Biomacromolecular structures: An introduction to EMBL-EBI resources

Joanna Argasinska [1]

- Proteins
- Beginner
- 1 hour

Introduction to structural biology and related resources at EMBL-EBI.

Learning objectives:

- Gain knowledge of biomacromolecular structures and their characteristics
- Be able to list common methods for obtaining biomacromolecular structure data
- Understand the protein structure resources which are available at the EBI

Biomacromolecular structures

Biomacromolecules [2] are large biological polymers, such as nucleic acids, proteins and carbohydrates, that are made up of monomers linked together.

For example, proteins are composed of monomers called amino acids. They are linked together to form a polypeptide chain, which folds into a three dimensional (3D) structure to constitute a functional protein (Figure 1). Often a functional macromolecule is composed of more than one such chain and sometimes requires small molecules [3] to assist in its function.

![Amino Acid Sequence and 3D Structure](image)

**Figure 1** The amino acid sequence (monomers) on the left and the corresponding 3D structure of ribosomal protein L2.
Although this course concentrates on the structural properties of biomacromolecules, it will also cover smaller molecules such as coenzyme [4], cofactor [5], prosthetic group [6], lipid [7], drug and metal ion because they might interact with and affect these biomacromolecules at the structural level.

When a small molecule binds specifically and reversibly to the biomacromolecule to form a complex and alters its activity or function it is called a ligand [8] (Figure 2).

![Figure 2 Example of a complex of a biomacromolecule with a ligand: the structure of human lysozyme. The ligand is a sugar molecule (NAG: N-Acetyl-D-Glucosamine) and is represented by red and blue spheres.](image)

Now let’s have a closer look at the different biomacromolecules (DNA, RNA and proteins) and some of the molecules they can interact with.

**DNA**

Genetic information is encoded in deoxyribonucleic acid (DNA) molecules. Therefore, DNA is an essential component of independently living organisms. Genes are the DNA segments that carry genetic information [9].

Some DNA sequences do not code for genes and have structural roles (for example, in structure of chromosomes), or are involved in regulating the use of the genetic information; for example, repressor sites are DNA sequences that allow binding of a repressor, which stops the process of gene expression [10].

DNA consists of two long polymers that run in opposite directions forming the regular geometry of the double helix. The monomers of DNA are called nucleotides. Nucleotides have three components: a base, a sugar (deoxyribose) and a phosphate residue. The four bases are adenine (A), cytosine (C), guanine (G) and thymine (T). The sugar and phosphate create a backbone down either side of the double helix. The bases interact via hydrogen bonds with complementary bases on the other DNA strand in the helix.

It is the sequence of these four bases that encodes genetic information. The interaction between two bases on opposite strands via hydrogen bonds is called base pairing. As shown in figure 3, adenine forms a base pair [11] with thymine and guanine forms a base pair with cytosine. These are the most common base pairing patterns but alternative patterns also are possible.
Figure 3 Chemical structure of DNA. Two polymers composed of phosphate-deoxyribose backbones and four bases: A, C, G, T linked via two (A-T) or three (G-C) hydrogen bonds. The two strands run in opposite directions. Image from wikipedia [12].

The majority of DNA in a cell is present in the so-called B-DNA structure. However, it can also adopt other 3D structures (Figure 4). Z-DNA, found in DNA bound to certain proteins, is a rarer structure. In Z-DNA, the bases have been chemically modified by methylation [13] and the strands turn in a left-handed helix, the opposite direction from that of the B form. Z-DNA formation is an important mechanism in modulating chromatin structure (2 [9]). The A-DNA structure, which has a wider right-handed helix, occurs only in dehydrated samples of DNA, such as those used in X-ray crystallography [14].
RNA

Like DNA, RNA (ribonucleic acid) is essential for all known forms of life. RNA monomers are also nucleotides. Unlike DNA, RNA in biological cells is predominantly a single-stranded molecule. While DNA contains deoxyribose, RNA contains ribose, characterised by the presence of the 2'-hydroxyl group on the pentose ring (Figure 5). This hydroxyl group make RNA less stable than DNA because it is more susceptible to hydrolysis. RNA contains the unmethylated\([16]\) form of the base thymine called uracil\([17]\) (U) (Figure 6), which gives the nucleotide uridine.
Figure 5 Chemical structure of RNA: nucleotides containing a ribose sugar (carbons numbered 1' through 5') with a base attached to the 1' position. From wikipedia [18].

