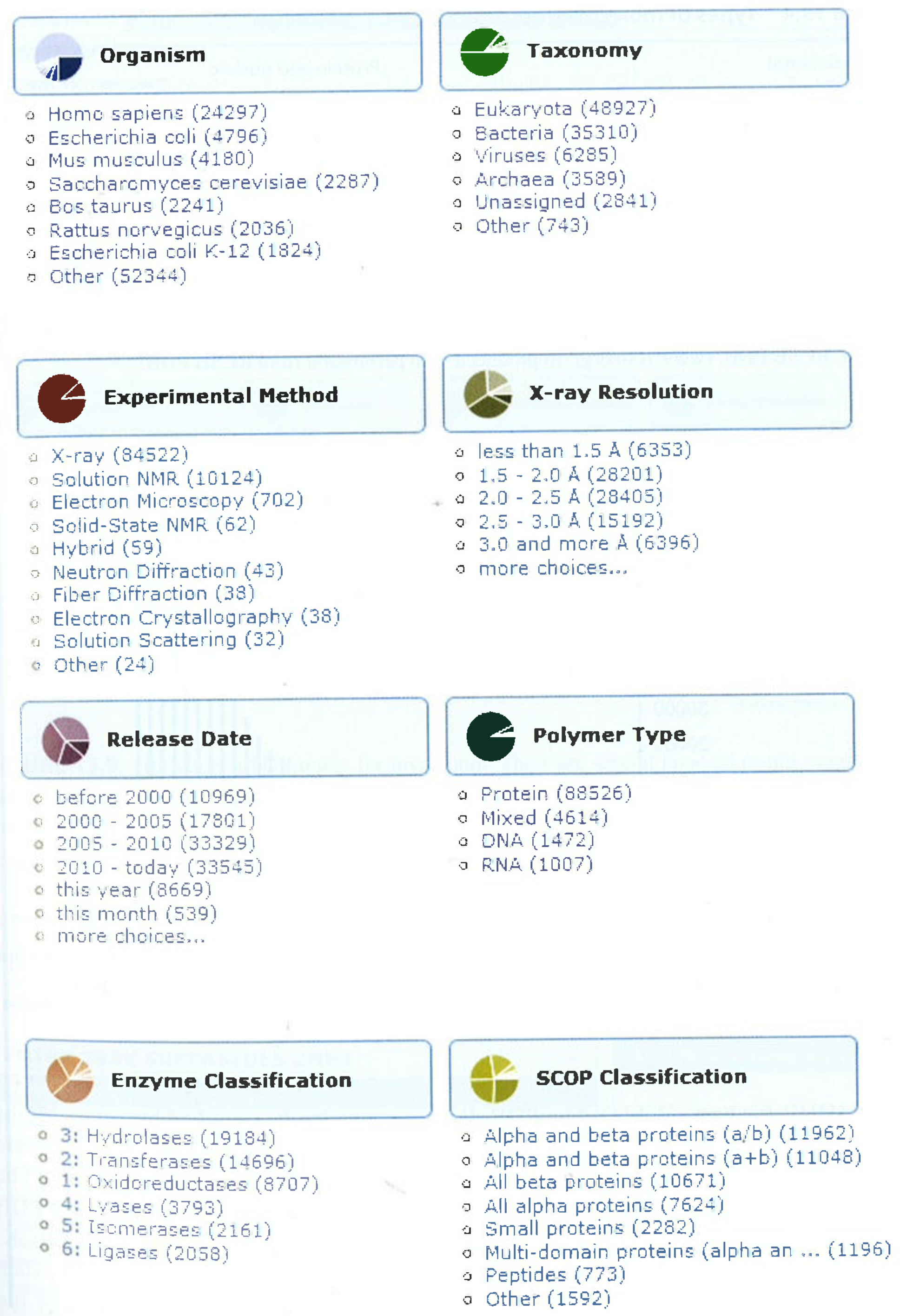
PROTEIN DATA BANK

The PDB was established at Brookhaven National Laboratories in Long Island in 1971. Initially, it contained seven structures. It moved to the Research Collaboratory for Structural Bioinformatics (RCSB) in 1998. PDB is accessed at ⊕ http://www.rcsb.org/pdb/ or ⊕ http://www.pdb.org (WebLink 13.14).

Once a protein sequence is determined, there is one principal repository in which the structure is deposited: the Protein Data Bank (PDB) (Rose *et al.*, 2013; reviewed in Berman, 2012; Berman *et al.*, 2013a–c; Goodsell *et al.*, 2013). A broad range of primary structural data is collected, such as atomic coordinates, chemical structures of cofactors, and descriptions of the crystal structure. The PDB then validates structures by assessing the quality of the deposited models and by how well they match experimental data.

The main page of the PDB website includes categories by which information may be accessed (**Fig. 13.7**). This database currently has over 100,000 structure entries (**Table 13.4**), with new structures being added at a rapid rate (**Fig. 13.8**). The database can be accessed directly by entering a PDB identifier into the query box on the main page, that is, by entering an accession number consisting of one number and three letters



Organism
- Homo sapiens (24297)
- Escherichia coli (4796)
- Mus musculus (4180)
- Saccharomyces cerevisiae (2287)
- Bos taurus (2241)
- Rattus norvegicus (2036)
- Escherichia coli K-12 (1824)
- Other (52344)

Taxonomy
- Eukaryota (48927)
- Bacteria (35310)
- Viruses (6285)
- Archaea (3589)
- Unassigned (2841)
- Other (743)

Experimental Method
- X-ray (84522)
- Solution NMR (10124)
- Electron Microscopy (702)
- Solid-State NMR (62)
- Hybrid (59)
- Neutron Diffraction (43)
- Fiber Diffraction (38)
- Electron Crystallography (38)
- Solution Scattering (32)
- Other (24)

X-ray Resolution
- less than 1.5 Å (6353)
- 1.5 - 2.0 Å (28201)
- 2.0 - 2.5 Å (28405)
- 2.5 - 3.0 Å (15192)
- 3.0 and more Å (6396)
- more choices...

Release Date
- before 2000 (10969)
- 2000 - 2005 (17801)
- 2005 - 2010 (33329)
- 2010 - today (33545)
- this year (8669)
- this month (539)
- more choices...

