Quarterly Reviews of Biophysics

cambridge.org/qrb

Hiding in plain sight: three chemically distinct α-helix types

Shuguang Zhang[1] and Martin Egli[2]

[1]Laboratory of Molecular Architecture, Media Lab, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA and [2]Department of Biochemistry, Vanderbilt University, School of Medicine, Nashville, Tennessee 37232-0146, USA

Review Article

Cite this article: Zhang S, Egli M (2022). Hiding in plain sight: three chemically distinct α-helix types. *Quarterly Reviews of Biophysics* **55**, e7, 1–10. https://doi.org/10.1017/S0033583522000063

Received: 26 April 2022
Revised: 9 June 2022
Accepted: 14 June 2022

Key words:
Type I-hydrophilic α-helix; Type II-hydrophobic α-helix; Type III-amphiphilic α-helix; QTY code; membrane protein design

Author for correspondence:
Shuguang Zhang,
E-mail: Shuguang@mit.edu

Abstract

Linus Pauling in 1950 published a three-dimensional model for a universal protein secondary structure motif which he initially called the alpha-spiral. Jack Dunitz, then a postdoc in Pauling's lab suggested to Pauling that the term helix is more accurate than spiral when describing the right-handed peptide and protein coiled structures. Pauling agreed, hence the rise of the alpha-helix, and, by extension, the 'double helix' structure of DNA. Although structural biologists and protein chemists are familiar with varying polar and apolar characters of amino acids in alpha-helices, to non-experts the three chemically distinct alpha-helix types classified here may hide in plain sight.

Table of contents

The rise of the α-helix

Linus Pauling in 1950 published a three-dimensional model for a universal protein secondary structure motif which he initially called the α-spiral (Pauling and Corey, 1950; Dunitz, 2001). Jack Dunitz, then a postdoc in Pauling's lab suggested to Pauling that the term helix is more accurate than spiral when describing the right-handed peptide and protein coiled structures. Pauling agreed, hence the rise of the α-helix (Pauling *et al.*, 1951; Dunitz, 2013; Egli and Zhang, 2022), and, by extension, the 'double helix' structure of DNA. Although structural biologists and protein chemists are familiar with varying polar and apolar characters of amino acids in α-helices (Eisenberg *et al.*, 1982; Eisenberg *et al.*, 1984; Eisenberg, 2003), to non-experts the three chemically distinct α-helix types classified here may hide in plain sight.

Three helix variants in proteins

There are three variants of helices found in proteins (Stryer, 1981; Brändén and Tooze, 1991; Fersht, 1998; Fodje and Al-Karadaghi, 2002; Armen *et al.*, 2003; Liljas *et al.*, 2017). Briefly, the most commonly observed α-helix has 3.6 residues per 360° helical turn (100° rotation per amino acid) and exhibits a translation of 1.5 Å along the helical axis with *i*, *i* + 4 hydrogen

bond (H-bond) formation (Pauling and Corey, 1950; Pauling *et al.*, 1951). There are two additional kinds of helices: The 3_{10}-helix has 3 residues per 360° helical turn (120° rotation per amino acid) and exhibits a translation of 2 Å along the helical axis with *i, i + 3* H-bonding (Bragg *et al.*, 1950; Perutz *et al.*, 1960). This 3_{10}-helix often occurs at the end of α-helices (Armen *et al.*, 2003). The π-helix has 4.4 residues per 360° helical turn (83° rotation per amino acid) and exhibits a translation of 1.2 Å along the helical axis with *i, i + 5* H-bonding (Fodje and Al-Karadaghi, 2002). The π-helix occurs in the middle of longer α-helices (Armen *et al.*, 2003). However, our analysis is not about the variations in the geometry of protein helices; interested readers should instead consult the original and more recent literature. Rather, we propose to classify structurally identical α-helices into three chemically distinct types, namely: Type I, hydrophilic α-helix; Type II, hydrophobic α-helix; Type III, amphiphilic α-helix.

Twenty amino acids with hydrophilic and hydrophobic characteristics

There are two general classes of amino acids, hydrophilic and hydrophobic (Stryer, 1981; Brändén and Tooze, 1991; Fersht, 1998; Liljas *et al.*, 2017). Hydrophilic amino acids are polar and readily water-soluble. They include Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q), Lysine (K), Arginine (R), Serine (S), Threonine (T), Histidine (H) and Tyrosine (Y). Conversely, hydrophobic amino acids are nonpolar and water-insoluble. They include Leucine (L), Isoleucine (I), Valine (V), Phenylalanine (F), Methionine (M), Tryptophan (W) and Alanine (A). Cysteine (C) and Glycine (G) are only weakly hydrophilic and hydrophobic, respectively. Proline (P) is a unique case, the nitrogen of Proline is part of the ring structure formed by the side chain and cannot participate in the helical H-bonds. Proline is frequently found at the N-terminus of helices, but if it occurs in the middle of an α-helix, it forces the helix to bend (Liljas *et al.*, 2017).

Several amino acids share striking structural similarities within their 1.5 Å-resolution electron density maps, despite their very different chemical properties (Fig. 1). These are: D, N, E and Q *versus* L; T *versus* V and I; and Y *versus* F. The side chains of D, N, E and Q can form four H-bonds with separate water molecules. The carboxylate oxygen atoms of D and E present acceptors for formation of four H-bonds. The amide moieties of N and Q present a donor for formation of two H-bonds (NH_2) and an acceptor for formation of two H-bonds (O). The hydroxyl group of S, T and Y can engage in three H-bonds, via two oxygen lone pairs and one O-H bond (Figure S1). Arginine can donate five H-bonds and lysine can donate three H-bonds in the protonated state of their side chains. Histidine has one H-bond donor and one acceptor, and W has one H-bond donor (Figure S1).

The atomic-resolution molecular structures of proteins unequivocally demonstrate that all 20 natural amino acids are found in α-helices, regardless of their hydrophilic and hydrophobic properties, although some amino acids have a higher propensity to form an α-helix (Chou and Fasman, 1974; Creighton, 1992).

A strong periodicity in the hydrophobic moment of amino acids in α-helices

David Eisenberg and colleagues in early 1980s had carried out quantitative studies of α-helices using a method they termed the hydrophobic moment (Eisenberg *et al.*, 1982; Eisenberg *et al.*, 1984). They defined the hydrophobic moment as follows: '*Periodicities in the polar/apolar character of the amino acid sequence of a protein can be examined by assigning to each residue a numerical hydrophobicity and searching for periodicity in the resulting one-dimensional function. The strength of each periodic component is the quantity that has been termed the hydrophobic moment. When proteins of known three-dimensional structure are examined, it is found that sequences that form a helix tend to have, on average, a strong periodicity in the hydrophobicity of 3.6 residues, the period of the α-helix.*' They analyzed 157 segments of α-helices from globular protein structures from the previous 25 years (1959–1984) and found that the average of these α-helical segments has a strong maximum at 100°. They observed that on average these primary structures are more amphiphilic arranged as α-helices than in other periodic secondary structures (Eisenberg *et al.*, 1982; Eisenberg *et al.*, 1984). They also pointed out that there are other factors influencing the hydrophobic periodicity, not all segments known to be α-helices give a profile with a maximum at 100°, especially if the hydrophobic residues are not uniformly distributed along one side parallel to the axis but rather arranged in a slanted region across the helix (Eisenberg *et al.*, 1982; Eisenberg *et al.*, 1984). However, they did not explicitly classify the three chemically distinct α-helix types.

