° for ϕ and −45° to 90° for ψ).[9] The average ϕ and ψ angles for the left-handed helices are also similar to those observed for right-handed helices (ignoring the sign changes).[29] The proteins that contain left-handed helices show no preference for overall secondary structure contents; they belong to the classes of mainly alpha, mainly-beta and mixed alpha-beta proteins. Most of the left-handed helices are located on the protein surface, but there are no obvious patterns in the types and spatial orientations of flanking secondary structure elements.

All amino acid types except proline were encountered in left-handed helices. Proline residues are unlikely to appear in helices in general, because their imido nitrogen atom cannot donate a hydrogen bond and because the bulky C^{δ} methylene group attached to the nitrogen introduces severe steric clashes. Moreover, due to the inability of proline residues to assume positive ϕ values, they are not expected to occur in any left-handed helices at all, and this expectation is borne out by our observations. The second position in the helix is occupied by asparagine in 11 cases and by glutamate or glutamine in six more cases. The last two positions in the helix are often occupied by two identical residues, usually Gly-Gly (14 cases), and an even higher prevalence of glycine is apparent at the last position in the helices (19 cases). This is most likely because glycine is achiral and lacks a side-chain, so that left-handed and right-handed conformations are equally favourable. Although the sample size is very small, tryptophan has a surprisingly high propensity for being in left-handed helices, especially compared to its propensity for assuming a left-handed backbone

† http://www.ncbi.nlm.nih.gov/omim/

Table 1. Left-handed helices encountered in a subset of the PDB

PDB code and chain ID[a]	Protein name	CATH classification[b]	Residues	Sequence	Helix type[c]	Reference
1bd0A	Alanine racemase	2.40.37.10	40–44	ANAYG	α	16
1autL	Activated protein C	2.10.25.10	101–104	DNGG	3_{10}	34
1b9wA	Merozoite surface protein 1 (*P. cynomolgi*)	–	52–55	KNGG	3_{10}	38
1g2lB	Coagulation factor X	2.10.25.10	258–261	DNGD	3_{10}	59
1kliL	Coagulation factor VII	2.10.25.10	94–97	ENGG	3_{10}	60
1n1iA	Merozoite surface protein 1 (*P. knowlesi*)	–	57–60	NNGG	3_{10}	61
1ob1C	Merozoite surface protein 1 (*P. falciparum*)	–	52–55	NNGG	3_{10}	62
1rfn_	Coagulation factor IX	2.10.25.10	91–94	KNGR	3_{10}	63
1pb5A	Lnr module from Notch	–	28–32	GWDGG	3_{10}	44
1kdgA	Cellobiose dehydrogenase	–	532–535	YENW	3_{10}	45
1hzmA	Protein phosphatase 6	–	61–64	IMLR	α	46
1gtxA	4-Aminobutyrate aminotransferase	3.30.70.160	70–73	SQIS	α	49
1qj5A	7, 8-Diaminopelargonic acid synthase	3.30.70.160	50–53	SSWW	3_{10}	48
2gsaA	Glutamate semialdehyde aminotransferase	3.30.70.160	65–68	GTWG	3_{10}	47
2oatA	Ornithine aminotransferase	3.30.70.160	83–86	SSYS	3_{10}	50
1bnlA	Endostatin (*Homo sapiens*)	3.10.100.10	135–138	CETW	α	52
1dy2A	Endostatin (*M. musculus*)	3.10.100.10	207–210	CEAW	α	54
1koeS	Endostatin (*M. musculus*)	3.10.100.10	266–269	CETW	α	53
1bqbA	Aureolysin	1.10.390.10	223–226	DNGG	α	64
1npcE	Neutral protease	1.10.390.10	227–230	DNGG	α	12
8tlnE	Thermolysin	1.10.390.10	226–229	DNGG	α	11
1j9qA	Nitrate reductase (*A. feacalis*)	2.60.40.240	105–108	ALGG	3_{10}	65
1nifA	Nitrate reductase (*A. cycloclates*)	2.60.40.240	105–108	ALGG	3_{10}	55
1oe1A	Nitrate reductase (*A. xylosoxidans*)	2.60.40.240	99–102	ALGG	3_{10}	66
1ptmA	4-Hydroxythreonine-4-phosphate dehydrogenase	–	211–216	HAGEGG	3_{10}	56
1ak0E	P1 nuclease	1.10.575.10	131–134	AVGG	α	57
1mzr_	2, 5-Diketo-D-gluconate reductase	–	191–194	AQGG	3_{10}	–[d]
1h21A	Split-soret cytochrome *C*	–	77–81	GGISD	3_{10}	67
1hxxA	Ompf porin	2.40.160.10	143–146	NFFG	3_{10}	68
1jv1A	Glcnac1p uridyltransferase	–	182–185	KYFG	3_{10}	69
1kwsA	Beta-1,3-glucuronyltransferase 3	3.90.550.10	298–301	AANC	α	70

[a] The four-character PDB code, followed by the chain identifier (where an underscore signifies a blank chain identifier). The entries are sorted in the order in which they are discussed in the text (i.e. grouped by function). Entries 1hzm and 1pb5 are NMR structures; all other structures were determined by X-ray crystallographic methods.
[b] The CATH[33] classification; a dash indicates that the PDB entry had not been classified by CATH yet.
[c] The type of the left-handed helices, as assigned manually based on hydrogen-bonding patterns.
[d] C. Abergel, S. Jeudy, V. Monchois and J.M. Claverie, Structure of DkgA from *Escherichia coli* at 2.13 Å resolution solved by molecular replacement (unpublished results).

conformation. Some additional data can be found on our web site†.