Polymer Type
- Protein (88526)
- Mixed (4614)
- DNA (1472)
- RNA (1007)

Enzyme Classification
- 3: Hydrolases (19184)
- 2: Transferases (14696)
- 1: Oxidoreductases (8707)
- 4: Lyases (3793)
- 5: Isomerases (2161)
- 6: Ligases (2058)

SCOP Classification
- Alpha and beta proteins (a/b) (11962)
- Alpha and beta proteins (a+b) (11048)
- All beta proteins (10671)
- All alpha proteins (7624)
- Small proteins (2282)
- Multi-domain proteins (alpha an ... (1196)
- Peptides (773)
- Other (1592)

FIGURE 13.7 The PDB is the main repository for three-dimensional structures of proteins and other macromolecules. Information about PDB holdings is organized into categories such as organism, taxonomy, experimental method (greater than 85% of which are derived from X-ray structure determination), and resolution (with less than 1.5 Å corresponding to the highest resolution structures). The home page of PDB allows queries such as a PDB identifier (e.g., 3RGK for a myoglobin structure) or a molecule name.

Source: RCSB PDB (www.rcsb.org). Reproduced with permission from RCSB PDB.

TABLE 13.4 Types of molecules, according to PDB Holdings.

Experimental technique	Proteins	Nucleic acids	Protein and nucleic acid complexes	Other	Total
X-ray diffraction	88,991	1,608	4,398	4	95,001
NMR	9,512	1,112	224	8	10,856
Electron microscopy	539	29	172	0	740
Hybrid	68	3	2	1	74
Other	164	4	6	13	187
Total	99,274	2,756	4,802	26	106,858

Source: RCSB PDB. (www.rcsb.org). Reproduced with permission from RCSB PDB.

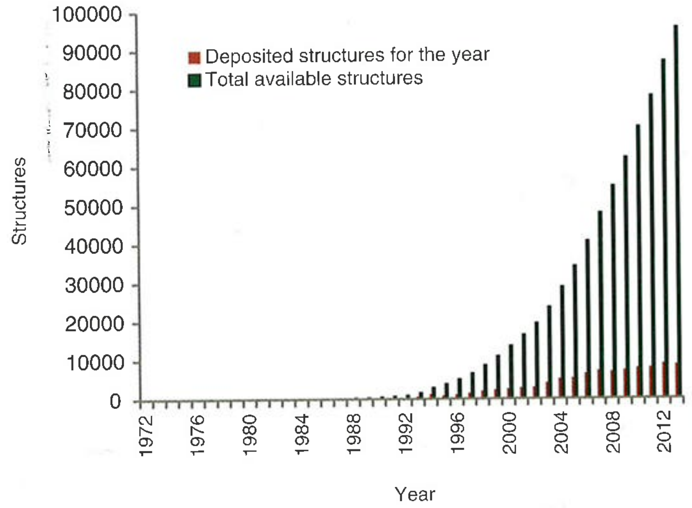
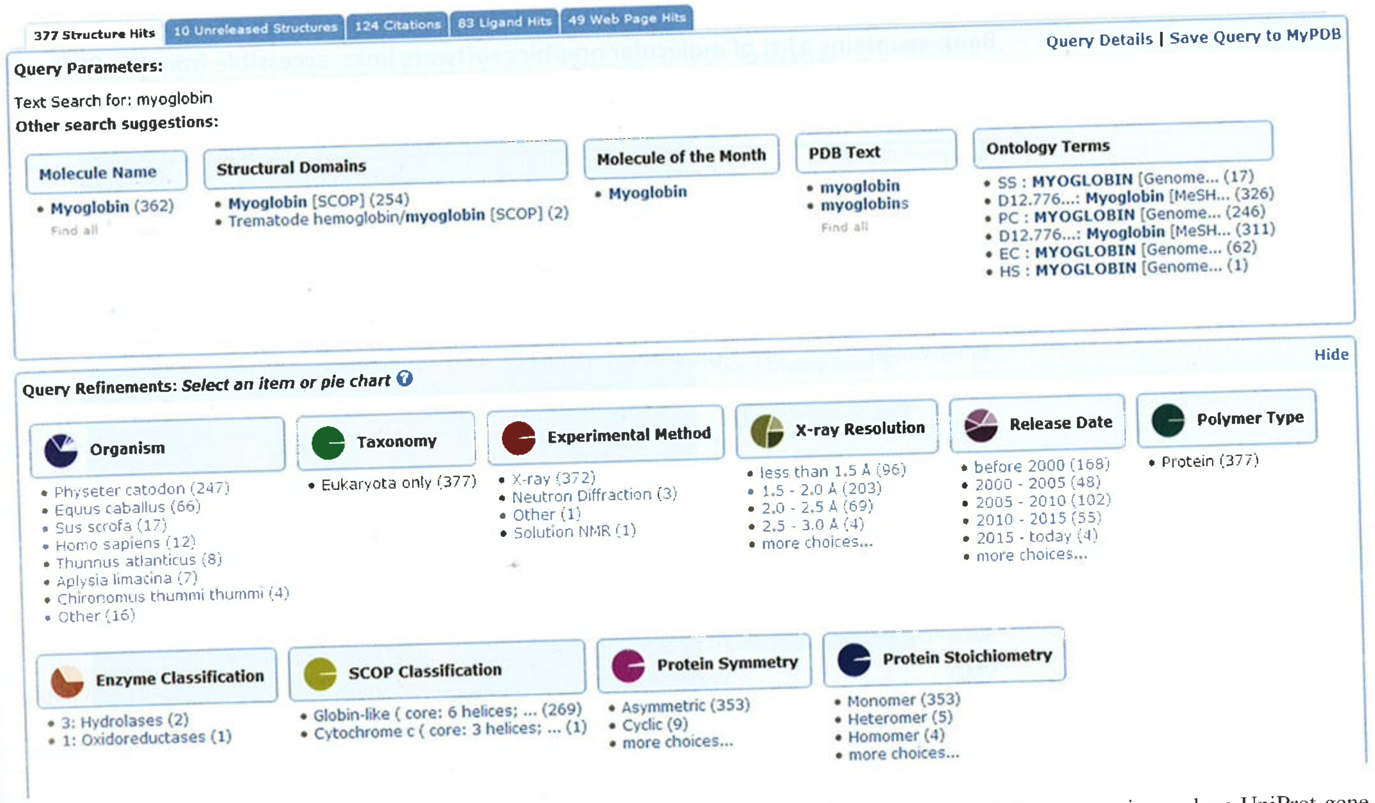


FIGURE 13.8 Number of searchable structures per year in PDB. The PDB database has grown dramatically in the past decade. The yearly (red) and total (green) numbers of structures are shown.