RNA performs a variety of functions in the cell. Messenger RNA (mRNA) carries the genetic information that directs the synthesis of proteins. Some viruses use RNA instead of DNA as their genetic material. Most of the RNA, however, does not code for proteins. These RNAs are called noncoding (ncRNA) and can be encoded by their own RNA genes or can derive from mRNA introns. Transfer RNA (tRNA) [19] and ribosomal RNA (rRNA) [20] are involved in the process of translation [21]. There are also non-coding RNAs involved in gene regulation [22], RNA processing and other processes.

Most RNA molecules contain short self-complementary sequences that fold and pair with each other into highly structured forms. These base-pairing interactions are part of RNA secondary structure [23]. The unpaired regions form structures such as hairpin loops, bulges and internal loops, which may be of functional importance (Figure 7). Examples include Rho-independent terminator [24] stem-loops and the tRNA cloverleaf [19].
Figure 7 Secondary and tertiary structure of tRNA. Unpaired regions are in grey and paired regions are in colour. Image from wikipedia [25].

The functional form of single-stranded RNA molecules, just like proteins, frequently requires a specific tertiary (3D) structure. RNA can also form RNA-RNA and DNA-RNA duplexes. Most RNA structures in the Protein Data Bank [26] (PDB [27]) (archive of macromolecular structural data) (3 [9]) contain double-stranded RNA folded into tertiary structures.

Some RNA structures provide binding sites for other molecules and have chemically active centres. An example, (Figure 8) is the molecular recognition of vitamin B12 by an RNA structure (4 [9]). Vitamin B12 binding to RNA regulates hepatitis C virus function (5 [9]).
Figure 8 The structure of vitamin B12 bound to RNA. The molecular recognition is achieved by the folding of an initially unstructured RNA around its ligand [8].

Proteins

Proteins are biomacromolecules [2] present in all organisms and they have a large variety of functions. Proteins are linear chains of L-L-α-amino acids [28] (Figure 9). There are 22 different genetically encoded (the genetic code is specified for them) amino acids.

Figure 9 Structure of an alpha amino acid. R represents amino-acid-specific side-chain [29], that is attached to the 7-carbon [30]. Each side chain [31] has different physico-chemical properties that can be exploited by proteins to perform different roles. Figure from wikipedia [32].
To be able to perform their biological function, proteins fold into specific spatial conformations. The formation of binding pockets and local 3D structures allow proteins to create different chemical environments through which they can specifically interact with other biomacromolecules, small molecules [3] or water. Examples of the interacting partners are: substrates, metal ions, prosthetic groups, cofactors and coenzymes.

Reversible structural changes, which create alternative structures of the same protein are referred to as different conformers. The transitions between them are called conformational changes.

Examples of functions of proteins include:

- **Structural**: offering stiffness and rigidity to fluid biological components. Collagen [33] is the most abundant structural protein in mammals and forms a triple helix.

- **Catalysis of chemical reactions** as enzymes [34]. Only a small region of an enzyme called the active site [35] binds the substrate and contains the catalytic residues. Example: Trypsin [36], found in the digestive system, hydrolyses [37] proteins.

- **Receptors**: they usually have a ligand [8]-binding site [38] on the cell surface and an effector domain [39] within the cell, which may have enzymatic activity or may undergo a conformational change. For example: G Protein-coupled receptor [40].

- **Channels** for molecules to pass through the cell membrane, for example the potassium channel.

- **Transport**: Those proteins bind small molecules and transport them to other locations in the cell or organism. For example haemoglobin [41] transports oxygen.

### Levels of Protein Structure

#### Primary

The primary protein structure refers to the sequence of amino acids and the location of disulfide bonds (Figure 10).

