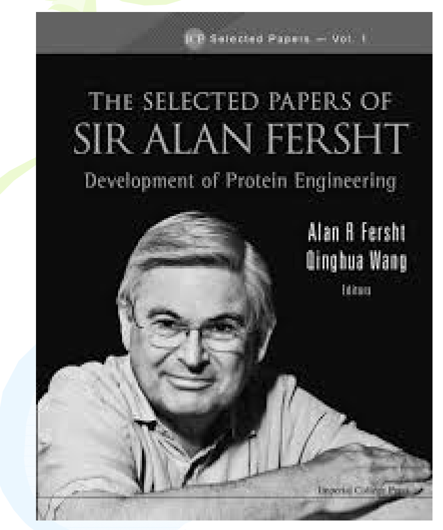
Fundamentals of Structural Biology

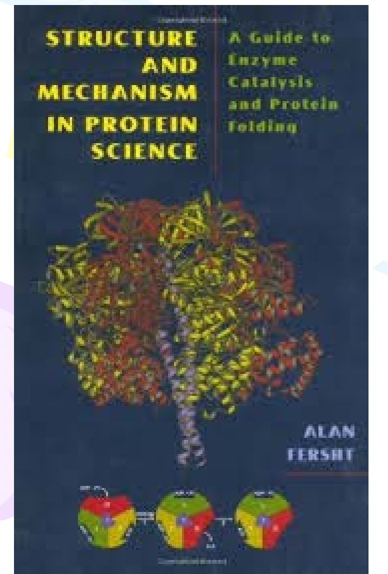
Jung-Hsin Lin (林榮信)
Research Center for Applied Sciences &
Institute of Biomedical Sciences, Academia Sinica
School of Pharmacy, National Taiwan University
College of Engineering, Chang Gung University

http://www.rcas.sinica.edu.tw/faculty/jhlin.html

2018 Frontiers in Computational Drug Design, Academia Sinica, March 16-20, 2018

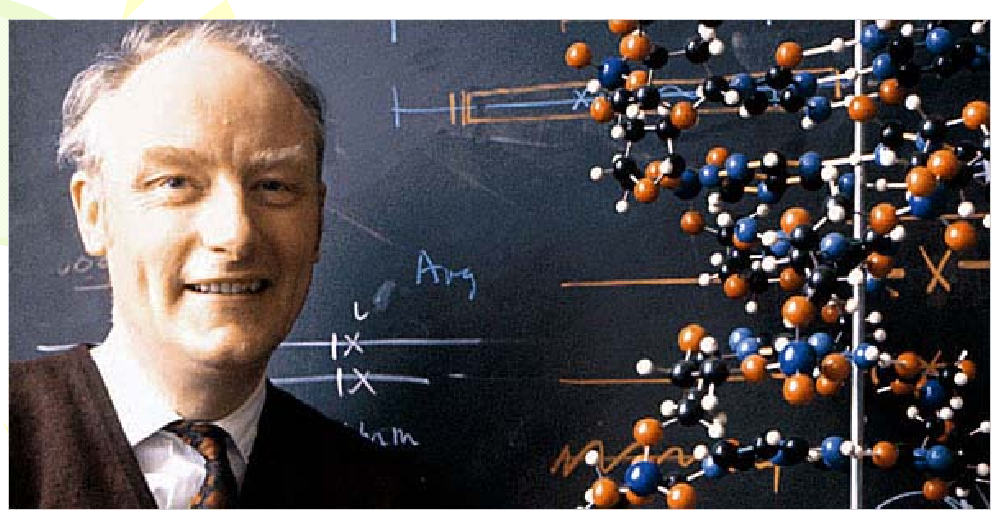
The ultimate goal of protein science …





…… Enzymes and receptors are the usual targets of drugs, either to restore function or to destroy infectious agents or cancers. **The ultimate goal of protein science is to be able to predict the structure and activity of a protein *de novo* and how it will bind to ligands.** When this is achieved, we will be able to design and synthesize novel catalysts, materials, and drugs that will eliminate disease and minimize ill health.

Alan Fersht, *Structure and Mechanism in Protein Science*, Ch. 1.