Source: RCSB PDB (www.rcsb.org). Reproduced with permission from RCSB PDB.

(e.g., 4HHB for hemoglobin). The PDB database can also be searched by keyword; the result of a keyword search for myoglobin is shown in **Figure 13.9**. In this case there are hundreds of results, and the list can be refined using options on the left sidebar. The result of searching for a specific hemoglobin identifier, 3RGK, links to a typical PDB entry (of which a portion is shown in **Fig. 13.10**). By clicking on an icon the 3RGK.pdb file can be downloaded locally for further analysis with a variety of tools such as DeepView. Information provided on the 3RGK page includes the resolution of the experimentally derived structure, the space group, and the unit cell dimensions of the crystals. There are links to a series of tools to visualize the three-dimensional structure, including Jmol (**Fig. 13.10**, arrow 2). **Table 13.5** lists some additional visualization software. Using Jmol does not require the installation of software (other than Java), and it is versatile (**Fig. 13.11**).

It is also possible to search within the PDB website using dozens of advanced search features (accessed via the top of the home page). This includes the use of BLAST or FASTA programs, allowing convenient access to PDB structures related to a query. Other advanced search features allow you to query based on properties of the molecule (e.g., its molecular weight), PubMed identifier, Medical Subject Heading (MeSH term; Chapter 2), deposit date, or experimental method.



Query Parameters:

Text Search for: myoglobin

Other search suggestions:

Molecule Name
- Myoglobin (362)

Find all

Structural Domains
- Myoglobin [SCOP] (254)
- Trematode hemoglobin/myoglobin [SCOP] (2)

Molecule of the Month
- Myoglobin

PDB Text
- myoglobin
- myoglobins

Find all

Ontology Terms
- SS : MYOGLOBIN [Genome... (17)
- D12.776...: Myoglobin [MeSH... (326)
- PC : MYOGLOBIN [Genome... (246)
- D12.776...: Myoglobin [MeSH... (311)
- EC : MYOGLOBIN [Genome... (62)
- HS : MYOGLOBIN [Genome... (1)

Query Refinements: ***Select an item or pie chart***

Organism
- Physeter catodon (247)
- Equus caballus (66)
- Sus scrofa (17)
- Homo sapiens (12)
- Thunnus atlanticus (8)
- Aplysia limacina (7)
- Chironomus thummi thummi (4)
- Other (16)

Taxonomy
- Eukaryota only (377)

Experimental Method
- X-ray (372)
- Neutron Diffraction (3)
- Other (1)
- Solution NMR (1)

X-ray Resolution
- less than 1.5 Å (96)
- 1.5 - 2.0 Å (203)
- 2.0 - 2.5 Å (69)
- 2.5 - 3.0 Å (4)
- more choices...

Release Date
- before 2000 (168)
- 2000 - 2005 (48)
- 2005 - 2010 (102)
- 2010 - 2015 (55)
- 2015 - today (4)
- more choices...

Polymer Type
- Protein (377)

Enzyme Classification
- 3: Hydrolases (2)
- 1: Oxidoreductases (1)

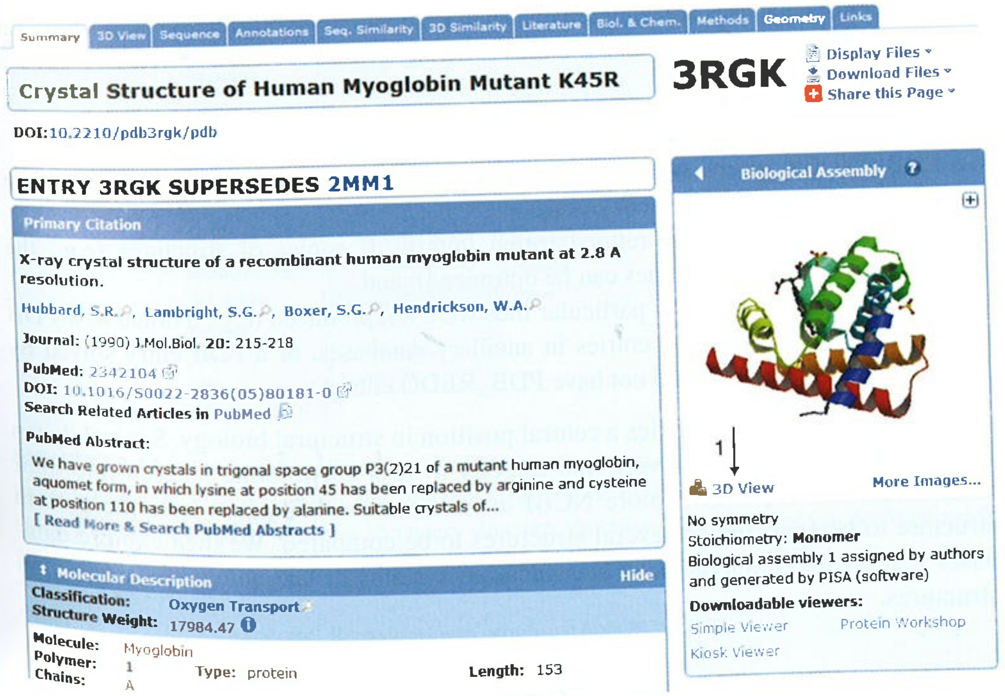
SCOP Classification
- Globin-like (core: 6 helices; ... (269)
- Cytochrome c (core: 3 helices; ... (1)

Protein Symmetry
- Asymmetric (353)
- Cyclic (9)
- more choices...

Protein Stoichiometry
- Monomer (353)
- Heteromer (5)
- Homomer (4)
- more choices...

FIGURE 13.9 Result of a PDB query for myoglobin. There are several hundred results organized into categories such as UniProt gene names, structural domains, and ontology terms. The search results further show how to explore myoglobin entries with the same categories shown in **Figure 13.7**.

Source: RCSB PDB (www.rcsb.org). Reproduced with permission from RCSB PDB.



Crystal Structure of Human Myoglobin Mutant K45R **3RGK**

DOI:10.2210/pdb3rgk/pdb

ENTRY 3RGK SUPERSEDES 2MM1

Primary Citation

X-ray crystal structure of a recombinant human myoglobin mutant at 2.8 A resolution.

Hubbard, S.R., Lambright, S.G., Boxer, S.G., Hendrickson, W.A.