Sometimes right-handed helices have a Schellman turn as a C-cap.[10,30] In such a motif, the helix is terminated by a residue with a left-handed conformation, and 5–2 or 6–1 hydrogen bonds are usually formed. Of the present set of left-handed helices, 13 contain a left-handed version of the Schellman turn. Whenever two consecutive residues have a helical conformation of opposite handedness (i.e. RL or LR), they form a motif called a nest.[31] Analysis of the 31 left-handed helices and their flanking residues revealed that six of them contain both an RL nest at their N terminus and an LR nest at their C terminus, whereas ten of them contain neither. One helix only contains an RL nest, whereas 14 helices contain only an LR nest.

To assess if the short amino acid sequences of the left-handed helices always occur in such secondary structure elements, all occurrences of each of the 31 sequences were located in a 95% sequence-identity subset of the PDB with the PATINPROT server.[28] As expected,[32] none of the sequences found in the left-handed helices showed any strong secondary structure preferences. Most were found to occur in all three secondary structure types (helix, strand and loop), e.g. the sequence AQGG originally found in P_1 nuclease (PDB code 1ak0) as a left-handed helix was found in ten other proteins in the PDB subset where it appeared as a helix, a strand, a loop and combinations thereof. The sequence of the longest left-handed helix (HAGEGG) was not found in any other protein in the PDB.

The extent to which left-handed helices are

† http://xray.bmc.uu.se/~marian/left

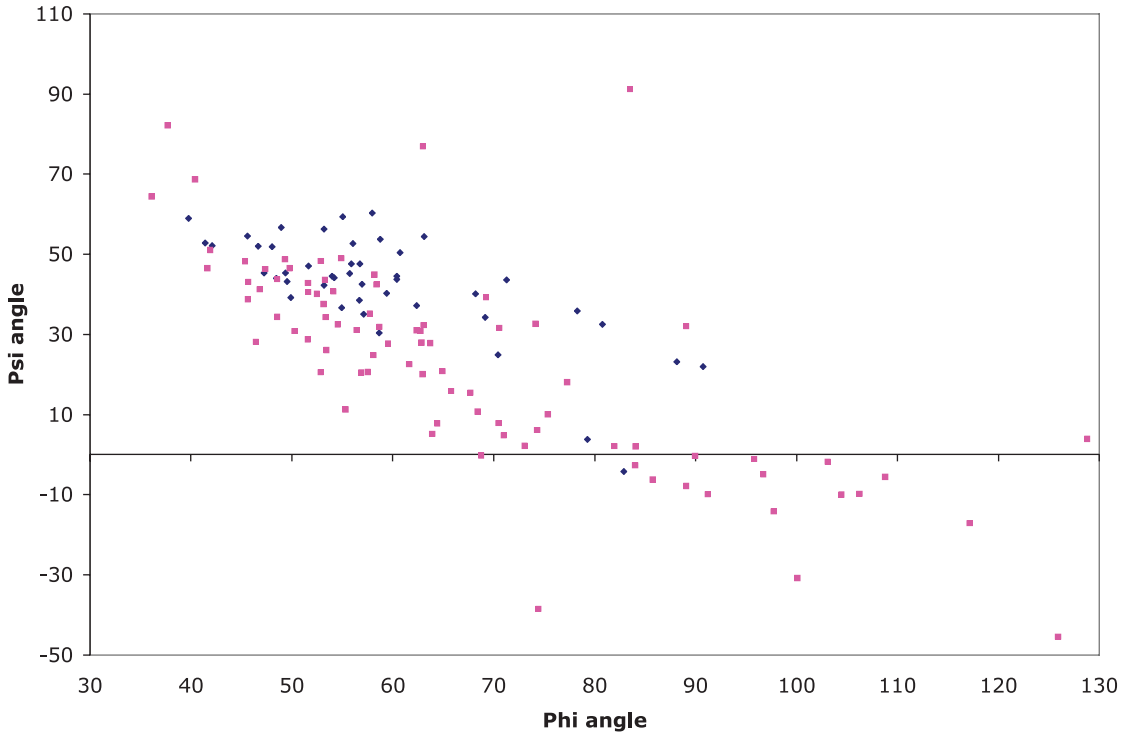

Figure 1. Distribution of ϕ and ψ torsion angles of residues found in left-handed helices. Pink squares represent residues in 3_{10}-helices and blue diamonds represent residues in α-helices. (For interpretation of the references to colour in this Figure legend, the reader is referred to the web version of this article.)

conserved between different sequence families at the same homologous superfamily level of the CATH classification[33] was also investigated. It was found that none of the helices that occurred in proteins that had been classified in CATH are conserved across all corresponding sequence families. This suggests that the left-handed helix motifs have evolved relatively recently and that they serve a specific purpose.

To assess the effect of using different cut-offs for the allowed level of sequence identity during the generation of reduced subsets of the PDB, the analysis was repeated with a 25% subset. Whereas the 90% subset yielded 31 left-handed helices, only 13 such helices were found in the 25% subset. However, if the set of 31 proteins that contain a left-handed helix is in turn reduced to a subset of proteins that have no more than 25% sequence identity, the number of hits is as high as 18. This discrepancy can be explained by realising that the two processes of selecting a subset of the PDB using a cut-off on the allowable sequence identity level and that of testing a set of proteins for a particular property (in this case, the presence of left-handed helices) are not commutative. In this case, 13 is the number of left-handed helices found in a 25% subset of the PDB, whereas 18 is the size of a 25% PDB subset of the proteins that are known to contain a left-handed helix.