The amino acids [42] when linked by peptide bonds [43] are referred to as residues. Short chains of amino acid [44] residues are often called (oligo-)peptides.
Figure 10 Chemical structure (bottom) and the 3D structure (top) of a peptide bond between two adjacent amino acid residues. The N-terminal end of a polypeptide contains a free amino group [45] and the C-terminal end contains a free carboxylate group [46]. Image from wikipedia.

Secondary

Protein structures are also classified by their secondary structure [23]. Secondary structure refers to regular, local structure of the protein backbone, stabilised by intramolecular and sometimes intermolecular hydrogen bonding of amide [47] groups.

There are two common types of secondary structure (figure 11). The most prevalent is the alpha helix [48].

The alpha helix (?) has a right-handed spiral conformation, in which every backbone N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues before it in the sequence.

The other common type of seconday structure is the beta strand [49]. A Beta strand (? strand) is a stretch of polypeptide chain, typically 3 to 10 amino acids long, with its backbone in an almost fully extended conformation. Two or more parallel or antiparallel adjacent polypeptide chains of beta strand stabilised by hydrogen bonds form a beta sheet. For example, the proteins in silk have a beta-sheet structure. Those local structures are stabilised by hydrogen bonds and connected by tight turns [50] and loose, flexible loops.
A common way for researchers to look at the conformations of amino acids in proteins is to use a Ramachandran plot [51] (Figure 12). If successive amino acids are found in particular so-called "favourable" regions of the plot, these tend to form particular secondary structures (sheets, strands and helices). However, it should be noted that loop and turn residues are also found in these areas and it is possible to find individual amino acids in "unfavourable" regions.

Figure 11 Alpha helix (blue) and antiparallel beta sheet composed of three beta strands (yellow and red).
The Ramachandran plot \cite{52} is a plot of the torsional angles (angles between two planes) - \( \psi \) \( (?) \) and \( \phi \) \( (?) \) \cite{53} - of amino acids contained in a peptide. It is used to show the ranges of angles that are permissible and the main types of structure adopted by a polypeptide chain (for example, \( \alpha \) helix, \( \beta \) sheet). By making a Ramachandran plot, protein structural scientists can determine which torsional angles are permitted and can obtain insight into the structure of peptides.

**Tertiary**

The spatial arrangement of \textit{secondary structure} \cite{23} elements results in the formation of the tertiary structure or fold of a protein. The tertiary structure is held together by \textit{non-covalent interactions} \cite{54} (hydrogen bonding, ionic interactions, van der Waals forces, and hydrophobic packing), disulphide bonds and metal ion coordination.

An example of the tertiary structure is a single-\textit{domain} \cite{39} globular protein. Globular proteins \cite{6} \cite{9} are sphere-like proteins that are more or less soluble in aqueous solutions (the other two protein classes are \textit{membrane} \cite{55} and \textit{fibrous proteins} \cite{56}) (Figure 13).
Quaternary

Some proteins form assemblies (units) with other molecules - this is called the quaternary structure (Figure 14).

Two examples are: haemoglobin [41] which is an assembly of four globular proteins and the actin microfilament [58], composed of many thousands actin [59] molecules.

In the PDB [27], assemblies of different proteins or different macromolecules are also referred to as the quaternary structure; for example, ribosome [60] is described as a quaternary structure.
Fold

Proteins form stable 3D structures or folds (Figure 15). However, it is generally accepted that there are a limited number of stable folds, probably on the order of thousands.

The main folding unit is a structural domain [39].

Folds are classified into 3 main types based on their secondary structure [23] content:

- All α: built almost exclusively from α-helixes
- All β: built almost exclusively from β-strands
- Mixed α and β: containing a mixture of both: either alternating α and β (?) or any order of α and β (?) + (?)

A classic example of a common mixed fold is the α/β barrel, also called the TIM barrel, consisting of eight α-helices and eight parallel β-strands that alternate along the peptide backbone. The structure is named after Triosephosphate Isomerase (TIM), a conserved glycolytic enzyme responsible for glucose degradation. This fold has been found in many different enzyme families catalysing completely unrelated reactions.