Journal: (1990) J.Mol.Biol. **20:** 215-218

PubMed: 2342104
DOI: 10.1016/S0022-2836(05)80181-0
Search Related Articles in PubMed

PubMed Abstract:

We have grown crystals in trigonal space group P3(2)21 of a mutant human myoglobin, aquomet form, in which lysine at position 45 has been replaced by arginine and cysteine at position 110 has been replaced by alanine. Suitable crystals of...
[Read More & Search PubMed Abstracts]

Molecular Description

Classification: Oxygen Transport
Structure Weight: 17984.47

Molecule: Myoglobin
Polymer: 1 **Type:** protein **Length:** 153
Chains: A

No symmetry
Stoichiometry: **Monomer**
Biological assembly 1 assigned by authors and generated by PISA (software)

Downloadable viewers:
Simple Viewer Protein Workshop
Kiosk Viewer

FIGURE 13.10 Result of a search for a myoglobin structure, 3RGK. The summary information includes a description of the resolution (2.8 Å), the space group, unit cell dimensions, ligands, and external database annotation. Available links include a variety of visualization software (including Jmol, arrow 1).

Source: RCSB PDB (www.rcsb.org). Reproduced with permission from RCSB PDB.

TABLE 13.5 Interactive visualization tools for protein structures. The Protein Data Bank maintains a list of molecular graphics software links, accessible from the PDB home page via software tools/molecular viewers at http://www.pdb.org/pdb/static.do?p=software/software_links/molecular_graphics.html (WebLink 13.40).

Tool	Comment	URL
Cn3D	From NCBI	http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml
JMol	Open-source Java viewer for chemical structures in 3D	http://jmol.sourceforge.net/
Kiosk Viewer	Uses Java Web Start	http://pdb.org/
Mage	Reads Kinemages	http://kinemage.biochem.duke.edu
Protein Workshop Viewer	Uses Java Web Start	http://pdb.org/
RasMol	Molecular graphics visualization tool	http://www.rasmol.org/
RasTop	Molecular visualization software adapted from RasMol	http://www.geneinfinity.org/rastop/
Simple Viewer	Uses Java Web Start	http://pdb.org/
SwissPDB viewer	At ExPASy	http://spdbv.vital-it.ch
VMD	Visual Molecular Dynamics; University of Illinois	http://www.ks.uiuc.edu/Research/vmd/

The PDB in Europe database (PDBe) is at http://www.ebi.ac.uk/pdbe/ (WebLink 13.15). PDB Japan is at http://pdbj.org (WebLink 13.16).

PDB is maintained by members of the WorldWide PDB. These include the RCSB PDB, the Protein Data Bank in Europe (operated by the European Bioinformatics Institute), and PDB Japan.

A series of databases are complementary to PDB and hold information corresponding directly to PDB entries. These include the following (Joosten *et al.*, 2011):

- DSSP includes secondary structure data;
- PDBREPORT includes data on structure quality and errors;
- PDBFINDER offers summaries of PDB content (information includes Enzyme Commission numbers for enzymes);
- PDB_REDO includes re-refined (often improved) copies of structures (e.g., the orientation of peptide planes can be optimized); and
- WHY_NOT explains why particular files were not produced (e.g., a brand new PDB entry might not yet have entries in ancillary databases, or a PDB entry solved by NMR spectroscopy would not have PDB_REDO entries).

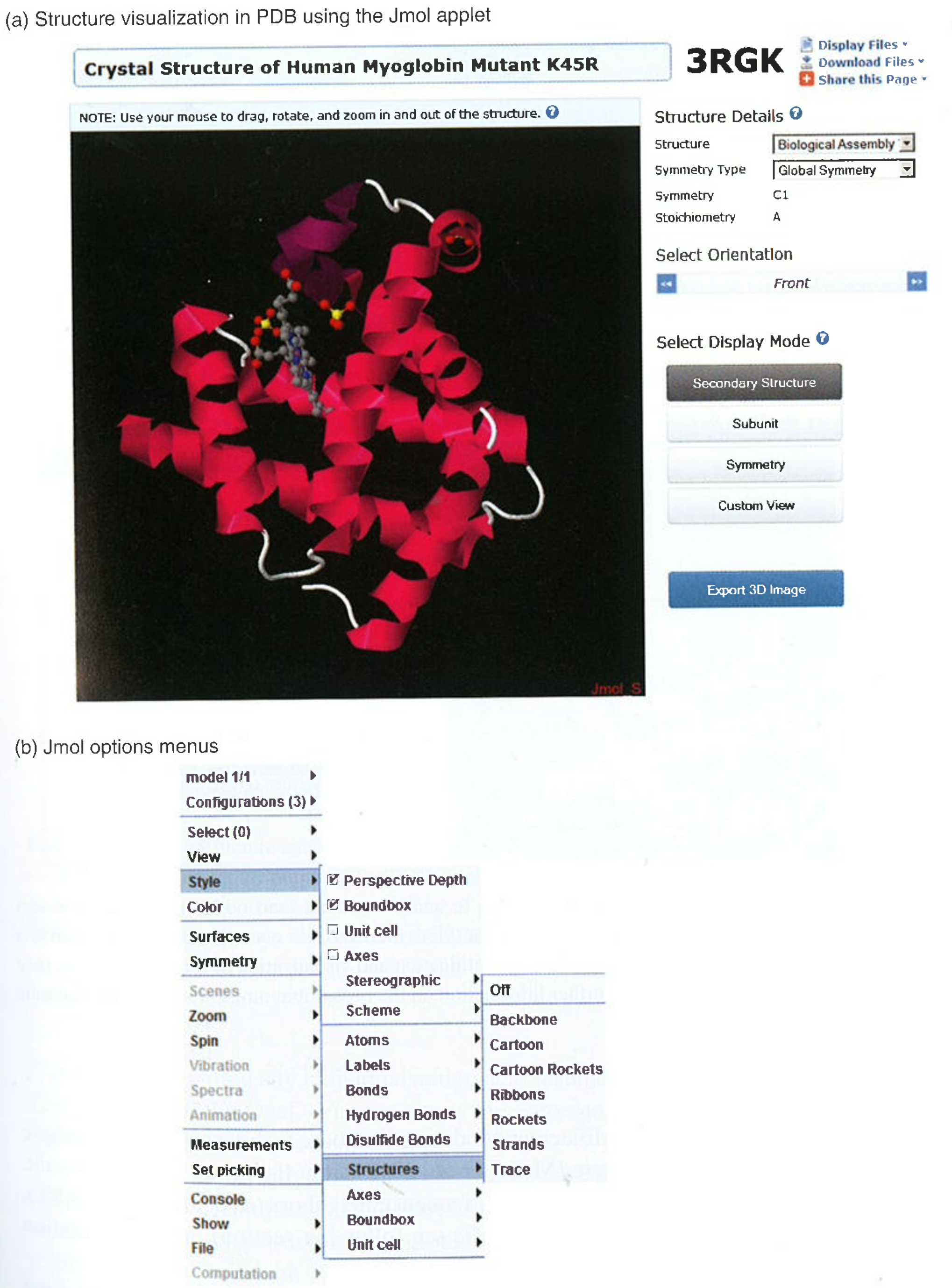
The PDB-related WHAT IF servers are available at http://swift.cmbi.ru.nl/servers/html/ (WebLink 13.17).