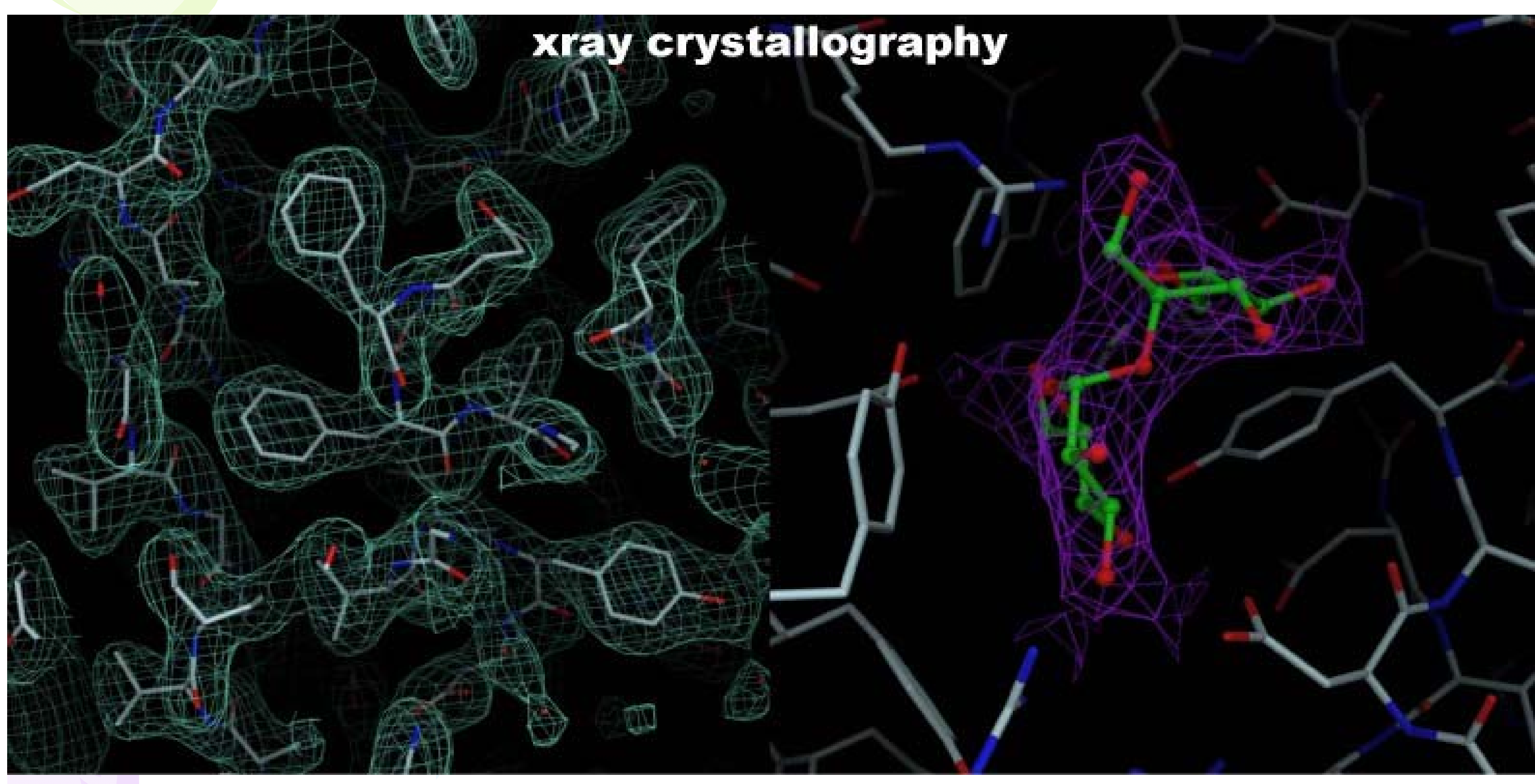
Francis Harry Compton Crick, (8 June 1916 – 28 July 2004) was an English molecular biologist, biophysicist, and neuroscientist, and most noted for being a co-discoverer of the structure of the DNA molecule in 1953 together with James D. Watson. He, Watson, and Maurice Wilkins were jointly awarded the 1962 Nobel Prize for Physiology or Medicine "for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material".

If you want to understand function, study structure.

What Mad Pursuit: A Personal View of Scientific Discovery (1988), 150

Where do protein structures come from?