Finally, four secondary structure assignment programs (DSSP, STRIDE, PROMOTIF and SecStruct) were tested for their ability to recognise the handedness of left-handed helices (Table 2). In addition, the assignments of these helices in the parent PDB files were retrieved. Significant differences in the results of the programs were observed. In fact, in none of the cases did the programs produce an identical result. Only two of the programs, DSSP and SecStruct, provided any information at all about the handedness of the secondary structure elements (both of them in a column labelled "chirality" in the output). The results obtained with the oldest program, DSSP, agree best with our own manual assignments. Good agreement was also obtained with the results of SecStruct, but the chirality assignments of this program were often misleading. Interestingly, the handedness of the helices appears to have escaped the attention of most of the depositors and annotators of the structures. As far as we could determine, only four of the 31 left-handed helices were described as such in the corresponding PDB file (1npc, 8tln) or original papers (1b9w, 1n1i).

Function of left-handed helices

For each of the 31 left-handed helices it was investigated if they are important for the structure or function of the parent protein. The results are as follows (the hits have been grouped according to sequence and structure similarity):

Alanine racemase

This enzyme mediates the interconversion of L and D-alanine, which is indispensable for cell wall

Table 2. Comparison of secondary structure assignments of left-handed helices by different methods

PDB code and chain ID[a]	Type [b]	PDB [c,d]	PROMOTIF [d,e]	DSSP [d,e]	STRIDE [d,e]	SECSTR [d,e]
1bd0A	H	HHHHH	HHHHH	HHHTT (L)	TGGGC	HHHH- (R)
1autL	G	GGGG	GGGG	GGGG (L)	GGGG	GGGG (L)
1b9wA	G	GGGG	GGGG	GGGG (L)	GGGG	GGG- (L)
1g2l_	G	GGGG	GGGG	GGGG (L)	GGGG	GGG- (L)
1kliL	G	GGGG	GGGG	GGGG (L)	GGGG	GGG- (L)
1n1iA	G	GGGG	GGGG	GGGG (L)	GGGG	GGG- (L)
1ob1C	G	GGGG	GGGG	GGGG (L)	GGGG	GGGG (L)
1rfn_	G	GGGG	GGGG	GGGG (L)	GGGG	GGGG (L)
1pb5A	G	GGGG	GGGG	GGGG (L)	GGGC	–
1kdgA	G	GGGG	GGGG	GGGG (L)	GGGG	GGGG (L)
1hzmA	H	HHHH	HHHH	HHHH (L)	TTTT	HHHH (L)
1gtxA	H	HHHH	HHHH	HHHH (L)	HHHH	HHHH (L)
1qj5A	G	–	–	TTTT (L)	HCCC	–
2gsaA	G	GGGG	HHHT	GGGT (L)	GGGT	–
2oatA	G	HHHH	GGGG	GGGG (L)	GGGG	GGGG(R)
1bnlA	H	HHHH	HHHH	HHHH (L)	TTTT	HHHH (L)
1dy2A	H	HHHH	HHHH	HHHH (L)	GGGT	HHHH (L)
1koeS	H	HHHH	HHHH	HHHH (L)	TTTT	HHHH (L)
1bqbA	H	HHHH	HHHH	HHHH (L)	GGGC	HHHH (R)
1npcE	H	HHHH/left	HHHH	HHHH (L)	GGGC	HHHH (R)
8tlnE	H	HHHH/left	HHHH	HHHH (L)	GGGC	HHHH (R)
1j9qA	G	GGGG	GGGG	GGGG (L)	GGGC	GGGG(R)
1nifA	G	GGGG	GGGG	GGGG (L)	GGGG	GGGG(R)
1oe1A	G	GGGG	GGGG	GGGG (L)	GGGG	GGGG(R)
1ptmA	G	GGGG–	GGGGGG	GGGGGG (L)	GGGGGG	GGGGGG (L)
1ak0E	H	HHH-	HHHT	HHHT (L)	TGGG	HHHH (L)
1mzr_	G	–	TTTT	TTTT (L)	TTTT	–
1h21A	G	GGG–	GGTTT	GGTTT (L)	GGTTT	G– (R)
1hxxA	G	HHHH	HHHH	HHHH (L)	TGGG	–
1jv1A	G	GGGG	GGGG	GGGG (L)	GGGG	GGGG (L)
1kwsA	H	HHHH	HHHH	HHHH (L)	TTTT	HHHH (L)

[a] The four-character PDB code is listed, followed by the chain identifier (where an underscore signifies a blank chain identifier). The entries are listed in the same order as in Table 1.
[b] The type of the left-handed helices is listed as assigned manually based on hydrogen-bonding patterns (H = α-helix; G = 3_{10}-helix).
[c] Secondary structure assignment for the residues in each left-handed helix in their parent PDB file.
[d] –, no secondary structure type was assigned, H is α-helix, G 3_{10} -helix, T turn, and S bend. The assigned handedness (if reported) is listed in parentheses as R for right-handed and L for left-handed.
[e] Secondary structure assignment for the residues in each left-handed helix according to the programs listed.

formation. Alanine racemase (PDB code 1bd0) was already known to contain a left-handed helix.[15] The catalytic residue is Lys39 and the left-handed α-helix covers residues 40 to 44, where Tyr43 is the covalent attachment site for the cofactor pyridoxal phosphate (Figure 2(a)). The carbonyl oxygen atom of Gly44 forms an interdomain hydrogen bond with Arg366.[16] The entire helix is part of the PROSITE pattern (PS00395) for the alanine racemase family.

EGF-like family

A left-handed helix was identified in seven proteins with EGF-like domains (PDB codes 1aut, 1b9w, 1g2l, 1kli, 1n1i, 1ob1 and 1rfn). Four of these are blood-clotting proteins (factor VII, factor IX, factor X and protein C) and three are merozoite surface proteins from various *Plasmodium* species. These two functionally distinct groups show weak but detectable sequence similarity, mainly due to conserved cysteine residues that form disulphide bridges, but also in a short stretch of sequence containing the left-handed helix. The hydrogen-bonding pattern of these helices classifies them as 3_{10}-helices rather than α-helices.