Figure 15 Example of a protein fold. Top view of a triosephosphateisomerase (TIM) barrel, colored from blue (N-terminus [61]) to red (C-terminus [62]) - (N-term/C-term). Have a look at this protein in PDBe [63]: 8TIM [64]

The two main resources that classify folds are: SCOP [65] and CATH [66]. Because of the differences in definitions and classification of folds the number of folds listed in each of them is different. (7, 8, 9 [9]). Each structure in PDBe is shown using both classifications.

One of the PDBe Services called Similar Structures (PDBe Fold [67]) can be used to compare the structures of proteins to identify conserved folds that may not be similar in sequence. An example of similar structure proteins is RsbT co-antagonist protein rsbRA (in PDBe 2bml [68]), which has a globin fold and shares a very similar fold structure with Sperm Whale Myoglobin (in PDBe 3o89 [69]). However, the sequence identity between these two
proteins is only 12.2% (as calculated by PDBFold).

**Domains**

Structural domains (the units of fold) are independently stable tertiary structures of proteins (Figure 16). They are distinct functional and/or structural units and can evolve, exist and function independently. Therefore, the same domain [39] can be a part of different proteins, as is the case of G-protein signalling domains described in the Train Online course: Introduction to Protein Classification. [70]

**Figure 16** Three domains of *Thermus aquaticus* elongation factor EF-Tu: in blue (all-?), red (?,?,?) and green (all-?). You can learn more about domains by taking the Train Online course: Introduction to Protein Classification [71].

**Structural Motifs**

Structural motifs are short segments of protein 3D structure, which are spatially close but not necessarily adjacent in the sequence. Structural motifs may be conserved in a large number of different proteins [10] [9]). Their role may be structural or functional.

An example of a structural motif [72] that generally performs a structural role is a beta-turn (Figure 17). A beta turn
consists of four consecutive residues where the polypeptide chain folds back on itself by nearly 180 degrees (11 [9]).

Figure 17 Beta turn structural motif. The four residues are shown in stick representation, the strands are as cartoon arrows. Nitrogen is shown in blue, oxygen in red and the chiral [73] carbon of Glycine residue in yellow.

An example of a structural motif that has an important functional role is the helix-turn-helix motif (Figure 18) which can bind DNA. This is a structural feature that is difficult to identify from the amino acid [44] sequence alone.
**Figure 18**: The transcription factor SATB1 (right: green and blue) contains a helix-turn-helix motif (blue) that can to bind DNA (left).

PDBe [63] offers a service specific for structural motifs and functional sites: Motifs and Sites (PDBMotif [74]). PDBeMotif can be used to examine the characteristics of the binding sites of single proteins or classes of proteins by combining protein sequence, chemical structure and 3D data in a single search.

---

**Ligand/small molecule**

In the PDB [27], small molecules [3] (low molecular weight organic compounds that are not part of polymers) are frequently associated with biomacromolecules [2].

Any substance that binds specifically and reversibly to a biomacromolecule to form a larger complex and alters its activity or function is called a ligand [8]. In the PDB, drugs, metals, and small molecules are also called ligands.

Here we show examples of ligands bound to protein (figure 19) and DNA (figure 20).

**Figure 19** An example of a ligand binding to cytochrome c-553 (in green, yellow, blue) is prosthetic group [6] Haem containing an iron ion.
Figure 20 An example of a ligand binding to DNA: drug, bleomycin a2.

Information about ligand interactions can be found in Motif [72] and Sites Services link on the PDBe [63] website at PDBMotif [74] and about ligands at PDBeChem [75].

Where do the data come from?

Where do the data come from?

There are three major methods to determine structures of biomacromolecules [2]: X-ray crystallography [14], NMR spectroscopy and cryo-electron microscopy [76].

Some data deposited in PDB [27] is captured using electron crystallography [77], fibre diffraction [78] and neutron diffraction [79].

Next, we will summarise the three main methods.