- X-ray crystallography
- Nuclear magnetic resonance (NMR)
- (cryo)Electron microscopy (cryoEM)
- Atomic force microscopy (AFM)
- Homology modeling
- Threading and fold recognition
- *Ab initio* prediction

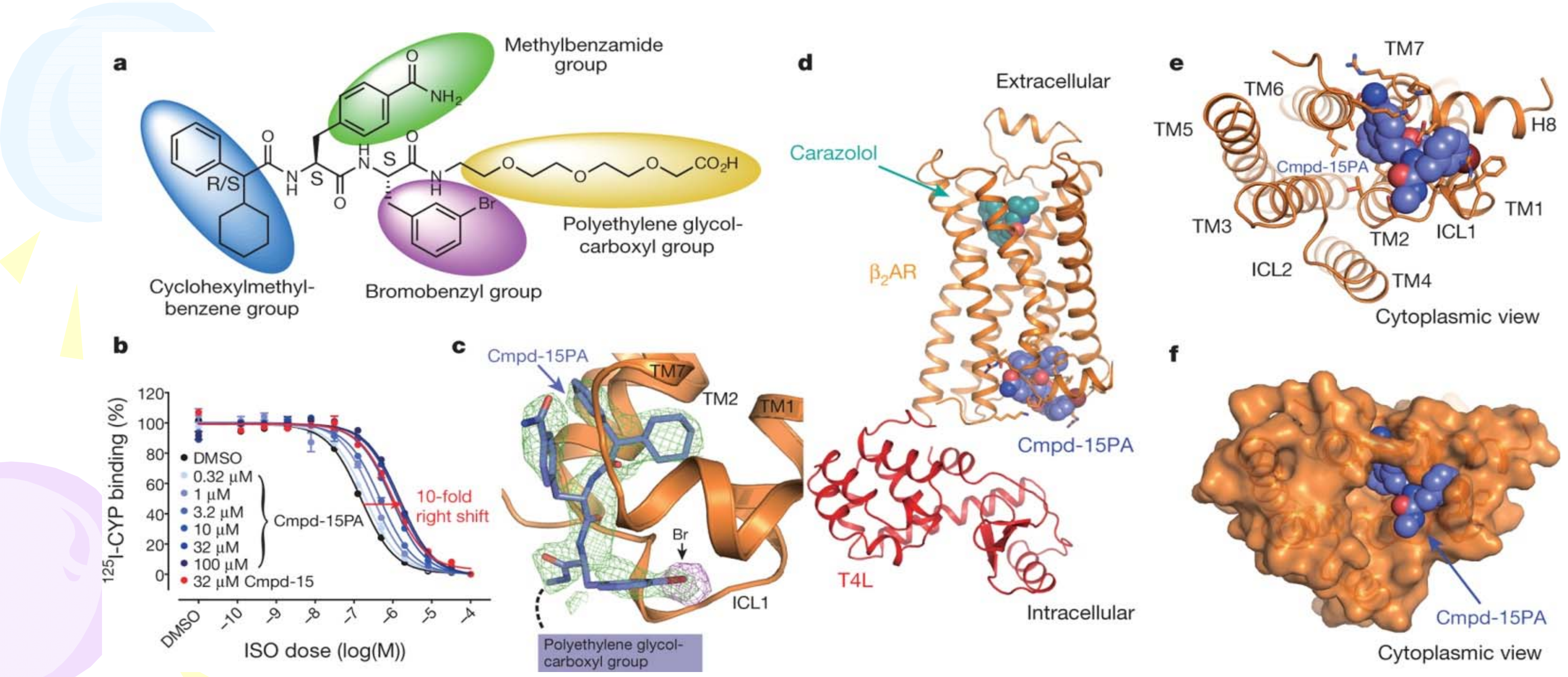

xray crystallography

LETTER

doi:10.1038/nature23652

Mechanism of intracellular allosteric β_2AR antagonist revealed by X-ray crystal structure

Xiangyu Liu[1]*, Seungkirl Ahn[2]*, Alem W. Kahsai[2], Kai-Cheng Meng[3], Naomi R. Latorraca[4,5], Biswaranjan Pani[2], A. J. Venkatakrishnan[4,5,6], Ali Masoudi[2], William I. Weis[7], Ron O. Dror[4,5], Xin Chen[3], Robert J. Lefkowitz[2,8,9] & Brian K. Kobilka[1,6]