Three chemically distinct α-helix types

The α-helix can be classified into three chemically distinct types (Fig. 2). Although they are significantly different in their chemical properties and water-solubility, they have nearly identical molecular structures, namely: (i) a 1.5 Å rise per amino acid, (ii) a 100° rotation per amino acid, (iii) 3.6 amino acids per helical turn, (iv) a 5.4 Å rise per helical turn, and (v) the key feature, i.e. each NH-group of an amino acid forms an H-bond with the C = O group of the amino acid 4 residues away. The latter, repeated *i, i + 4* H-bonding pattern is the most prominent characteristic of an α-helix (Stryer, 1981; Brändén and Tooze, 1991; Fersht, 1998; Liljas *et al.*, 2017).

The Type I α-helix is mostly comprised of hydrophilic amino acids, including D, E, N, Q, K, R, S, T and Y, and is commonly found on the outer layer in water-soluble globular proteins, and in some cases, in the inner layer of membrane helices, away from the lipid bilayer. The Type II α-helix is mostly comprised of hydrophobic amino acids, including L, I, V, F, M, A, W and P, and is commonly found in the helical transmembrane segments of membrane proteins and buried in the interior of water-soluble proteins. The Type III amphiphilic α-helix is almost equally comprised of hydrophilic and hydrophobic amino acids that are sometimes partitioned on the hydrophobic face and the hydrophilic face. Using an analogy, we can think of the two faces as the front and back of our fingers. The Type III α-helix is sometimes attached to the surface of membrane lipid bilayers, or partially buried in the hydrophobic core and partially exposed on the surface of water-soluble globular proteins, or in the integral membrane pores that face away from the hydrophobic lipid bilayer. In the literature glycine (G) is often counted as hydrophobic because its side chain hydrogen does not engage in H-bonds, but is only weakly hydrophobic at the same time. If we were to exclude glycine from the list of hydrophobic amino acids, Type I and Type III α-helices would exhibit different overall percentages of hydrophobicity.

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press

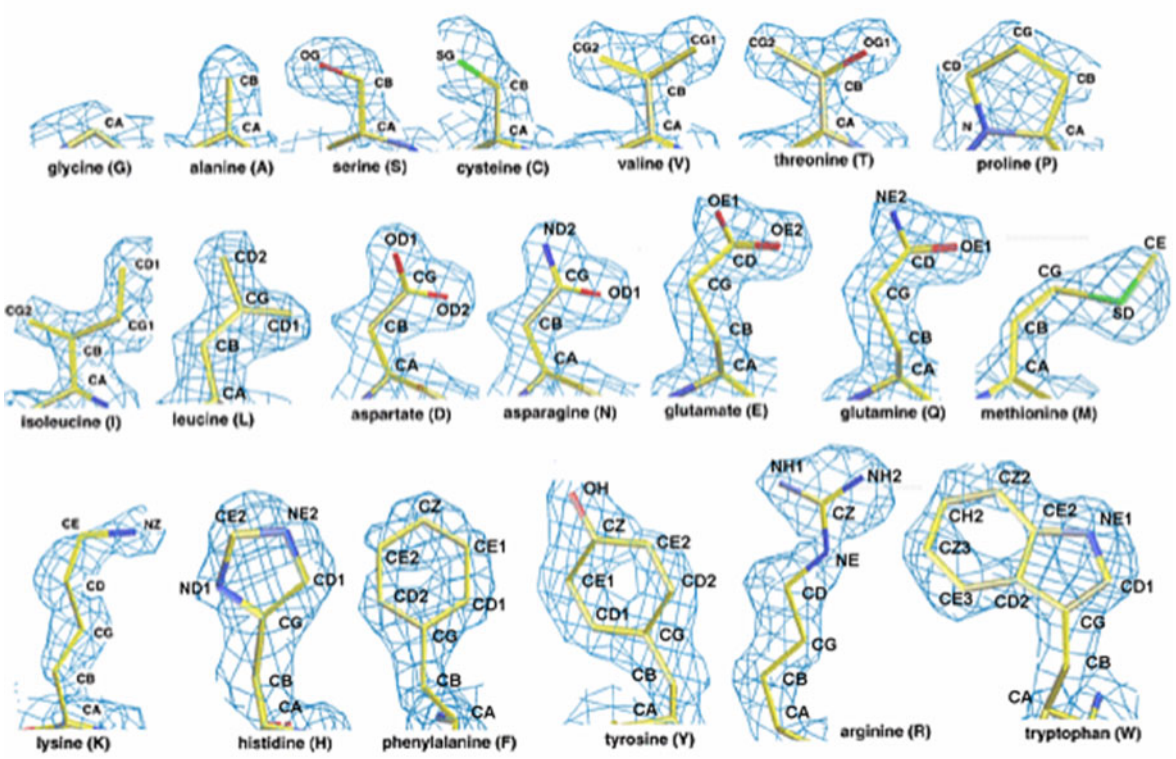

Fig. 1. Experimental X-ray electron density maps (~1.5 Å resolution) of 20 amino acids arranged by size. This figure is provided by Dr. Mike Sawaya (UCLA), and used with permission in order to show the individual amino acid electron density maps at high resolution. The density maps demonstrate similar shapes of V and T; L, D, N, E and Q, and F and Y. Please see Dr. Mike Sawaya's original website: http://people.mbi.ucla.edu/sawaya/m230d/Modelbuilding/modelbuilding.html (courtesy of Dr. Michael R. Sawaya of University of California, Los Angeles, CA, USA).

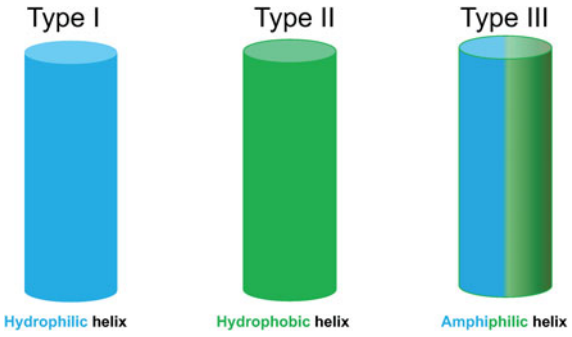

Fig. 2. Schematic illustrations of three chemically distinct types of α-helices: Type I hydrophilic, Type II hydrophobic, Type III Janus (% hydrophilic/hydrophobic). The Type I α-helix is mostly comprised of hydrophilic amino acids including Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q), Lysine (K), Arginine (R), Serine (S), Threonine (T), and Tyrosine (Y) that are commonly found on the outer layer in water-soluble globular proteins. The Type II α-helix is mostly comprised of hydrophobic amino acids Leucine (L), Isoleucine (I), Valine (V), Phenylalanine (F), Methionine (M), Tryptophan (W) and Alanine (A) that are commonly found in helical transmembrane segments in membrane proteins. The Type III amphiphilic α-helix is mostly comprised of hydrophilic and hydrophobic amino acids that are partitioned on two faces: hydrophobic face and hydrophilic face.