X-ray crystallography

X-ray crystallography [14] uses the diffraction [80] of X-rays [81] to determine the structure of macromolecules. Over 85% of the structures in the PDB [27] derive from this experimental method. Many macromolecules and macromolecular complexes form crystals under the right conditions. These are routinely frozen to cryogenic [82] temperatures and exposed to synchrotron [83] X-ray radiation to give a diffraction pattern [84] (see Figure 21). Once electron density [85] is calculated for a crystal then a structural model can be fitted into it. Both the individual diffraction points and the structural model are archived in the PDB.
Figure 21 Workflow for X-ray crystallography. From Wikipedia. [86]

**Nuclear magnetic resonance spectroscopy**

Nuclear magnetic resonance [87] (NMR) spectroscopy is the second most common method of structure determination, providing ~14% of all entries in the PDB [27]. It utilises the fact that some atomic nuclei are magnetically active and can emit radiofrequency signals when placed in a strong external magnetic field (on the order of 10-20 Tesla, which is almost a million times stronger than the Earth's magnetic field on the surface). Typical data collection may take 2-3 weeks for a small soluble protein, but can be substantially longer for larger systems.

The measurements in NMR spectroscopy are a number of different complex spectra that report, among other things, on the chemical environment for the magnetically active nuclei (most commonly $^1$H, $^{13}$C and $^{15}$N), on chemical bond connections between nuclei and on short distances between specific atoms. These short distances constitute constraints for molecular dynamics (MD) simulation software, which attempts to satisfy as many of them as possible. The outcome of MD simulation is an ensemble of structures (usually 10-20) which, when combined, best satisfy the experimental data. The whole ensemble is deposited in the PDB.


In order to provide you with a real case scenario, we illustrate an example below (Figure 22):
Figure 22 15N-HSQC (heteronuclear single quantum coherence) spectrum of a protein. It is a two-dimensional spectrum, where each peak corresponds to an N-H (amide [47]) group and essentially labels a residue of the protein. The HSQC spectrum is therefore often called the "fingerprint" experiment, as each protein will have a unique pattern [84] of peaks. The horizontal axis gives the chemical shifts of hydrogens, while the vertical - that of nitrogens. Chemical shift is a parameter [91] which is very sensitive to the exact chemical environment of a particular atom, and can therefore act as a "label" or "reporter" for that atom. The protein sample is enriched with the 15N isotope of nitrogen, which interacts with the magnetic field stronger than the more common 14N isotope.

Cryo-electron microscopy

Electron microscopy [76] uses beams of electrons to form images of macromolecules and complexes. Electron microscopy can reveal smaller details than light microscopes owing to the wave length differences between electrons and photons. However, the resolution [92] is lower than that of X-ray crystallography [14] or NMR spectroscopy. Images of macromolecules can be obtained in artificially-stained states or frozen in glassy 'vitreous' ice [93] layers (Figure 23). Images from a large number of orientations of a macromolecule can be combined computationally to give a 3D reconstruction of the electron density. The reconstructed 3D density map is deposited in the Electron Microscopy Data Bank (EMDB) [94]. In many cases a higher resolution X-ray or NMR structure can be positioned in this electron density volume to allow its interpretation. It is these positioned 'fitted' structures that are then made available in the PDB [27].
Biomacromolecular structure resources at the EBI

PDBe [63]

EMBL-EBI's Protein Data Bank in Europe (PDBe) is the European resource for the collection, organisation and dissemination of data about biological macromolecular structures. PDBe is one of four partners in the Worldwide Protein Data Bank (wwPDB).

PDBe also develops and maintains several tools to search, retrieve, explore and analyse the data.

For an overview of the PDBe resources to search and visualise data see the PDBe quick guide [98]. For a quick search, simply go to the PDBe website [97] and use the search box.

PDBe tool bar [97] gives quick access to all PDBe services, tool for browsing enzymes, folds, protein families etc. and advanced searches including sequence or structure similarity search. Structures can also be deposited here and the latest PDB [27] and EMDB [99] entries can be viewed.

Summary

- Biomacromolecules [2] are large biological polymers, such as nucleic acids, proteins and carbohydrates, that are made up of monomers. Some of them interact with smaller molecules called ligands.

- DNA and RNA are composed of nucleotides and form specific three-dimensional structures. DNA generally forms a double helix and can be bound by a number of macromolecules including proteins (e.g. helix-turn-
RNA is mostly a single stranded polymer that can fold into a vast number of complex three-dimensional shapes. mRNA carries genetic information that directs protein synthesis. Non-coding RNAs, for example tRNA and rRNA, are functional RNAs that are not translated into proteins.