6

structural NMR

nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-018-0022-z

Solution structure and elevator mechanism of the membrane electron transporter CcdA

Yunpeng Zhou[1*] and John H. Bushweller[1,2*]

Membrane oxidoreductase CcdA plays a central role in supplying reducing equivalents from the bacterial cytoplasm to the envelope. It transports electrons across the membrane using a single pair of cysteines by a mechanism that has not yet been elucidated. Here we report an NMR structure of the *Thermus thermophilus* CcdA (TtCcdA) in an oxidized and outward-facing state. CcdA consists of two inverted structural repeats of three transmembrane helices (2×3-TM). We computationally modeled and experimentally validated an inward-facing state, which suggests that CcdA uses an elevator-type movement to shuttle the reactive cysteines across the membrane. CcdA belongs to the LysE superfamily, and thus its structure may be relevant to other LysE clan transporters. Structure comparisons of CcdA, semiSWEET, Pnu, and major facilitator superfamily (MFS) transporters provide insights into membrane transporter architecture and mechanism.

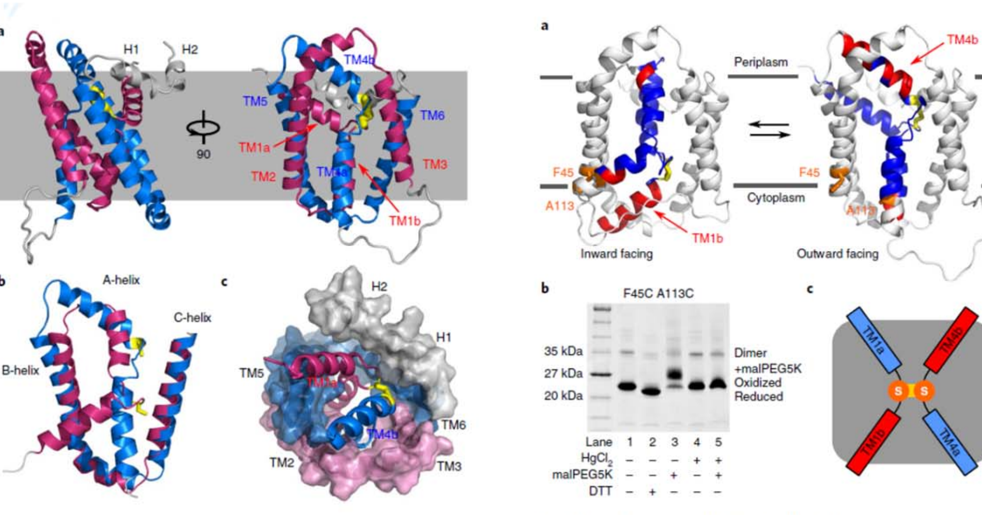


Fig. 2 | Structure of oxidized TtCcdA. a, Structure of oxidized TtCcdA

Fig. 3 | Inward-facing model of oxidized TtCcdA. a, Cartoon

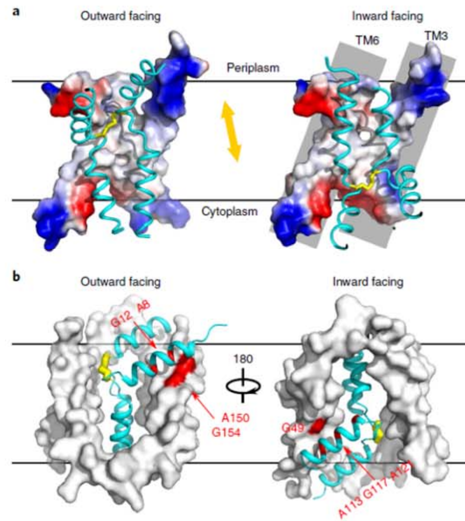


Fig. 4 | Elevator-type movement of the transport helices. a, Cartoon

8

electron microscopy

LETTER

doi:10.1038/nature25502

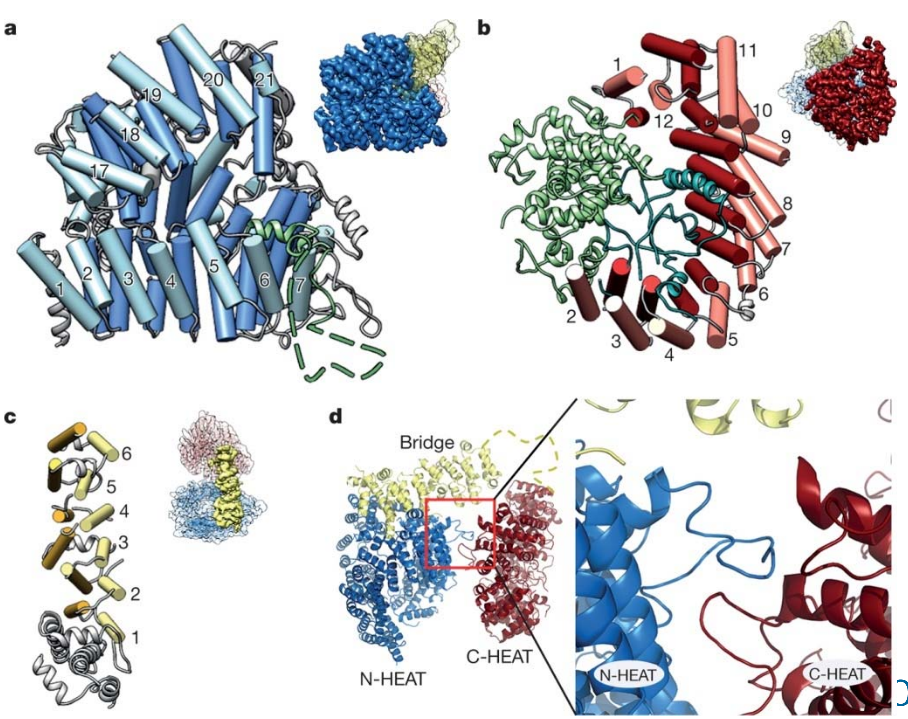
The cryo-electron microscopy structure of huntingtin