The seven proteins have a conserved Asn residue in their left-handed helices that forms an important hydrogen bond[34] either with residue $i+5$ (protein C and factor VII) or $i+7$ (merozoite surface proteins) in the EGF-like domain, or with residues in the protease domain (factor IX and factor X). There is also some evidence that the left-handed helix in factor IX is involved in the binding of factor VIII, which is its natural interaction partner in the blood-clotting cascade. Double mutation of Asn89 and Asn92 to alanine completely abolishes the interaction between factor IX and factor VIII.[35] However, the single mutation of Asn89 to alanine has only a marginal effect on binding.[36,37]

The two consecutive glycine residues in the left-handed helix have been implicated as protein-stabilising factors in the merozoite surface proteins.[38] The first amino acid in the helix in protein C (Asp101) participates in an unusual hydrogen bonding interaction with another acidic residue (Glu85).[39] Mutations of three residues in the left-handed helix, namely Asp101, Gly103 and Gly104, cause protein C deficiency which may lead to recurrent venous thrombosis, which in turn may cause neonatal death.[40–42]

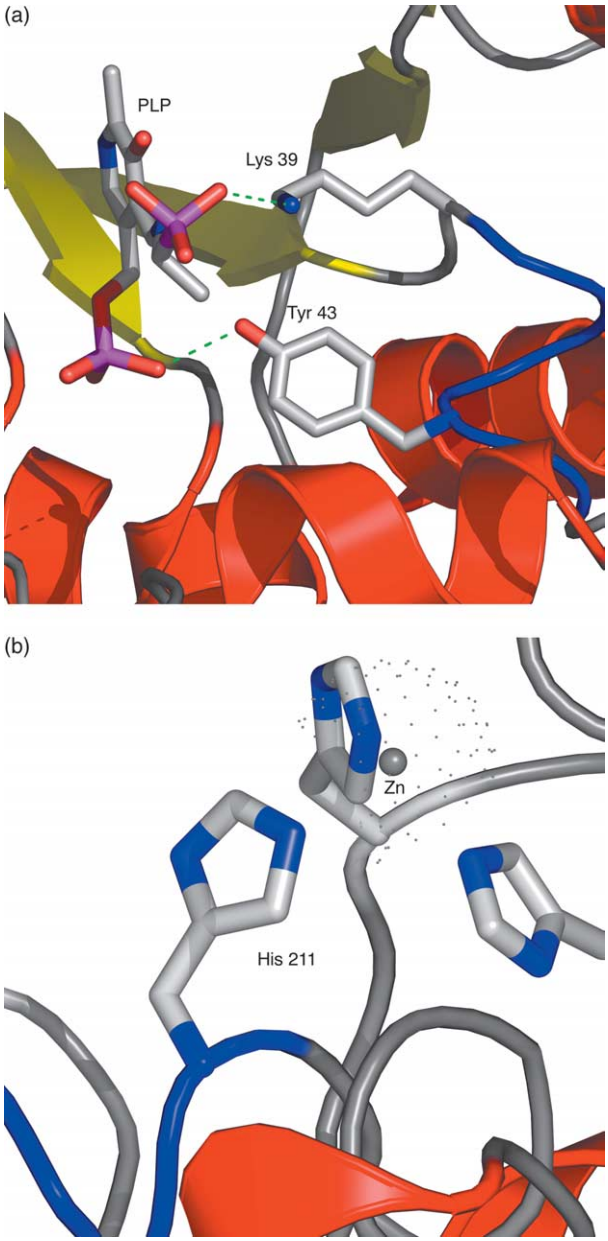

Figure 2. (a) The left-handed helix (in blue) in alanine racemase (PDB code 1bd0) directly follows the active site residue Lys39. Moreover, Tyr43 in the left-handed helix binds the cofactor pyridoxal phosphate (PLP). Both panels of this Figure were created with PyMol[58] (http://www.pymol.org/). (b) His211 in the left-handed helix (in blue) of PdxA (PDB code 1ptm) participates in the coordination of a Zn^{2+}, together with two histidine residues from the other subunit of this protein. (For interpretation of the references to colour in this Figure legend, the reader is referred to the web version of this article.)

Lin12-Notch repeat module of human Notch protein

Notch is a transmembrane receptor for many signals that initiate conserved signalling pathways in a variety of cellular responses (e.g. cellular growth or differentiation). Notch consists of several domains, one of them being the Lin 12-Notch repeat module (LNR), which has been suggested to keep the Notch protein in the idle state.[43] A left-handed 3_{10}-helix (residues 28–32) was identified in this module (PDB code 1pb5). It appears to be important for folding, because Asp30 and Asp33 (immediately following the helix) participate in the coordination of a Ca^{2+}, whose binding is highly specific and necessary for folding of LNR.[44] The hydrophobic residue Trp29 lies on the surface of the module and may be involved in inter-domain contacts with a second LNR.[44]

Flavin domain of cellobiose dehydrogenase

Cellobiose dehydrogenase (CDH) is a lignin and cellulose-decomposing enzyme. Residues 532 to 535 in the flavin domain of this protein (PDB code 1kdg) form a left-handed 3_{10}-helix and line the entrance to the active site of the protein, although they are not in contact with the substrate.[45] It has been suggested that the left-handed helical segment could interact with the cytochrome domain of CDH (C. Divne, personal communication).