Two classes of proteins: hydrophilic and hydrophobic

Natural proteins can be generally divided into two classes: Class I is hydrophilic and Class II is hydrophobic (Stryer, 1981; Brändén and Tooze, 1991; Fersht, 1998; Liljas *et al.*, 2017). Class I hydrophilic proteins include water-soluble proteins that reside in the cytoplasm, such as hemoglobin, as well as metabolic enzymes and circulating proteins outside cells, such as growth factors, hormones, and antibodies. Class II hydrophobic proteins comprise the integral membrane proteins that are embedded in cellular and other membranes. They include G protein-coupled receptors (GPCRs), membrane transporters and ion channels as well as the photosynthesis machinery. The Class I proteins are generally water-soluble, and the Class II proteins, as integral membrane proteins, are generally water-insoluble. In order to solubilize Class II proteins, various detergent/surfactants are required after isolating them from their lipid-bilayer membrane environment (Lin and Guidotti, 2009; Linke, 2009; Duquesne and Sturgis, 2010).

Long α-helices in proteins

Some proteins are almost entirely made up of α-helices, including trompomyosin (Doran *et al.*, 2020) and the helical parts of keratins (Lee *et al.*, 2020) (Fig. 3). These proteins are often involved in cytoskeletal structural scaffolds, similar to the scaffolds used to construct buildings. A single α-helix structure of trompomyosin consists of 164 amino acids (~244 Å long, or ~24.4 nm) (Doran *et al.*, 2020). In crystal structures determined for portions of keratins, a single α-helix is ~90-amino acids long (Lee *et al.*, 2020). It has been suggested that the α-helical parts (~330 aa) of keratins are likely to be as long as ~500 Å (~50 nm) (Hanukoglu and Fuchs, 1983), but currently no molecular structure is available. Another long α-helix is hemagglutinin of influenza virus, which has 52 amino acids (~77 Å) and is crucial for viral fusion with the host membrane to confer entrance into the cell for infection during pH change (Gamblin *et al.*, 2004). All 20 amino acids are found in these α-helices, regardless of their chemical

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press

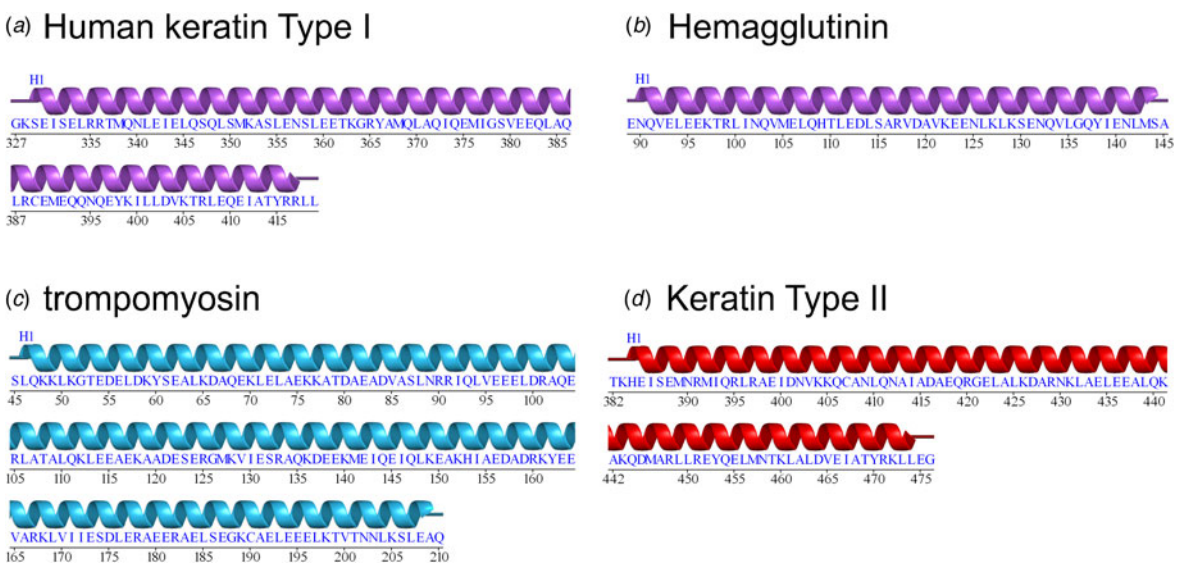

Fig. 3. Long α-helices in proteins. There are some examples of long α-helices in proteins that contain most of the 20 amino acids. (*a*) Human keratin Type I (K1C14) where the α-helix is made of 90 aa; (*b*) The α-helix of hemagglutinin of influenza virus has 52 amino acids and is crucial for viral fusion with the host membrane to confer entrance into the cell for infection. (*c*) The α-helix crystal structure of trompomyosin (TPM1, 4.25 Å) is comprised of 164 aa. (*d*) Keratin Type II (K2C5) where the α-helix is made of 91 aa. It has been suggested that keratin helices could be as long as 50 nm comprising 330 aa.

properties. In the long 164-amino acid α-helix example of pig trompomyosin, 18 kinds of amino acids are present with only P and W missing (Doran *et al.*, 2020). However, in the transmembrane helices of glucose transporter GLUT1 (Custódio *et al.*, 2021), there are two W and five P residues. Thus, all 20 amino acids are observed in α-helices.

Type I α-helix: hydrophilic and highly water-soluble

The Type I hydrophilic α-helix is mostly comprised of hydrophilic amino acids D, E, N, Q, K, R, S, T, Y (Table 1, Fig. 4). G has also been found in Type I α-helices. It is counted as hydrophobic because its 'side chain' H atom does not form H-bonds. At the same time this amino acid is only weakly hydrophobic. If G is not counted as hydrophobic, the overall % hydrophobicity of G-containing helices will have to be adjusted.

It is difficult to find long Type I α-helices with only hydrophilic residues. However, it is common to find the Type I α-helix in water-soluble proteins. Figure 4 shows Type I α-helices from proteins with between 73-82% hydrophilic amino acids (Table 1). For example, (i) yeast Zuotin α-helix H5: 13/16 = 81.3% hydrophilic (Zhang *et al.*, 1992; Lee *et al.*, 2016); (ii) Troponin T (TNNT) α-helix 2: 14/17 = 82.2% hydrophilic (Takeda *et al.*, 2003); (iii) Troponin I (TNNI3) α-helix 2: 13/17 = 76.5% hydrophilic (Yaguchi *et al.*, 2017); (iv) Troponin T (TNNT) α-helix 1: 16/22 = 72.7% hydrophilic (Takeda *et al.*, 2003). It is clear that countless Type I α-helices exist in proteins.

Type II α-helix: hydrophobic and spanning membranes

The Type II hydrophobic α-helix is mostly comprised of hydrophobic amino acids L, I, V, F, M, P, W and A (Fig. 5 and Table 1). Therefore, Type II α-helices are mostly found embedded in lipid bilayer membranes. They constitute transmembrane α-helices that contain between 81% and 91% hydrophobic amino acids. For example, (i) GLUT1 (TM9): 21/25 = 84% hydrophobic (Custódio *et al.*, 2021); (ii) GLUT3: 16/18 = 88.9% hydrophobic (Deng *et al.*, 2015); (iii) pufL, L-chain (TM1): 22/24 = 91.6% hydrophobic (Xu *et al.*, 2004); (iv) pufM, M-chain (TM5): 18/22 = 81.8% hydrophobic (Xu *et al.*, 2004); e) CCR5 (TM1): 25/31 = ~81% hydrophobic (Tan *et al.*, 2013). Again, it is uncommon to find a completely hydrophobic α-helix, even in transmembrane domains. Other Type II α-helices are also found in interior domains of water-soluble proteins, i.e. in the hydrophobic core.