The PDB database occupies a central position in structural biology. Several dozen other databases and web servers link directly to it or incorporate its data into their local resources. We next explore NCBI and other sites that allow a single protein structure to be analyzed or several structures to be compared. We then explore databases that create comprehensive classification systems or taxonomies for all protein structures.

Accessing PDB Entries at NCBI Website

There are three main methods of finding a protein structure in the NCBI databases:

1. Text searches allow access to PDB structures. These searches can be performed on the structure page or through Entrez, and they can consist of keywords or PDB identifiers.


(a) Structure visualization in PDB using the Jmol applet

Crystal Structure of Human Myoglobin Mutant K45R

3RGK

NOTE: Use your mouse to drag, rotate, and zoom in and out of the structure.

Structure Details

Structure: Biological Assembly

Symmetry Type: Global Symmetry

Symmetry: C1

Stoichiometry: A

Select Orientation: Front

Select Display Mode

(b) Jmol options menus

FIGURE 13.11 Jmol applet software permits the visualization and analysis of macromolecular structures. (a) View of a human myoglobin structure. This can be manipulated (e.g., zoomed or rotated), colored according to criteria such as secondary structure, visualized (e.g., to show van der Waal radii), and analyzed (e.g., by measuring interatomic distances). (b) Right-clicking (on a PC) opens a menu of Jmol viewing options.

Source: RCSB PDB (www.rcsb.org). Reproduced with permission from RCSB PDB.

A keyword search of Entrez structures for hemoglobin yields a list of ~1300 proteins with four-character PDB identifiers. If you know a PDB identifier of interest, such as 3RGK for myoglobin, use it as a search term and to find an NCBI Structure entry with useful links, including to the Molecular Modeling Database (**Fig. 13.12**), the Cn3D viewer, the VAST comparison tool (see below), and the Conserved Domain

PROTEIN STRUCTURE PREDICTION

Structure prediction is a major goal of proteomics. There are three principal ways to predict the structure of a protein (**Fig. 13.20**; Cozzetto and Tramontano, 2008; Pavlopoulou and Michalopoulos, 2011). First, for a protein target that shares substantial

TABLE 13.8 Proteins having different numbers of domains assigned by SCOP, CATH, and DALI. Values are the number of domains assigned by each database. Data from CATH, SCOP, and DALI from the Protein Data Bank (http://www.pdb.org).

Name	PDB accession	SCOP	CATH	DALI
Glycogen phosphorylase	1gpb	1	2	3
Annexin V	1avh_A	1	4	4
Submaxillary renin	1smr_A	1	2	1
Fructose-1,6-bisphosphatase	5fbp_A	1	2	2

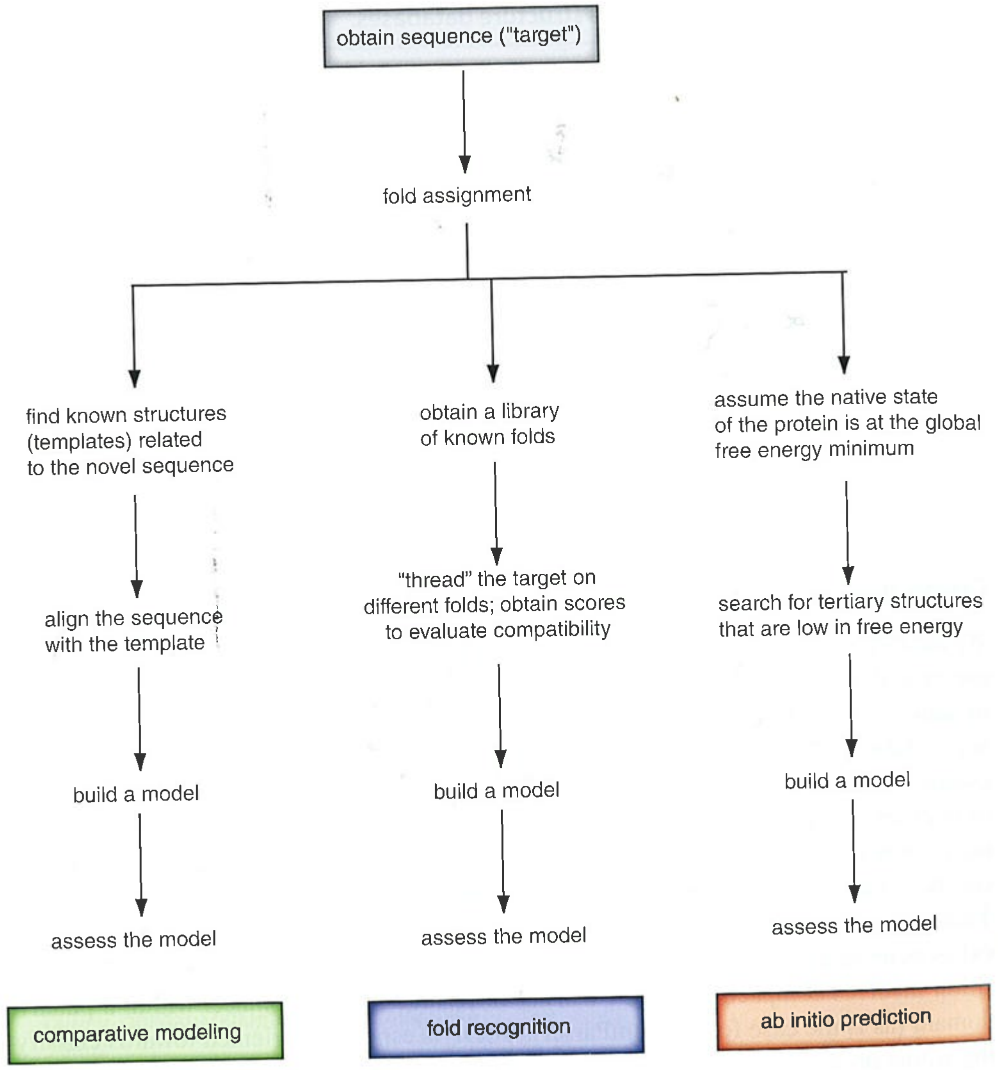

obtain sequence ("target")
fold assignment
find known structures (templates) related to the novel sequence
obtain a library of known folds
assume the native state of the protein is at the global free energy minimum
align the sequence with the template
"thread" the target on different folds; obtain scores to evaluate compatibility
search for tertiary structures that are low in free energy
build a model
build a model
build a model
assess the model
assess the model
assess the model
comparative modeling
fold recognition
ab initio prediction