Erk-2 binding domain of MAPK phosphatase MPK-3

MPK-3 negatively regulates the activity of mitogen-activated protein kinases (MAPK) by dephosphorylating their threonine and tyrosine residues. MPK-3 consists of a C-terminal catalytic domain and an N-terminal domain that is responsible for binding of MAPK. Residues 61–64 (PDB code 1hzm) have a left-handed α-helical conformation, where residues Leu63, Arg64 and Arg65 are critical for the binding of MAPK.[46] Residues Ile61 and Met62 stabilise the conformation of the helical segment by interacting with the hydrophobic core of the protein.[46]

Type I PLP-dependent aspartate aminotransferase-like fold

Four proteins with this fold were found to contain a left-handed helix: ornithine amino transferase (PDB code 2oat), aminobutyrate aminotransferase (1gtx), glutamate-1-semialdehyde aminomutase (2gsa) and 7,8-diaminopelargonic acid synthase (1qj5). The third residue in the four-residue long α-helix (1gtx) or 3_{10}-helix (the other three structures) contributes to the substrate specificity of the enzymes.[47–50] A structural segment, equivalent to the left-handed helical region of these transferases, in another protein with a type I PLP-dependent aspartate aminotransferase-like fold, namely dialkylglycine decarboxylase (PDB code 1d7r), shows some similarity with them. Three of the four residues (29–32) have positive ϕ and ψ values, but the third residue lies within the right-handed helical region of the Ramachandran plot.[51] Hence, this segment of the protein does not form a left-handed helix, but a turn. Interestingly, the third

residue is the substrate-binding residue in this family of proteins as well.

Endostatins

Left-handed α-helical segments (with identical sequences) were identified in three different endostatins (PDB codes 1bnl, 1dy2, 1koe). A conserved cysteine residue in these helices participates in a disulphide bridge that is crucial for the stability of these proteins.[52–54]

Proteins from peptidase family M14

A left-handed α-helix was found in three related proteases, namely, in thermolysin (PDB code 8tln), aurolysin (1bqb) and neutral protease (1npc), and the sequence of the helix is identical in all three. The side-chain of the first amino acid in the helix, aspartate, forms a hydrogen bond with a histidine that coordinates the zinc ion in the active site.[12] There is also a hydrogen bond between the backbone of this aspartate and the zinc-coordinating histidine, which further underscores the importance of the helical segment for the catalytic mechanism of these proteases.

Nitrate reductase

Nitrate reductases mediate the second reaction in the N_2-fixation pathway, the reduction of nitrate to NO. Three bacterial nitrate reductases were found to contain a left-handed 3_{10}-helix (PDB codes 1j9q, 1nif and 1oe1). The second amino acid in the helix is a leucine residue that is part of the active site pocket where it also hydrogen bonds to an ordered water molecule.[55]

4-Hydroxythreonine-4-phosphate dehydrogenase (PdxA)

PdxA (PDB code 1ptm) is a homodimeric enzyme in the pathway leading to pyridoxal 5′-phosphate, an essential cofactor for many enzymes in amino acid metabolism. PdxA contains the longest left-handed helix encountered in this survey, a six-residue long 3_{10}-helix that is located in the dimer interface. A histidine residue in the helix (His211), together with two histidine residues from the other monomer, coordinates a Zn^{2+} that is required for catalytic activity of the enzyme (Figure 2(b)).[56] After the substrate 4-hydroxy-l-threonine-phosphate (HTP) has entered the protein, a hydrogen bond is formed between the O^{γ} oxygen atom of HTP and the $N^{\delta 2}$ atom of His211. Thus, this histidine residue participates in both metal-binding and substrate-binding. It is conserved in all 74 known sequences of the PdxA family[56].

Nuclease P_1

Nuclease P_1 (PDB code 1ak0) is a phosphodiesterase that degrades both single-stranded RNA and DNA. The enzyme contains a short left-handed α-helix (Ala131 to Gly134) that forms a part of the active site cleft of the enzyme. Asn135, immediately following the helix, forms a hydrogen bond with the substrate.[57] This enzyme contains zinc ions, and one of them is coordinated by His126, which lies five residues upstream of the left-handed helix.

2,5-Diketo-D-gluconic acid reductase (Dkg A)

DkgA (PDB code 1mzr) is an enzyme in the metabolic pathway to vitamin C. A short 3_{10}-helix was identified involving amino acids 191 to 194. Leu190 of this enzyme coordinates the phosphate group of the cofactor NADPH (S. Jeudy, personal communication).

Remaining cases

For the remaining proteins that contain a left-handed helix (PDB codes 1hxx, 1kws, 1jv1 and 1h21) no indication could be found as to their possible structural or functional roles.

Concluding remarks

An extensive survey of left-handed helices in known protein structures has been carried out. The results show that left-handed helices are rare motifs in protein structures (31 cases in 7284 proteins). Amino acid residues in these 31 left-handed helices have ϕ angles between 30° and 130° and ψ angles between −50° and 100°. These ranges agree well with previous definitions.[9] Most of the left-handed helices were only four residues in length, and they occur as both α-helices and 3_{10}-helices. Most of them could be shown to play an important role either for the stability or for the function of the protein (e.g. determining substrate specificity, mediating protein–protein interactions, or binding a cofactor of an enzymatic reaction). Consideration of the handedness of helices could therefore help pinpoint functionally interesting parts of the protein. This can become especially important as we enter the structural genomics era and the need for automatic annotation of structures grows, in particular for structures of proteins of unknown function.

Our results also show that secondary structure assignment is not a trivial problem and that different answers can be obtained for a given piece of structure when different programs are used. The same is true for the designation of hydrogen bonds, where discrepancies between the results of DSSP, Deep View and HBPLUS were observed (data not shown).

The Uppsala Electron Density Server[22] was a helpful tool to assess which hits were genuine and which were artefactual left-handed helices. Of the 56 putative left-handed helices that were identified, 33 occurred in crystal structures and for 19 of these electron density maps were available from EDS.