Type III α-helix: amphiphilic

The Type III amphiphilic α-helix is comprised of almost equal numbers of hydrophilic and hydrophobic amino acids that are partitioned on hydrophobic and hydrophilic faces (Fig. 6). The Type III α-helix is sometimes attached to the surface of lipid bilayer membranes, or partially buried in the hydrophobic core and partially exposed on the surface of water-soluble globular proteins (Stryer, 1981; Brändén and Tooze, 1991; Fersht, 1998; Liljas *et al.*, 2017). Examples include α-helices in, (i) human hemoglobin beta subunit, 50% hydrophilic/hydrophobic amino acids, (ii) T4 lysozyme with 47.4% hydrophilic and 52.6% hydrophobic amino acids (Mooers and Matthews, 2004), (iii) alcohol dehydrogenase with 53.8% hydrophilic and 46.2% hydrophobic amino acids (Niederhut *et al.*, 2001), (iv) Cytochrome b562, a coiled-coil tetramer (H4) with 46% hydrophilic and 54% hydrophobic amino acids (Lederer *et al.*, 1981), and (v) the designed 29-amino acids, trimeric coiled-coil VALD with 51.7% hydrophilic and 48.3% hydrophobic amino acids (Ogihara *et al.*, 1997). There are also many Type III α-helices in proteins that reside on the surfaces of membranes, whereby one side interacts with the hydrophobic lipids from the membrane and the other is exposed to the aqueous environment.

Conversion of a hydrophobic α-helix to a hydrophilic α-helix using a simple QTY code

Over a decade ago, in 2010, Alexander Rich posed a question to one of us (S.Z.) *'Can you convert a hydrophobic α-helix to a*

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press

Table 1. Three chemically distinct types of α-helices: hydrophilic, hydrophobic and amphiphilic

Name	Type	aa	Sequence (Helix number in proteins)	%
Zuotin	I	16	RIKMFKEEEKKEKERR (H5)	81.2%
Troponin T	I	17	KRQTEREKKKKILAERR (H2)	82.3%
Troponin I	I	17	DKVDEERYDIEAKVTKN (H2)	76.5%
Troponin I	I	22	KQELEREAEERRGEKGRALSTR (H1)	~73%
Troponin C	I	11	TEDDIEELMKD (H8)	~73%
GLUT1	II	25	RRLHLIGLAGMAGCAILMTIALALL (TM9)	85%
GLUT3	II	18	MSFVCIGAILVFVAFFEI (TM 10)	~89%
pufL	II	24	GFFGVATFFFAALGIILIAWSAVL (TM1)	91.6%
pufM	II	22	GFFGVATFFFAALGIILIAWSAVL (TM5)	~82%
CCR5	II	31	GIHRWAIWMAVLVTLTGGIGIL (TM1)	~81%
Hemoglobin Beta subunit	III	34	KQIAARLLPPLYSLVFIFGFVGNMLVILILI (H6)	50%/50%
T4 lysozyme	III	19	DEAEKLFNQDVDAAVRGIL (H1)	47.4%/52.6%
ADH1A	III	13	FEKINEGFDLLHS (H16)	53.8%/46.2%
Cytochrom b562	III	24	EMKDFRHGFDILVGQIDDALKLAN (H3)	46%/54%
Coiled-coil	III	29	EVEALEKKVAALESKVQALEKKVEALEHG	51.7%/48.3%

The α-helix can be classified into three chemically distinct Types. The Type I hydrophilic α-helix is mostly comprised of hydrophilic amino acids D, E, N, Q, K, R, S, T, Y; Type II hydrophobic α-helix is mostly comprised of hydrophobic amino acids L, I, V, F, M, P and A; Type III amphiphilic alpha-helix is comprised of both hydrophilic and hydrophobic amino acids, the hydrophobic face and the hydrophilic face. The Type III α-helix is sometimes attached to the surface of the membrane lipid bilayer, or partially buried in the hydrophobic core and partially exposed on the surface of water-soluble globular proteins. Glycine (G) is counted as hydrophobic because its side chain does not engage in H-bonding, although it is only very weakly hydrophobic. If G is not counted as hydrophobic, the percentages will be different.

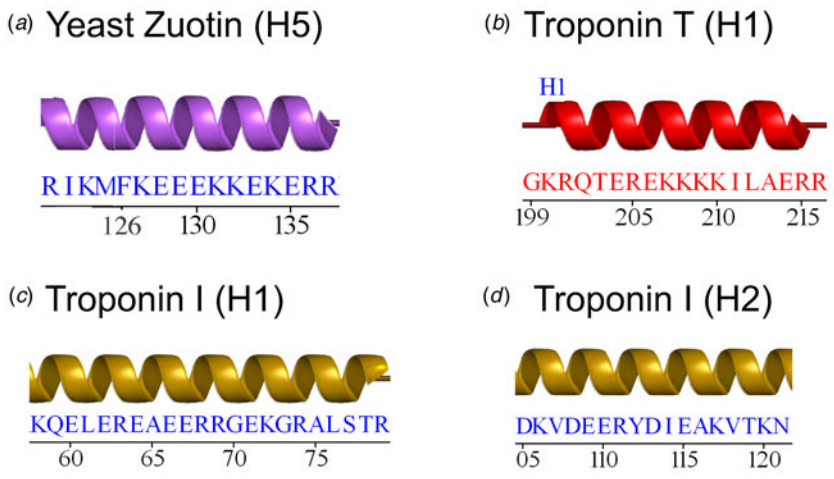

Fig. 4. Type I α-helix comprises between 73%-82% hydrophilic amino acids (Also see Table 1). (*a*) Yeast Zuotin α-helix 5 (13/16 = ~81% hydrophilic), (*b*) Troponin T α-helix 2 (TNNT), (14/17 = 82.2% hydrophilic), (*c*) Troponin I (TNNI3), α-helix 2, (13/17 = 76.5%), (*d*) Troponin T α-helix 1, (6/22 = 73% hydrophilic.

hydrophilic one?' which prompted the immediate answer '*Yes, because hemoglobin is composed mostly of α-helices and it is one of the most water-soluble proteins*' But how? Several previous attempts were not successful, e.g. (Mitra *et al.*, 2002). Others succeeded by using sophisticated approaches, including computer-assisted modeling, simulations and introduction of specific mutations in a given protein (Slovic *et al.*, 2003, 2004, 2005; Zhang SQ *et al.*, 2018*a*; Zhang S *et al.*, 2018*b*). For each water-insoluble protein, it is necessary to go through a rigorous and time-consuming exercise to render it water-soluble. There was no straightforward and shared code to simplify such hydrophobic to hydrophilic protein conversion.

A simple QTY code (Fig. 7) was devised by considering the 1.5 Å-resolution electron density maps of the 20 amino acids (Fig. 1). It is clear that several hydrophobic amino acids closely resemble those of hydrophilic amino acids in terms of shape (Fig. 1). Thus, the QTY code is based on the fact that the electron density (computed at sufficient resolution) of the hydrophobic L is similar to that of the hydrophilic N and Q. The electron density of the hydrophobic I and V are similar to that of the hydrophilic T. Lastly, the electron density of the hydrophobic F is similar to that of the hydrophilic Y (Fig. 1). Although water molecules also form H-bonds with D (- charge), E (−), K (+) and R (+), these residues introduce charge, thereby altering the surface property of proteins. Thus, they are not considered in the QTY code. Since the sidechain of N is proximate to the backbone of the polypeptide chain and sometimes interferes with H-bonds to the amide moiety, and often maps to turns, it was not included in the code.