FIGURE 13.20 Approaches to predicting protein structures (adapted from Baker and Sali, 2001). Comparative modeling is the most powerful approach when a target sequence has any indications of homology with a known structure. Threading is used to compare segments of a protein to a library of known folds. In the absence of homologous structures, *ab initio* prediction is used to model protein structure. Adapted from Baker and Sali (2001).

similarity to other proteins of known structure, homology modeling (also called comparative modeling) is applied. Second, for proteins that share folds but are not necessarily homologous, threading is a major approach. Proteins that are analogous (related by convergent evolution rather than homology) can be studied this way. Third, for targets lacking identifiable homology (or analogy) to proteins of known structure, *ab initio* approaches are applied.

Homology Modeling (Comparative Modeling)

While over 100,000 protein structures have been deposited in PDB, over half a million protein sequences have been deposited in the SwissProt database and 84 million more in TrEMBL (Chapter 12). For the vast majority of proteins, the assignment of structural models relies on computational biology approaches rather than experimental determination. As protein structures continue to be solved by X-ray crystallography and NMR spectroscopy, the most reliable method of modeling and evaluating new structures is by comparison to previously known structures (Baker and Sali, 2001; Jones, 2001). This is the method of comparative modeling of protein structure, also called homology modeling. This method is fundamental to the field of structural genomics.

Comparative modeling consists of four sequential steps (Marti-Renom *et al.*, 2000).

1. Template selection and fold assignment are performed. This can be accomplished by searching for homologous protein sequences and/or structures with tools such as BLAST and DELTA-BLAST. The target can be queried against databases described in this chapter, such as PDB, CATH, and SCOP. As part of this analysis, structurally conserved regions and structurally variable regions are identified. It is common for structurally variable regions to correspond to loops and turns, often at the exterior of a protein.
2. The target is aligned with the template. As for any alignment problem, it is especially difficult to determine accurate alignments for distantly related proteins. For 30% sequence identity between a target and a template protein, the two proteins are likely to have a similar structure if the length of the aligned region is sufficient (e.g., more than 60 amino acids). The use of multiple sequence alignments (Chapter 6) can be especially useful.
3. A model is built. A variety of approaches are employed, such as rigid-body assembly and segment matching.
4. The model must be evaluated (see below).

There are several principal types of errors that occur in comparative modeling (see Marti-Renom *et al.*, 2000):

- errors in side-chain packing;
- distortions within correctly aligned regions;
- errors in regions of a target that lack a match to a template;
- errors in sequence alignment; and
- use of incorrect templates.

The accuracy of protein structure prediction is closely related to the percent sequence identity between a target protein and its template (**Fig. 13.21**). When the two proteins share 50% amino acid identity or more, the quality of the model is usually excellent. For example, the root-mean-square deviation (RMSD) for the main-chain atoms tends to be 1 Å in such cases. Model accuracy declines when comparative models rely on 30–50% identity, and the error rate rises rapidly below 30% identity. *De novo* models are able to generate low-resolution structure models.

In Chapter 3, we discussed the importance of the length of the alignment in considering percent identity between two proteins.

Many web servers offer comparative modeling including quality assessment, such as SWISSMODEL at ExPASy, MODELLER, and the PredictProtein server (**Table 13.9**). After a model is generated it is necessary to assess its quality. The goal is to assess whether a particular structure is likely, based on a general knowledge of protein structure principles. Criteria for quality assessment may include whether the bond lengths and angles are appropriate; whether peptide bonds are planar; whether the carbon backbone conformations are allowable (e.g., following a Ramachandran plot); whether there are appropriate local environments for hydrophobic and hydrophilic residues; and solvent accessibility. Quality assessment programs include VERIFY3D, PROCHECK, and WHATIF at CMBI (Netherlands; **Table 13.9**).

Fold Recognition (Threading)

While there are currently >100,000 entries in the Protein Data Bank, there may be only 1000–2000 distinct folds in nature. Fold recognition, also called threading, is useful when