Thus, even low-resolution structures (that are otherwise often excluded in the generation of subsets through strict criteria for resolution or *R*-value) could be included in the PDB subset. Conversely, one of the putative left-handed helices was rejected on the basis of its poor electron density, even though the parent structure had been determined at 1.9 Å resolution.

This study was concerned exclusively with left-handed helices (i.e. motifs of at least four consecutive residues that form at least one turn). However, a cursory analysis of motifs that consist of only three consecutive residues with left-handed helical conformations was also carried out (data not shown). About 600 such motifs were identified in the 90% sequence-identity subset used throughout this study. The large number of hits precluded analysis and validation at the same level of detail and accuracy that was used in the rest of this study. However, the distribution of torsion angles and amino acid frequencies were examined. The distribution of ϕ and ψ torsion angles turned out to be similar to the one observed for the residues in left-handed helices. The amino acid frequencies are more similar to the frequencies observed for the amino acids in left-handed conformations than those in the verified left-handed helices. For a random subset of 52 of the three-residue motifs, taken exclusively from X-ray structures, the secondary structure assignment of the motifs and their neighbouring residues according to Promotif[25,27] was inspected. Most of the motifs had been classified as 3_{10}-helices (33 instances), but turns were also frequent (14 instances), and five of the motifs were considered α-helical. A few of these motifs had been annotated in PDBsum as part of an active, metal-binding or ligand-binding site or were found in close proximity to such a site. Therefore, while annotating structures, it would appear to be useful to carefully examine any motif with a left-handed conformation or, indeed, any motif with unusual main-chain conformation.[6,7]

Acknowledgements

The authors thank Professor T. Alwyn Jones (Uppsala University) for a critical reading of the manuscript, and Terese Bergfors (Uppsala University) for indispensable linguistic comments and corrections. We also thank an anonymous referee for several useful suggestions. M.N. is supported through grants from the Linnaeus Centre for Bioinformatics (Uppsala) and Uppsala University. G.J.K. is a Royal Swedish Academy of Sciences (KVA) Research Fellow, supported through a grant from the Knut and Alice Wallenberg Foundation. He is further supported by Uppsala University, the Linnaeus Centre for Bioinformatics, and the Swedish Structural Biology Network (SBNet).

References

1. Cintas, P. (2002). Chirality of living systems: a helping hand from crystals and oligopeptides. *Angew. Chem. Int. Ed. Engl.* **41**, 1139–1145.
2. Brändén, C.-I. & Tooze, J. (1999). *Introduction to Protein Structure*, Garland Publishing, New York.
3. Zawadzke, L. E. & Berg, J. M. (1993). The structure of a centrosymmetric protein crystal. *Proteins: Struct. Funct. Genet.* **16**, 301–305.
4. Hung, L. W., Kohmura, M., Ariyoshi, Y. & Kim, S. H. (1998). Structure of an enantiomeric protein, D-monellin at 1.8 Å resolution. *Acta Crystallog. sect. D*, **54**, 494–500.
5. Fodje, M. N. & Al-Karadaghi, S. (2002). Occurrence, conformational features and amino acid propensities for the pi-helix. *Protein Eng.* **15**, 353–358.
6. Herzberg, O. & Moult, J. (1991). Analysis of steric strain in the polypeptide backbone of protein models. *Proteins: Struct. Funct. Genet.* **11**, 223–229.
7. Petock, J. M., Torshin, I. Y., Weber, I. T. & Harrison, R. W. (2003). Analysis of protein structures reveals regions of rare backbone conformation at functional sites. *Proteins: Struct. Funct. Genet.* **53**, 872–879.
8. Takano, K., Yamagata, Y. & Yutani, K. (2001). Role of non-glycine residues in left-handed helical conformation for the conformational stability of human lysozyme. *Proteins: Struct. Funct. Genet.* **44**, 233–243.
9. Gunasekaran, K., Nagarajaram, H. A., Ramakrishnan, C. & Balaram, P. (1998). Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals. *J. Mol. Biol.* **275**, 917–932.
10. Schellman, C. (1980). The αL-conformation at the ends of helice. In *Protein Folding* (Jaenicke, R., ed.), pp. 53–61, Elsevier/New Holland, New York.
11. Holmes, M. A. & Matthews, B. W. (1982). Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol.* **160**, 623–639.
12. Stark, W., Pauptit, R. A., Wilson, K. S. & Jansonius, J. N. (1992). The structure of neutral protease from *Bacillus cereus* at 0.2-nm resolution. *Eur. J. Biochem.* **207**, 781–791.
13. 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.
14. Lindqvist, Y. (1989). Refined structure of spinach glycolate oxidase at 2 Å resolution. *J. Mol. Biol.* **209**, 151–166.
15. Kleywegt, G. J. (1999). Recognition of spatial motifs in protein structures. *J. Mol. Biol.* **285**, 1887–1897.
16. Stamper, G. F., Morollo, A. A., Ringe, D. & Stamper, C. G. (1998). Reaction of alanine racemase with 1-aminoethylphosphonic acid forms a stable external aldimine. *Biochemistry*, **37**, 10438–10445.
17. Madsen, D. & Kleywegt, G. J. (2002). Interactive motif and fold recognition in protein structures. *J. Appl. Crystallog.* **35**, 137–139.
18. Wang, G. & Dunbrack, R. L., Jr (2003). PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589–1591.
19. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
20. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

21. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
22. Kleywegt, G. J., Harris, M. R., Zou, J. Y., Taylor, T. C., Wählby, A. & Jones, T. A. (2004). The Uppsala electron-density server. *Acta Crystallog. sect. D*, **60**, 2240–2249.
23. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E. *et al.* (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucl. Acids Res.* **31**, 365–370.
24. Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K. & Bairoch, A. (2002). The PROSITE database, its status in 2002. *Nucl. Acids Res.* **30**, 235–238.
25. Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L. & Thornton, J. M. (1997). PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* **22**, 488–490.
26. Frishman, D. & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: Struct. Funct. Genet.* **23**, 566–579.
27. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF-a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.
28. Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**, 147–150.
29. Creighton, T. (1993). *Structures and Molecular Properties* (2nd edit.), W.H. Freeman, New York.
30. Penel, S., Morrison, R. G., Mortishire-Smith, R. J. & Doig, A. J. (1999). Periodicity in α-helix lengths and C-capping preferences. *J. Mol. Biol.* **293**, 1211–1219.
31. Watson, J. D. & Milner-White, E. J. (2002). A novel main-chain anion-binding site in proteins: the nest. A particular combination of φ,ψ values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. *J. Mol. Biol.* **315**, 171–182.
32. Kabsch, W. & Sander, C. (1984). On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proc. Natl Acad. Sci. USA*, **81**, 1075–1078.
33. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). CATH—a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.
34. Mather, T., Oganessyan, V., Hof, P., Huber, R., Foundling, S., Esmon, C. & Bode, W. (1996). The 2.8 Å crystal structure of Gla-domainless activated protein C. *EMBO J.* **15**, 6822–6831.
35. Chang, Y. J., Wu, H. L., Hamaguchi, N., Hsu, Y. C. & Lin, S. W. (2002). Identification of functionally important residues of the epidermal growth factor-2 domain of factor IX by alanine-scanning mutagenesis. Residues Asn(89)-Gly(93) are critical for binding factor VIIIa. *J. Biol. Chem.* **277**, 25393–25399.
36. Celie, P. H., Van Stempvoort, G., Fribourg, C., Schurgers, L. J., Lenting, P. J. & Mertens, K. (2002). The connecting segment between both epidermal growth factor-like domains in blood coagulation factor IX contributes to stimulation by factor VIIIa and its isolated A2 domain. *J. Biol. Chem.* **277**, 20214–20220.
37. Wilkinson, F. H., London, F. S. & Walsh, P. N. (2002). Residues 88–109 of factor IXa are important for assembly of the factor X activating complex. *J. Biol. Chem.* **277**, 5725–5733.
38. Chitarra, V., Holm, I., Bentley, G. A., Petres, S. & Longacre, S. (1999). The crystal structure of C-terminal merozoite surface protein 1 at 1.8 Å resolution, a highly protective malaria vaccine candidate. *Mol. Cell.* **3**, 457–464.
39. Mather, T., Oganessyan, V., Hof, P., Huber, R., Foundling, S., Esmon, C. & Bode, W. (1996). The 2.8 Å crystal structure of Gla-domainless activated protein C. *EMBO J.* **15**, 6822–6831.
40. Gandrille, S. & Aiach, M. (1995). Identification of mutations in 90 of 121 consecutive symptomatic French patients with a type I protein C deficiency. The French INSERM Network on Molecular Abnormalities Responsible for Protein C and Protein S deficiencies. *Blood*, **86**, 2598–2605.
41. Zheng, Y. Z., Sakata, T., Matsusue, T., Umeyama, H., Kato, H. & Miyata, T. (1994). Six missense mutations associated with type I and type II protein C deficiency and implications obtained from molecular modelling. *Blood Coagul. Fibrinolysis*, **5**, 687–696.
42. Ireland, H., Thompson, E. & Lane, D. A. (1996). Gene mutations in 21 unrelated cases of phenotypic heterozygous protein C deficiency and thrombosis. Protein C Study Group. *Thromb. Haemost.* **76**, 867–873.
43. Greenwald, I. & Seydoux, G. (1990). Analysis of gain-of-function mutations of the lin-12 gene of *Caenorhabditis elegans*. *Nature*, **346**, 197–199.
44. Vardar, D., North, C. L., Sanchez-Irizarry, C., Aster, J. C. & Blacklow, S. C. (2003). Nuclear magnetic resonance structure of a prototype Lin12-Notch repeat module from human Notch1. *Biochemistry*, **42**, 7061–7067.
45. Hallberg, B. M., Henriksson, G., Pettersson, G. & Divne, C. (2002). Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *J. Mol. Biol.* **315**, 421–434.
46. Farooq, A., Chaturvedi, G., Mujtaba, S., Plotnikova, O., Zeng, L., Dhalluin, C. *et al.* (2001). Solution structure of ERK2 binding domain of MAPK phosphatase MKP-3: structural insights into MKP-3 activation by ERK2. *Mol. Cell.* **7**, 387–399.
47. Hennig, M., Grimm, B., Contestabile, R., John, R. A. & Jansonius, J. N. (1997). Crystal structure of glutamate-1-semialdehyde aminomutase: an alpha2-dimeric vitamin B6-dependent enzyme with asymmetry in structure and active site reactivity. *Proc. Natl Acad. Sci. USA*, **94**, 4866–4871.
48. Käck, H., Sandmark, J., Gibson, K., Schneider, G. & Lindqvist, Y. (1999). Crystal structure of diaminopelargonic acid synthase: evolutionary relationships between pyridoxal-5′-phosphate-dependent enzymes. *J. Mol. Biol.* **291**, 857–876.
49. Storici, P., Capitani, G., De Biase, D., Moser, M., John, R. A., Jansonius, J. N. & Schirmer, T. (1999). Crystal structure of GABA-aminotransferase, a target for antiepileptic drug therapy. *Biochemistry*, **38**, 8628–8634.
50. Storici, P., Capitani, G., Muller, R., Schirmer, T. & Jansonius, J. N. (1999). Crystal structure of human ornithine aminotransferase complexed with the highly specific and potent inhibitor 5-fluoromethylornithine. *J. Mol. Biol.* **285**, 297–309.
51. Malashkevich, V. N., Strop, P., Keller, J. W., Jansonius, J. N. & Toney, M. D. (1999). Crystal structures of dialkylglycine decarboxylase inhibitor complexes. *J. Mol. Biol.* **294**, 193–200.