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press

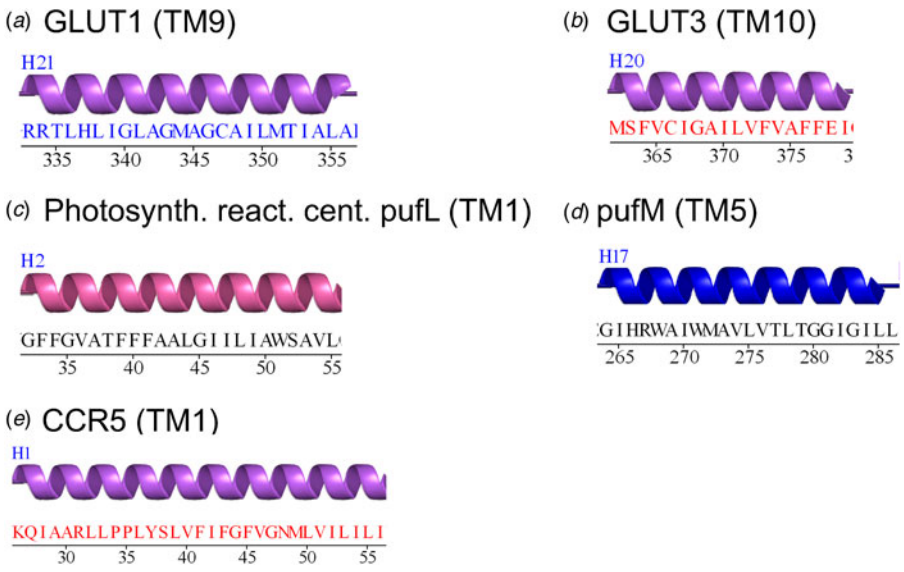

Fig. 5. Type II α-helix is a transmembrane helix comprising 81%-91% hydrophobic amino acids. (*a*) GLUT1 (TM9) 21/25 = 85% hydrophobic, (*b*) GLUT3, 16/18 = 89% hydrophobic, (*c*) pufL, L-chain (TM1) (22/24 = 91.6% hydrophobic), (*d*) pufM, M-chain (TM5), 18/22 = (82%) hydrophobic, (*e*) CCR5 (TM1), (25/31 = ~81% hydrophobic).

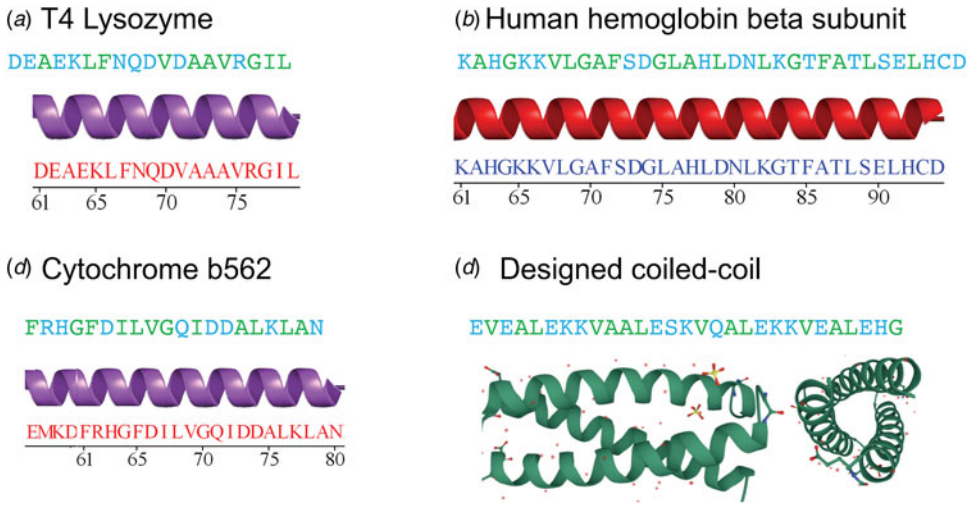

Fig. 6. Type III α-helix is an amphiphilic helix and comprises both hydrophilic and hydrophobic amino acids. (*a*) T4 lysozyme with 47.4% hydrophilic and 52.6% hydrophobic amino acids, (*b*) alcohol dehydrogenase with 53.8% hydrophilic and 46.2% hydrophobic amino acids, (*c*) Cytochrome b562 is a coiled-coil tetramer (H4) with 46% hydrophilic and 54% hydrophobic amino acids, (*d*) Designed 29 aa trimeric coiled-coil VALD (1COI, 2.10 Å) with 51.7% hydrophilic and 48.3% hydrophobic amino acids.

The QTY code was initially applied to convert transmembrane (TM) α-helices of seven G protein-coupled receptors (GPCRs), namely the chemokine receptors CCR5, CCR9, CCR10, CXCR2, CXCR4, CXCR5, CXCR7 (Zhang *et al.*, 2018*a*, 2018*b*; Qing *et al.*, 2019; Hao *et al.*, 2020; Tegler *et al.*, 2020; Skuhersky *et al.*, 2021; Tao *et al.*, 2022). Despite ~20–30% overall amino acid changes (46–58% transmembrane domain amino acid changes), these water-soluble QTY variant receptors still retained their overall structure, with a similar amount of α-helical content, and folded into stable proteins (Fig. 8). Most importantly, they were able to bind their respective ligands, with affinities that were similar to those of the natural receptors. After using the QTY code to replace the hydrophobic amino acids L, I, V and F in the transmembrane domain, the hydrophobicity on the surface of membrane proteins is reduced significantly. They in turn became more hydrophilic for several GPCRs, including CCR5, CCR9, CXC2 and CXCR4 (Fig. 8). Currently, experimental determinations of the structures of QTY variant receptors are in progress in collaborations with other laboratories.

We later successfully applied the QTY code to four cytokine receptors interleukins IL4Rα, IL10Rα and interferon INFγR1 and INFλR1 that have a single transmembrane α-helix (Hao *et al.*, 2020). Other laboratories also independently reproduced the water-soluble QTY variant receptors, carried out ligand binding experiments and obtained similar results. Recently, a water-soluble CCR10^{QTY} variant has been used as antigen to generate monoclonal antibodies.