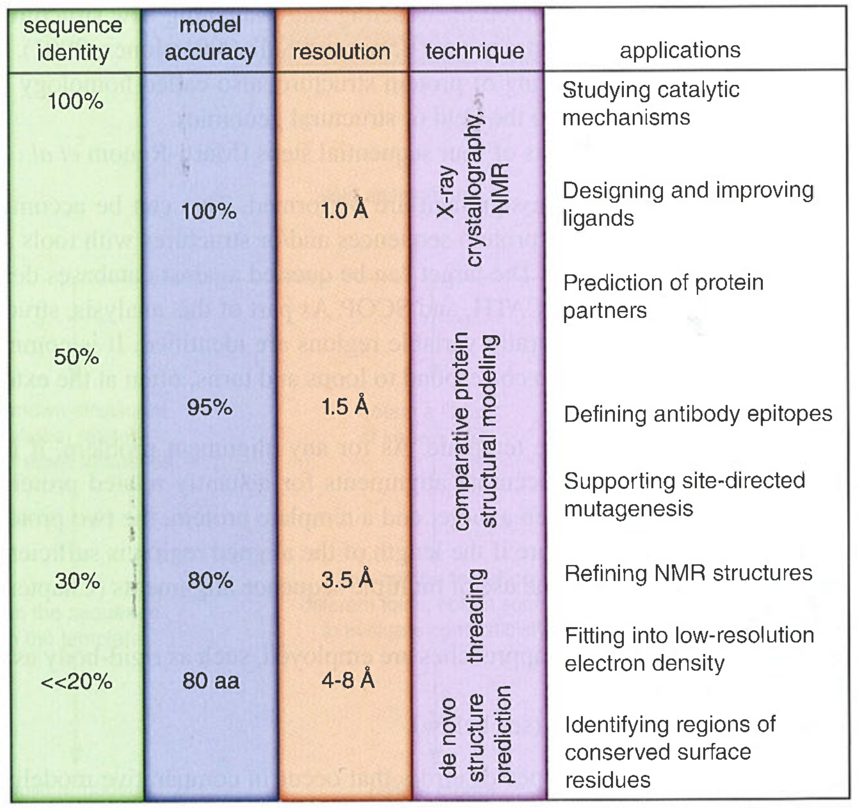

sequence identity
model accuracy
resolution
technique
applications
100%
100%
1.0 Å
X-ray crystallography, NMR
Studying catalytic mechanisms
Designing and improving ligands
Prediction of protein partners
50%
95%
1.5 Å
comparative protein structural modeling
Defining antibody epitopes
Supporting site-directed mutagenesis
30%
80%
3.5 Å
threading
Refining NMR structures
Fitting into low-resolution electron density
<<20%
80 aa
4-8 Å
de novo structure prediction
Identifying regions of conserved surface residues

FIGURE 13.21 Protein structure prediction and accuracy as a function of the relatedness of a novel structure to a known template. Modified from Baker and Sali (2001). aa: amino acids. Used with permission.

a target sequence of interest lacks identifiable sequence matches and yet may have folds in common with proteins of known structure. The target might assume a fold that occurs in a characterized protein because of convergent evolution, or because the two proteins are homologous but extremely distantly related. An input sequence is parsed into subfragments and "threaded" onto a library of known folds. Scoring functions allow an assessment of how compatible the sequence is with known structures. A variety of web servers provide automatic threading.

Websites for fold recognition include 3D-PSSM (⊕ http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html, WebLink 13.26) and its successor PHYRE (⊕ http://www.sbg.bio.ic.ac.uk/~phyre/, WebLink 13.27), FUGUE (⊕ http://tardis.nibio.go.jp/fugue/, WebLink 13.28).

TABLE 13.9 Websites for structure prediction by comparative modeling, and for quality assessment.

Website	Comment	URL
3D-JIGSAW	Laboratory of Paul Bates	http://bmm.cancerresearchuk.org/~3djigsaw/
Geno3D	POLE	http://pbil.ibcp.fr/htm/index.php
MODELLER	From Andrej Sali's group	http://www.salilab.org/modeller/
PredictProtein	Laboratory of Burkhard Rost	http://www.predictprotein.org/
SWISS-MODEL	ExPASy	http://swissmodel.expasy.org/
PROCHECK	Quality assessment	http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
VERIFY3D	Quality assessment	http://nihserver.mbi.ucla.edu/Verify_3D/
WHATIF	Quality assessment	http://swift.cmbi.ru.nl/whatif/

Ab Initio Prediction (Template-Free Modeling)

In the absence of detectable homologs, protein structure may be assessed by *ab initio* (or *de novo*) structure prediction. "*Ab initio*," meaning "from the beginning," is the most difficult approach to structure prediction (Osguthorpe, 2000; Simons *et al.*, 2001; Jothi, 2012). It is based on two assumptions: (1) all the information about the structure of a protein is contained in its amino acid sequence; and (2) a globular protein folds into the structure with the lowest free energy. Finding such a structure requires both a scoring function and a search strategy. While the resolution of *ab initio* methods is generally low, this approach is useful to provide structural models.

The Rosetta method is one of the most successful *ab initio* strategies (Simons *et al.*, 2001; Rohl *et al.*, 2004; Adams *et al.*, 2013). The target protein is evaluated in fragments of nine amino acids. These fragments are compared to known structures in PDB. From this analysis, structures can be inferred for the entire peptide chain. Typically, models generated with Rosetta have accuracies of 3–6 Å root mean square deviation from known structures for aligned segments of 60 or more amino acids (Rohl *et al.*, 2004). Bonneau *et al.* (2002) used the Rosetta method to model the structure of all Pfam-A sequence families (Chapter 6) for which three-dimensional structures are unknown. By calibrating their method on known structures, they estimated that for 60% of the proteins studied (80 of 131), one of the top five ranked models successfully predicted the structure within 6.0 Å RMSD.

The Robetta server from David Baker's lab, located at ⊕ http://robetta.bakerlab.org/ (WebLink 13.29), applies the Rosetta method (Kim *et al.*, 2004).

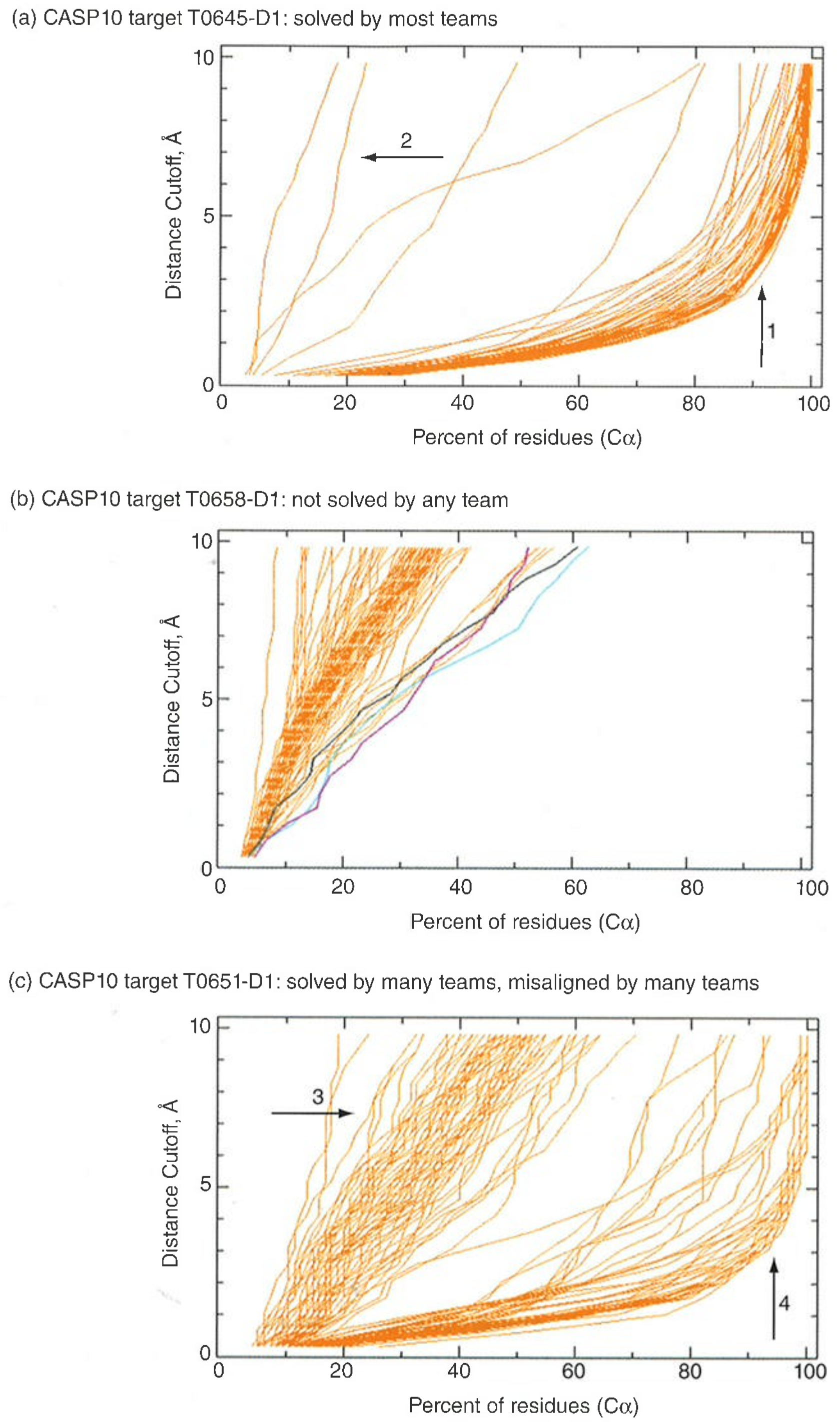
A Competition to Assess Progress in Structure Prediction