Proteins are linear chains composed of amino acids. To be able to perform their biological functions such as chemical catalysis, structural roles, transport and intracellular signalling, proteins fold into specific spatial conformations.

There are four levels of protein structure: primary, secondary, tertiary and quaternary.

The tertiary and quaternary structural data submitted to PDB is determined mainly by X-ray crystallography, but also by NMR spectroscopy and cryo-electron microscopy methods.

Protein Data Bank in Europe (PDBe) is a part of the Wordwide Protein Data bank (wwPDB). PDBe is a European resource for the collection, organisation and dissemination of data about nucleic acid and protein structure as well as molecules that may interact with them.

PDBe also develops and maintains several tools to explore and analysis biomacromolecular data.

Quiz: Biomacromolecular structures

Questions: 5
Attempts allowed: Unlimited
Available: Always
Pass rate: 75%
Backwards navigation: Allowed

Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Learn more

Find out more

Quips: "Quite Interesting PDB Structures" are short articles and tutorials on structures picked from the PDB archive.

Wizard provides help in finding information you are looking for.

Recommended literature


Recommended courses

EMBO Practical Course: Computational structural biology - from data to structure to function

Date: 16 - 20 April 2012
Venue: EMBL-EBI, Hinxton

References


Contributors

Joanna Argasinska [1]
EMBL-EBI
Intern

Joanna Argasinska completed her PhD in the Department of Biochemistry, University of Cambridge in 2004. She studied proteins controlling translation of stored maternal mRNAs during Xenopus oogenesis in Dr Standart's laboratory. She then joined Prof Smith group at the Gurdon Institute, Cambridge to study functions of transcription factors in Xenopus development. After working for a Cambridge-based biotech company, she joined the Training Group at EBI for an internship working on Train Online.

Martyn Symmons [127]
EMBL-EBI
Scientific Curator

Martyn F. Symmons received his PhD in protein crystallography from Birkbeck College University of London in 2004. This was for his structural solution of the first known PNPase which is a key bacterial and mitochondrial RNase controlling virulence and cell senescence. He was then awarded an Oppenheimer Research Fellowship at the University of Cambridge to work on bacterial antibiotic-efflux pumps in the Department of Pathology. He is currently a Curator in the PDBe [128] team at EBI. In his spare time he continues to develop methods for
assembling macromolecules from distance restraint data. Most recently he applied FRET restraints to assemble the core 5'UTR [129] region of HIV-1 genomic RNA.

James Watson [130]
EMBL-EBI
Senior Scientific Outreach Officer

James Watson studied biochemistry at the University of Glasgow and obtained his PhD on structural motifs in proteins. He joined the EBI as a researcher in the Thornton group investigating the prediction of protein function from three-dimensional structure. As part of his work with the Midwest Center for Structural Genomics [131], he has been involved in the development and testing of the ProFunc server - a fully automated protein function prediction server. He is now the Senior Scientific Training Officer in the Training group, joining in 2008, and is responsible for organising and teaching in-house training events as well as the day to day running of the IT training facilities. He also trains on a variety of resources at EBI Roadshows across Europe and beyond.

Aleksandras Gutmanas [132]
EMBL-EBI
Scientific Officer

Aleksandras Gutmanas defended his PhD in biophysics with emphasis on Nuclear Magnetic Resonance (NMR) Spectroscopy and Molecular Dynamics Simulations in 2003 at the Dept. of Biochemistry and the Swedish NMR Centre, Gothenburg University, Sweden. He then moved on as a postdoc and staff scientist in the group of Dr. C. H. Arrowsmith, Toronto, Canada, part of the Northeast Structural Genomics Consortium (NESG). Since June 2010, he joined the PDBe group at EBI as NMR Project Leader. His interest is in NMR Spectroscopy, Macromolecular structures and structure validation.
Gerard Kleywegt

EMBL-EBI
PDBe Team Leader

Gerard Kleywegt received his PhD in Chemistry from the University of Utrecht (The Netherlands) in 1991. After 6 months with a California-based software company he moved to the University of Uppsala (Sweden) to do post-doctoral research in the laboratory of Alwyn Jones. He became an independent investigator in 1996 and played a pivotal role in the Swedish Structural Biology Network (SBNet) for more than a decade, first as its coordinator and later as its programme director. He was appointed Professor of Structural Molecular Biology in 2009. In the same year, he moved to Cambridge and joined EMBL-EBI as Head of the Protein Data Bank in Europe (PDBe).