Currently, we are interested in studying glucose transporters (12TM) since they are directly involved in proliferation and metastasis of a wide spectrum of cancer cells that constantly demand energy supply (Barron *et al.*, 2016). The QTY code has also been applied to design water-soluble glucose transporter

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press

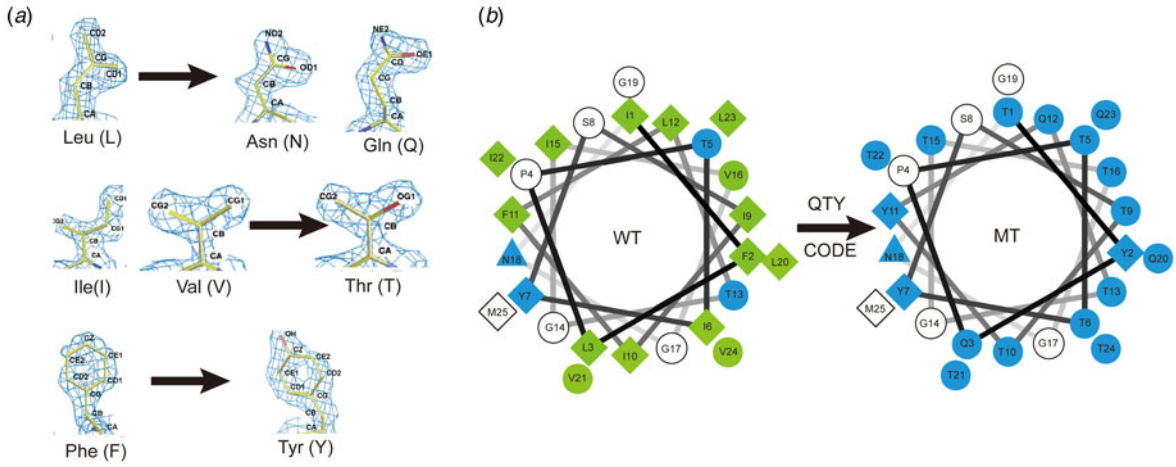

Fig. 7. The QTY code and how it replaces L, V, I and F with Q, T and Y. (A) Crystallographic electron density maps of the following amino acids: Leucine (L), Asparagine (N), Glutamine (Q), Isoleucine (I), Valine (V), Threonine (T), Phenylalanine (F) and Tyrosine (Y). The density maps of L, N and Q are very similar. Likewise, the density maps of I, V and T are similar, and the density maps of F and Y are similar. The side chains of L, V, I, and F cannot form any H-bonds with water, thus rendering them water-insoluble. On the other hand, N and Q can form four H-bonds with four water molecules, two as H-bond donors and two as H-bond acceptors (SI Appendix Fig. S7). Likewise, three water molecules can form H-bonds with the –OH (two H-bond donors and one H-bond acceptor) of Thr (T) and Tyr (Y). Both L and Q have high tendencies to form α-helices, but N frequently occurs at turns. Thus, Q was used to replace L, but not N. I, V and T are all β-branched amino acids and their density maps are very similar, indicating similar shapes. (B) Helical wheels before and after applying the QTY code to transmembrane helical segment 1 (TM1) of CXCR4. Amino acids that interact with water molecules are light blue in color. The QTY code conversions render the α-helical segment water-soluble.

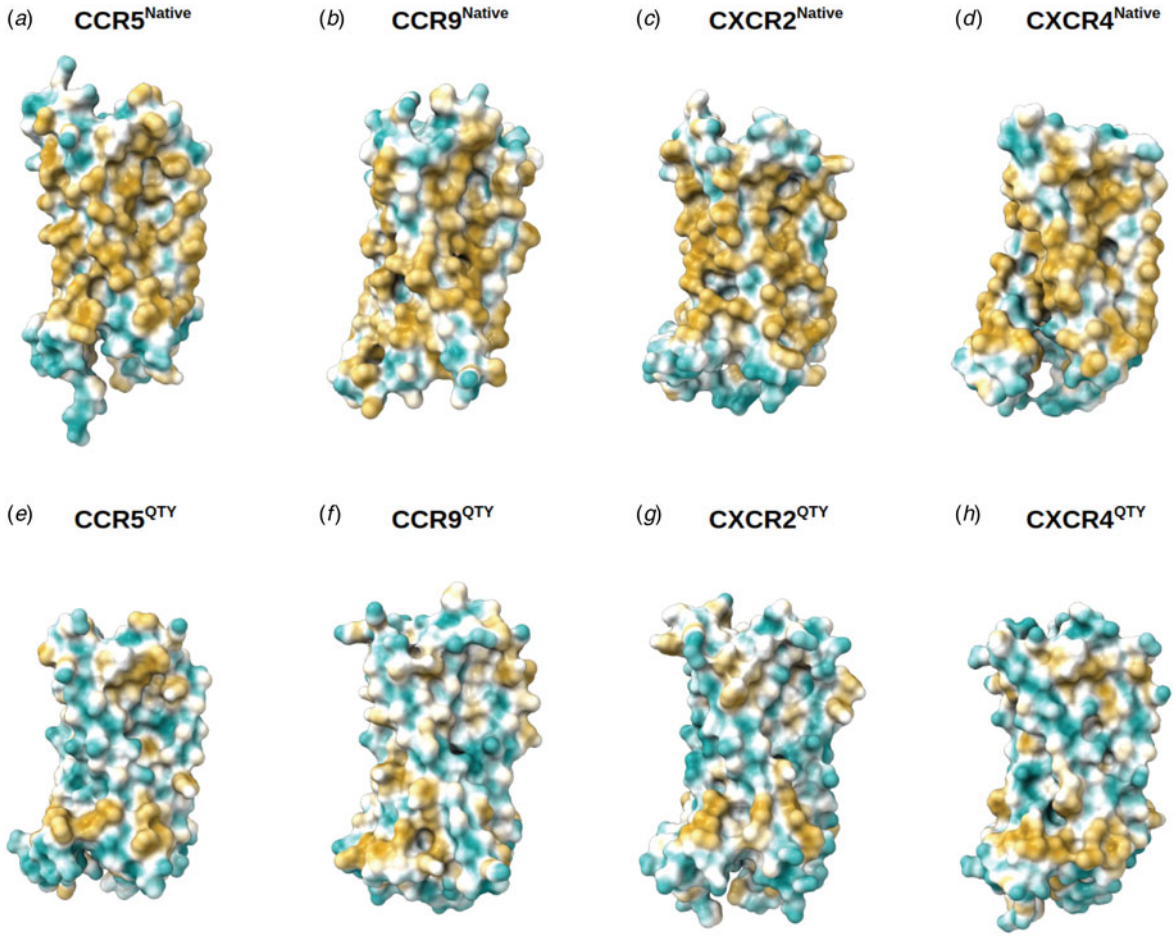

Fig. 8. Surface hydrophobic patches of X-ray crystal structures of native chemokine receptors and AlphaFold2 predicted water-soluble QTY variants. The native GPCR receptors mostly expose hydrophobic residues leucine (L), isoleucine (I), valine (V) and phenylalanine (F) to the hydrophobic lipid bilayer of the cell membrane. After replacing L, I, V, F with polar amino acids, glutamine (Q), threonine (T) and tyrosine (Y), the surfaces are much less hydrophobic. The large surface hydrophobic patch (yellow color) of the native receptors determined by X-ray crystallography: (*a*) CCR5, (*b*) CCR9, (*c*) CXCR2 and (*d*) CXCR4. The hydrophobic patch is significantly reduced on the transmembrane domains for the AlphaFold2 predicted water-soluble QTY variants: (*e*) $CCR5^{QTY}$, (*f*) $CCR9^{QTY}$, (*g*) $CXCR2^{QTY}$, (*h*) $CXCR4^{QTY}$. These QTY variants become water-soluble without any detergent. The N- and C-termini are removed for clarity.