52. Ding, Y. H., Javaherian, K., Lo, K. M., Chopra, R., Boehm, T., Lanciotti, J. *et al*. (1998). Zinc-dependent dimers observed in crystals of human endostatin. *Proc. Natl Acad. Sci. USA*, **95**, 10443–10448.
53. Hohenester, E., Sasaki, T., Olsen, B. R. & Timpl, R. (1998). Crystal structure of the angiogenesis inhibitor endostatin at 1.5 Å resolution. *EMBO J.* **17**, 1656–1664.
54. Sasaki, T., Larsson, H., Tisi, D., Claesson-Welsh, L., Hohenester, E. & Timpl, R. (2000). Endostatins derived from collagens XV and XVIII differ in structural and binding properties, tissue distribution and anti-angiogenic activity. *J. Mol. Biol.* **301**, 1179–1190.
55. Adman, E. T., Godden, J. W. & Turley, S. (1995). The structure of copper-nitrite reductase from *Achromobacter cycloclastes* at five pH values, with NO_2- bound and with type II copper depleted. *J. Biol. Chem.* **270**, 27458–27474.
56. Sivaraman, J., Li, Y., Banks, J., Cane, D. E., Matte, A. & Cygler, M. (2003). Crystal structure of *Escherichia coli* PdxA, an enzyme involved in the pyridoxal phosphate biosynthesis pathway. *J. Biol. Chem.* **278**, 43682–43690.
57. Romier, C., Dominguez, R., Lahm, A., Dahl, O. & Suck, D. (1998). Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. *Proteins: Struct. Funct. Genet.* **32**, 414–424.
58. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA, USA.
59. Nar, H., Bauer, M., Schmid, A., Stassen, J. M., Wienen, W., Priepke, H. W. *et al*. (2001). Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure*, **9**, 29–37.
60. Sichler, K., Banner, D. W., D'Arcy, A., Hopfner, K. P., Huber, R., Bode, W. *et al*. (2002). Crystal structures of uninhibited factor VIIa link its cofactor and substrate-assisted activation to specific interactions. *J. Mol. Biol.* **322**, 591–603.
61. Garman, S. C., Simcoke, W. N., Stowers, A. W. & Garboczi, D. N. (2003). Structure of the C-terminal domains of merozoite surface protein-1 from *Plasmodium knowlesi* reveals a novel histidine binding site. *J. Biol. Chem.* **278**, 7264–7269.
62. Pizarro, J. C., Chitarra, V., Verger, D., Holm, I., Petres, S., Dartevelle, S. *et al*. (2003). Crystal structure of a Fab complex formed with PfMSP1-19, the C-terminal fragment of merozoite surface protein 1 from *Plasmodium falciparum*: a malaria vaccine candidate. *J. Mol. Biol.* **328**, 1091–1103.
63. Hopfner, K. P., Lang, A., Karcher, A., Sichler, K., Kopetzki, E., Brandstetter, H. *et al*. (1999). Coagulation factor IXa: the relaxed conformation of Tyr99 blocks substrate binding. *Struct. Fold. Des.* **7**, 989–996.
64. Banbula, A., Potempa, J., Travis, J., Fernandez-Catalan, C., Mann, K., Huber, R. *et al*. (1998). Amino-acid sequence and three-dimensional structure of the *Staphylococcus aureus* metalloproteinase at 1.72 Å resolution. *Structure*, **6**, 1185–1193.
65. Boulanger, M. J. & Murphy, M. E. (2001). Alternate substrate binding modes to two mutant (D98N and H255N) forms of nitrite reductase from *Alcaligenes faecalis* S-6: structural model of a transient catalytic intermediate. *Biochemistry*, **40**, 9132–9141.
66. Ellis, M. J., Dodd, F. E., Sawers, G., Eady, R. R. & Hasnain, S. S. (2003). Atomic resolution structures of native copper nitrite reductase from *Alcaligenes xylosoxidans* and the active site mutant Asp92Glu. *J. Mol. Biol.* **328**, 429–438.
67. Abreu, I. A., Lourenco, A. I., Xavier, A. V., LeGall, J., Coelho, A. V., Matias, P. M. *et al*. (2003). A novel iron centre in the split-Soret cytochrome c from *Desulfovibrio desulfuricans* ATCC 27774. *J. Biol. Inorg. Chem.* **8**, 360–370.
68. Phale, P. S., Philippsen, A., Widmer, C., Phale, V. P., Rosenbusch, J. P. & Schirmer, T. (2001). Role of charged residues at the OmpF porin channel constriction probed by mutagenesis and simulation. *Biochemistry*, **40**, 6319–6325.
69. Peneff, C., Ferrari, P., Charrier, V., Taburet, Y., Monnier, C. & Zamboni, V. (2001). Crystal structures of two human pyrophosphorylase isoforms in complexes with UDPGlc(Gal)NAc: role of the alternatively spliced insert in the enzyme oligomeric assembly and active site architecture. *EMBO J.* **20**, 6191–6202.
70. Pedersen, L. C., Darden, T. A. & Negishi, M. (2002). Crystal structure of beta 1,3-glucuronyltransferase I in complex with active donor substrate UDP-GlcUA. *J. Biol. Chem.* **277**, 21869–21873.

Edited by J. Thornton

(Received 8 September 2004; received in revised form 17 January 2005; accepted 17 January 2005)