Source URL: https://www.ebi.ac.uk/training/online/course/biomacromolecular-structures-introduction-ebi-reso

Links
[1] https://www.ebi.ac.uk/training/online/trainers/joanna
[2] https://www.ebi.ac.uk/training/online/glossary/biomacromolecules
[3] https://www.ebi.ac.uk/training/online/glossary/small-molecules
[4] https://www.ebi.ac.uk/training/online/glossary/coenzyme
[5] https://www.ebi.ac.uk/training/online/glossary/cofactor
[6] https://www.ebi.ac.uk/training/online/glossary/prosthetic-group
[7] https://www.ebi.ac.uk/training/online/glossary/lipid
[8] https://www.ebi.ac.uk/training/online/glossary/ligand
[9] https://www.ebi.ac.uk/training/online/course/biomacromolecular-structures-introduction-ebi-reso/references
[10] https://www.ebi.ac.uk/training/online/glossary/gene-expression
[11] https://www.ebi.ac.uk/training/online/glossary/base-pair
[13] https://www.ebi.ac.uk/training/online/glossary/methylation
[14] https://www.ebi.ac.uk/training/online/glossary/x-ray-crystallography
[22] https://www.ebi.ac.uk/training/online/glossary/gene-regulation
[23] https://www.ebi.ac.uk/training/online/glossary/secondary-structure
[29] https://www.ebi.ac.uk/training/online/glossary/side-chain
[33] https://www.ebi.ac.uk/training/online/glossary/active-site
[34] http://en.wikipedia.org/wiki/Trypsin
[36] https://www.ebi.ac.uk/training/online/glossary/binding-site
[37] https://www.ebi.ac.uk/training/online/glossary/domain
[44] https://www.ebi.ac.uk/training/online/glossary/amino-acid
[47] https://www.ebi.ac.uk/training/online/glossary/amide
[48] https://www.ebi.ac.uk/training/online/glossary/alpha-helix
[49] https://www.ebi.ac.uk/training/online/glossary/beta-strand
[51] https://www.ebi.ac.uk/training/online/glossary/ramachandran-plot
[53] http://proteopedia.org/wiki/index.php/Psi_and_Phi_Angles
[61] https://www.ebi.ac.uk/training/online/glossary/n-terminus
[62] https://www.ebi.ac.uk/training/online/glossary/c-terminus
[63] https://www.ebi.ac.uk/training/online/glossary/pdbe
[65] http://scop.mrc-lmb.cam.ac.uk/scop/
[67] http://www.ebi.ac.uk/msd-srv/ssm/
[70] https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/protein-classification/family-and-domain-based-protei
[71] https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi
[72] https://www.ebi.ac.uk/training/online/glossary/motif
[74] http://www.ebi.ac.uk/pdbe-site/pdbemotif/
[75] http://www.ebi.ac.uk/pdbe-srv/pdbechem/
[76] https://www.ebi.ac.uk/training/online/glossary/electron-microscopy
[84] https://www.ebi.ac.uk/training/online/glossary/pattern
[87] https://www.ebi.ac.uk/training/online/glossary/nuclear-magnetic-resonance
[88] http://www.ebi.ac.uk/pdbe-apps/nmr/olderado/
[89] https://www.ebi.ac.uk/training/online/glossary/clustering
[90] http://www.ebi.ac.uk/pdbe-apps/nmrviewer/
[91] https://www.ebi.ac.uk/training/online/glossary/parameter
[94] http://www.ebi.ac.uk/pdbe/emdb/
[97] http://www.ebi.ac.uk/pdbe/
[98] https://www.ebi.ac.uk/training/online/course/pdbe-quick-tour
[99] https://www.ebi.ac.uk/training/online/glossary/emdb