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press

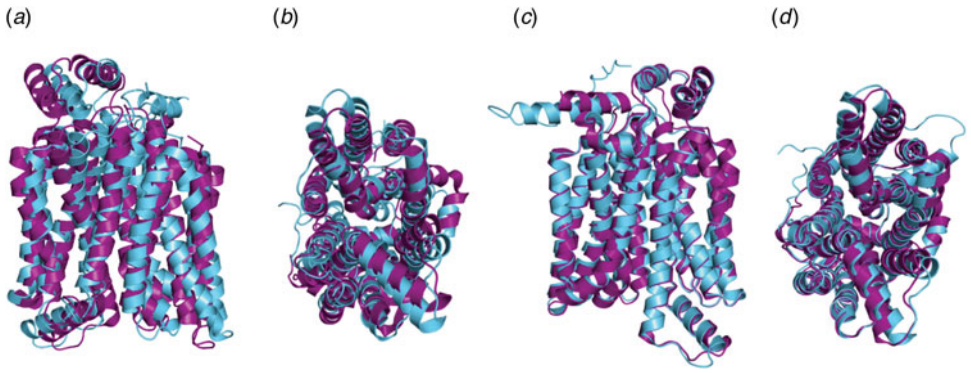

Fig. 9. Superimpositions of two glucose transporters GLUT1 and GLUT3 and their AlphaFold2 predicted QTY water-soluble variants. For each superimposition, the structures are shown in the side view (left), and the top view (right). The X-ray crystal structures of natural GLU1 (6THA, 2.4 Å), GLUT3 (4ZW9, 1.5 Å). (*a* and *b*) The crystal structure of native GLUT1 (magenta) is overlaid on the AlphaFold2 predicted water-soluble variant GLUT1^{QTY} (cyan). The RMSD is 1.55 Å for GLUT1 and GLUT1^{QTY}. (*c* and *d*) The GLUT3 CryoEM structure (magenta) is overlaid on the AlphaFold2 predicted water-soluble variant GLUT3^{QTY} (cyan). The RMSD is 1.03 Å for GLUT3 and GLUT3^{QTY}. For clarity, long N- and C-termini are removed.

Fig. 10. Two simple codes. (*a*) DNA code: base-pairing specificity and complementarity. The DNA code is bi-directional and reversible. (*b*) QTY code: matching shapes of hydrophobic and hydrophilic amino acid side chains. The QTY code is also bi-directional and reversible.

QTY variants (Fig. 9). Despite significant amino acid changes (overall >24% and TM >44%), the crystal structure of GLUT1 and the predicted structure of its QTY variant GLUT1^{QTY} superpose well (RMSD = 1.55 Å). Likewise, the crystal structure of GLUT3 closely resembles that of the predicted structure of its variant GLUT3^{QTY} (RMSD = 1.03 Å) (Smorodina *et al.*, 2022). We believe that this simple QTY code could be widely applicable to diverse membrane proteins that have TM α-helices.

Leonardo da Vinci remarked that '*Simplicity is the ultimate sophistication*'. Such a simple QTY code may stimulate researchers to systematically convert water-insoluble or aggregated proteins into water-soluble variants, not only to further protein structural studies, but also for biotechnological and nanotechnological applications and beyond.

Two simple molecular codes

Complementary DNA base pairing is governed by a simple code: A pairs with T and G pairs with C and vice versa (Fig. 10). Once the sequence of one DNA strand is identified, that of its complement is automatically known. When the right-handed model of the DNA double helix was first discovered in the early 1950s, based on X-ray fiber diffraction experiments and base-pairing considerations, it was thought it would exhibit limited conformational variation beyond the two forms apparent in fiber diffraction images (Frank-Kamenetskii, 1993). Through systematically studying single-crystal structures of DNA alone and in complex with proteins at high resolution, it was revealed that the double helix is not uniform at all (Rich, 1983; Frank-Kamenetskii, 1993). Thus, DNA not only forms a left-handed double helix (Z-form) (Wang *et al.*, 1979), but it also adopts a wide range of conformational variants, including A-form, B-form, C-form, D-form, E-form, H-form, writhing and kinked species, and more (Egli, 2022). In retrospect, it is not surprising that DNA structure exhibits many variations, consistent with its diverse biological activities and functions in cells.

The three chemical types of α-helix, although known to structural biologists (Eisenberg *et al.*, 1982, 1984), are much less known to non-structural biologists and non-protein scientists. We here explicitly classify three chemically distinct α-helix types to put a spotlight on them. This classification will likely provide us with an opportunity to fully understand subtle variations in the three chemical types of α-helix in high-resolution structures, especially when combined with the results of artificial intelligence and machine learning tools. We believe that the explicit classification of the three chemically distinct types of α-helix combined with the simple QTY code (Fig. 10) will likely stimulate scientists not only to further study protein structure and function, but also use the concept for protein design.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033583522000063.

Acknowledgements. We thank Eva Smorodina for help with preparing Figs 8 and 9 and Figure S1. We are grateful to Ander Liljas for helpful discussion, suggestions and corrections. We also thank Dorrie Langsley for careful English editing.

References

Armen R, Alonso DO and Daggett V (2003) The role of alpha-, 3(10)-, and pi-helix in helix-->coil transitions. *Protein Science* **12**, 1145–1157.

Barron CC, Bilan PJ, Tsakiridis T and Tsiani E (2016) Facilitative glucose transporters: implications for cancer detection, prognosis and treatment. *Metabolism: Clinical and Experimental* **65**, 124–139.

Bragg L, Kendrew J and Perutz M (1950) Polypeptide chain configurations in crystalline proteins. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **203**, 321–357.

Brändén CI and Tooze J (1991) Introduction to protein structure. *Garland Science*, First edition. p. 15.

Chou PY and Fasman GD (1974) Prediction of protein conformation. *Biochemistry* **13**, 222–245.

Creighton TE (1992) *Proteins: Structure and Molecular Properties*, 2nd Edn. San Francisco: WH Freeman and Company.

Custódio TF, Paulsen PA, Frain KM and Pedersen BP (2021) Structural comparison of GLUT1 to GLUT3 reveal transport regulation mechanism in sugar porter family. *Life Science Alliance* **4**, e202000858.

Deng D, Sun P, Yan C, Ke M, Jiang X, Xiong L, Ren W, Hirata K, Yamamoto M, Fan S and Yan N (2015) Molecular basis of ligand recognition and transport by glucose transporters. *Nature* **526**, 391–396.

Doran MH, Pavadai E, Rynkiewicz MJ, Walklate J, Bullitt E, Moore JR, Regnier M, Geeves MA and Lehman W (2020) Cryo-EM and molecular docking shows myosin loop 4 contacts actin and tropomyosin on thin filaments. *Biophysical Journal* **119**, 821–830.

Dunitz JD (2001) Pauling's left-handed α-helix. *Angewandte Chemie International Edition* **40**, 4167–4173.

Dunitz JD (2013) La primavera, an autobiographical essay. *Helvetica Chimica Acta* **96**, 545–563.

Duquesne K and Sturgis JN (2010) Membrane protein solubilization. *Methods in Molecular Biology* **601**, 205–217.

Egli M (2022) DNA and RNA structure. In Blackburn GM, Egli M, Gait MJ and Watts JK (eds), *Nucleic Acids in Chemistry and Biology*, 4th Edn. Cambridge, UK: R. Soc. Chem., pp. 20–95, ISBN: 978-1-78801-904-0.