How well can the community predict the structures of proteins, particularly those with novel folds? The state-of-art protein prediction is assessed by the structural genomics community at Critical Assessment of Techniques for Protein Structure Prediction (CASP; Kryshtafovych *et al.*, 2014a). This structure prediction experiment (or competition) has occurred every two years since the first competition in 1996. While 35 groups participated in CASP1, over 200 prediction servers and manual groups joined CASP10 in 2012, coming from dozens of countries. Approximately 100 experimentally determined targets were evaluated, and tens of thousands of models were deposited with a team of assessors. The structures of the targets were known but withheld from publication so that the community could perform predictions in a blind fashion (Kryshtafovych *et al.*, 2014b). Predictors consisted of either scientists who performed modeling of each target, or automatic servers that produced predictions in a short time period (48 hours) without human intervention. By 2014, CASP11 generated nearly 60,000 predictions.

The CASP targets include those that require:(1) comparative modeling with close evolutionary relationships (e.g., those identifiable by BLAST); (2) comparative modeling to distantly related targets (e.g., those requiring PSI- or DELTA-BLAST or hidden Markov models to detect relationships of a template to proteins having known structure); (3) threading; (4) template-free modeling; (5) refinement of protein models; or (6) assessment of intramolecular residue-residue contacts (Monastyrskyy *et al.*, 2014a; Nugent *et al.*, 2014; Taylor *et al.*, 2014). Kryshtafovych *et al.* (2014a) reviewed the overall progress of CASP. In its first 10 years (CASP1 through CASP5) there was substantial improvement in model quality. In the second decade, improvements through CASP10 have been more modest, with overall model accuracy being comparable to that in CASP5. There are several reasons for this. Each target undergoes comparative modeling using an existing experimental structure as a guide that may be superimposed on the target. There has been progress in the ability to identify best templates (with 10% improvement in the past decade), partly through the development of methods involving multiple templates. The increased availability of known structures has however (surprisingly) made it more difficult to identify best templates in some cases. Major challenges include: the need for improved alignments; the need for models of close evolutionary relationships to approach the accuracy obtained by experimental structure determination; the need to better refine models of remote evolutionary relationships; and the need to discriminate among the best template-free models (Moult, 2005; Tai *et al.*, 2005; Moult *et al.*, 2007).

The CASP website provides detailed results of the competition. One criterion for the accuracy of a prediction is the GDT_TS metric which compares the difference in position of the main chain Cα atoms in a model relative to the position in the experimentally determined structure. **Figure 13.22** shows examples of an easy protein target from CASP10 that was solved by most groups (**Fig. 13.22a**) and a difficult target that no group solved (**Fig. 13.22b**). **Figure 13.22c** depicts an example of a target that was aligned either very well or very poorly by many groups; those with poor results misaligned the sequence of the target, highlighting the difficulty of correctly aligning a target sequence onto available template structures for template-based models.

The Protein Structure Prediction Center organizes CASP information (⊕ http://predictioncenter.org/, WebLink 13.30) including results from each CASP competition.


(a) CASP10 target T0645-D1: solved by most teams

(b) CASP10 target T0658-D1: not solved by any team

(c) CASP10 target T0651-D1: solved by many teams, misaligned by many teams

FIGURE 13.22 Examples of results from the CASP10 competition. Each plot (called a GDT plot or "Hubbard plot") shows the percent of CA or Cα residues (i.e., the percent of the modeled structure; x axis) versus the distance cutoff in Ångstroms (from 0 Å to 10 Å; y axis). Each line represents a summary of a single prediction of that protein's structure; multiple lines are from the many groups that submitted predictions. (a) Example of a protein target (T0645) whose structure was modeled extremely well by many teams participating in the CASP competition. Note that a very high percentage of the residues in the predictions that could be overlaid on the correct structure (*x* axis values approaching 100%) with only a very small RMSD (distance cutoff, *y* axis) as indicated by arrow 1. A small number of predictions were wrong (arrow 2) because they correctly matched the true structure over only a small percent of residues even at large distance cutoffs. (b) Example of a protein target (T0658) whose true structure was not predicted by any group in the CASP competition. Several groups' predictions (colored lines from the Seok, Jiang, and Zhang groups) were better than all others. (c) Example of a target (T0651) that was predicted incorrectly by many teams (arrow 3) but correctly by others (arrow 4). Such a broad discrepancy in prediction accuracy is often attributable to incorrect sequence alignments in homology modelling.

Source: CASP10 results at ⊕ http://www.predictioncenter.org. Reproduced with permission from University of California, Davis.