Egli M and Zhang S (2022) How the α-helix got its name. *Nature Reviews Molecular Cell Biology* **23**, 165.

Eisenberg D (2003) The discovery of the α-helix and β-sheet, the principle structural features of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11207–11210.

Eisenberg D, Weiss RM and Terwilliger TC (1982) The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature* **299**, 371–374.

Eisenberg D, Weiss RM and Terwilliger TC (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 140–144.

Fersht A (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. San Francisco: WH Freeman and Company.

Fodje MN and Al-Karadaghi S (2002) Occurrence, conformational features and amino acid propensities for the pi-helix. *Protein Engineering* **15**, 353–358.

Frank-Kamenetskii M (1993) *Unraveling DNA: The Most Important Molecule of Life*. Hoboken, NJ: John Wiley & Sons.

Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC and Skehel JJ (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838–1842.

Hanukoglu I and Fuchs E (1983) The cDNA sequence of a type II cytoskeletal keratin reveals constant and variable structural domains among keratins. *Cell* **33**, 915–924.

Hao SL, Jin D, Zhang S and Qing R (2020) QTY code-designed water-soluble Fc-fusion cytokine receptors bind to their respective ligands. *QRB Discovery* **1**, e4.

Lederer F, Glatigny A, Bethge PH, Bellamy HD and Matthew FS (1981) Improvement of the 2.5 Å resolution model of cytochrome b562 by re-determining the primary structure and using molecular graphics. *Journal of Molecular Biology* **148**, 427–448.

Lee K, Sharma R, Shrestha OK, Bingman CA and Craig EA (2016) Dual interaction of the Hsp70 J-protein cochaperone Zuotin with the 40S and 60S ribosomal subunits. *Nature Structural & Molecular Biology* **23**, 1003–1010.

Lee CH, Kim MS, Li S, Leahy DJ and Coulombe PA (2020) Structure-Function analyses of a keratin heterotypic complex identify specific keratin regions involved in intermediate filament assembly. *Structure* **28**, 355–362, e4.

Liljas A, Liljas L, Ash M-R, Lindblom G, Nissen P and Kjeldgaard M (2017) Textbook of Structural Biology 2nd Edtion (Series in Structural Biology) World Scientific Publishing Company, Singapore.

Lin SH and Guidotti G (2009) Purification of membrane proteins. *Methods in Enzymology* **463**, 619–629.

Linke D (2009) Detergents: an overview. *Methods in Enzymology* **463**, 603–617.

Mitra K, Steitz TA and Engelman DM (2002) Rational design of 'water-soluble' bacteriorhodopsin variants. *Protein Engineering* **15**, 485–492.

Mooers BH and Matthews BW (2004) Use of an ion-binding site to bypass the 1000-atom limit to structure determination by direct methods. *Acta Crystallographica Section D Biological Crystallography* **60**, 1726–1737.

Niederhut MS, Gibbons BJ, Perez-Miller S and Hurley TD (2001) Three-dimensional structures of the three human class I alcohol dehydrogenases. *Protein Science* **10**, 697–706.

Ogihara NL, Weiss MS, Degrado WF and Eisenberg D (1997) The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. *Protein Science* **6**, 80–88.

Pauling L and Corey RB (1950) Two hydrogen-bonded spiral configurations of the polypeptide chain. *Journal of the American Chemical Society* **72**, 5346–5347.

Pauling L, Corey RB and Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America* **37**, 205–211.

Perutz M, Rossmann MG, Cullis AF, Muirhead H and Will G (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-Ray analysis. *Nature* **185**, 416–422.

Qing R, Han Q, Fei T, Skuhersky M, Badr M, Chung H, Schubert T and Zhang S (2019) QTY Code designed thermostable and water-soluble chimeric chemokine receptors with tunable ligand-binding activities. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 25668–25676.

Rich A (1983) Right-Handed and Left-Handed DNA: Conformational Information in Genetic Material. *Cold Spring Harbor Symp. on Quant. Biol.*, vol. **47**, pp. 1–12.

Skuhersky M, Tao F, Qing R, Smorodina E, Jin D and Zhang S (2021) Comparing native crystal structures and AlphaFold2 predicted water-soluble G protein-coupled receptor QTY variants. *Life* **11**, 1285, 10.3390/life11121285.

Slovic AM, Summa CM, Lear JD and DeGrado WF (2003) Computational design of a water-soluble analog of phospholamban. *Protein Science* **12**, 337–348.

Slovic AM, Kono H, Lear JD, Saven JG and DeGrado WF (2004) Computational design of water-soluble analogues of the potassium channel KcsA. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1828–1833.

Slovic AM, Stayrook SE, North B and Degrado WF (2005) X-ray structure of a water-soluble analog of the membrane protein phospholamban: sequence determinants defining the topology of tetrameric and pentameric coiled coils. *Journal of Molecular Biology* **348**, 777–787.

Smorodina E, Tao F, Qing R, Jin D, Yang S and Zhang S (2022) Comparing 2 crystal structures and 12 AlphaFold2 predicted human membrane glucose transporters and their water-soluble QTY variants. *QRB Discovery* **3**, e5, 1–11.

Stryer L (1981) *Biochemistry*, 2nd Edn. San Francisco: WH Freeman and Company.

Takeda S, Yamashita A, Maeda K and Maeda Y (2003) Structure of the core domain of human cardiac troponin in the Ca(2+)-saturated form. *Nature* **424**, 35–41.

Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q and Wu B (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* **341**, 1387–1390.

Tao F, Tang H, Zhang S, Li M and Xu P (2022) Enabling QTY server for designing water-soluble α-helical transmembrane proteins. *MBio* **13**, e03604–21.

Tegler LT, Corin K, Skuhersky M, Pick H, Vogel H and Zhang S (2020) G protein-coupled receptor CXCR4 designed by the QTY code becomes more hydrophilic and retains cell-signaling activity. *Scientific Reports* **10**, 21371.

Wang AH, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, van der Marel G and Rich A (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686.

Xu Q, Axelrod HL, Abresch EC, Paddock ML, Okamura MY and Feher G (2004) X-Ray structure determination of three mutants of the bacterial photosynthetic reaction centers from Rb. sphaeroides; altered proton transfer pathways. *Structure* **12**, 703–715.

Yaguchi S, Yaguchi J and Tanaka H (2017) Troponin-I is present as an essential component of muscles in echinoderm larvae. *Scientific Reports* **7**, 43563.

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press

Zhang S, Lockshin C, Herbert A, Winter E and Rich A (1992) Zuotin, a putative Z-DNA binding protein in *Saccharomyces cerevisiae*. *EMBO Journal* **11**, 3787–3796.

Zhang SQ, Huang H, Yang J, Kratochvil HT, Lolicato M, Liu Y, Shu X, Liu L and DeGrado WF (2018*a*) Designed peptides that assemble into cross-α amyloid-like structures. *Nature Chemical Biology* **14**, 870–875.

Zhang S, Tao F, Qing R, Tang H, Skuhersky M, Corin K, Tegler L, Wassie A, Wassie B, Kwon Y, Suter B, Entzian C, Schubert T, Yang G, Labahn J, Kubicek J and Maertens B (2018*b*) QTY Code enables design of detergent-free chemokine receptors that retain ligand-binding activities. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E8652–E8659.

https://doi.org/10.1017/S0033583522000063 Published online by Cambridge University Press