Qiang Guo[1]*, Bin Huang[2]*, Jingdong Cheng[3], Manuel Seefelder[2], Tatjana Engler[2], Günter Pfeifer[1], Patrick Oeckl[4], Markus Otto[4], Franziska Moser[5], Melanie Maurer[5], Alexander Pautsch[5], Wolfgang Baumeister[1], Rubén Fernández-Busnadiego[1] & Stefan Kochanek[2]

Architecture of the HTT–HAP40 complex

a b c d e
Bridge
C-HEAT
HAP40
N-HEAT
90°
180°
30 Å
0 500 1,000 1,500 2,000 2,500 3,000
HTT
HAP40
Unresolved sequence Polyglutamine N-HEAT Bridge C-HEAT HAP40

Structure of HTT domains



atomic force microscopy

Rhodopsin dimers in native disc membranes

Neat rows of paired photon receptors are caught on camera in their natural state.

In vertebrate retinal photoreceptors, the rod outer-segment disc membranes contain densely packed rhodopsin molecules for optimal light absorption and subsequent amplification by the visual signalling cascade[1], but how these photon receptors are organized with respect to each other is not known. Here we use infrared-laser atomic-force microscopy to reveal the native arrangement of rhodopsin, which forms paracrystalline arrays of dimers in mouse disc membranes. The visualization of these closely packed rhodopsin dimers in native membranes gives experimental support to earlier inferences about their supramolecular structure[2,3] and provides insight into how light signalling is controlled.

When activated by a single photon, rhodopsin induces dissociation of the subunits of transducin molecules, which are heterotrimeric G proteins that amplify the light signal. The structure of rhodopsin, which has been solved[4], indicates that it is a prototypical member of subfamily A of G-protein-coupled receptors (GPCRs), which represents about 90% of all GPCRs.

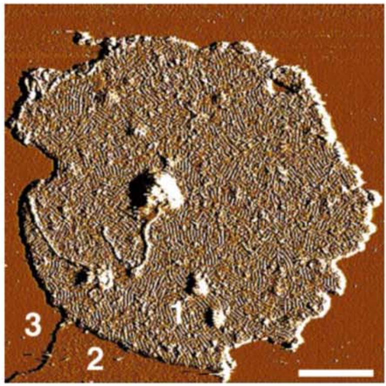
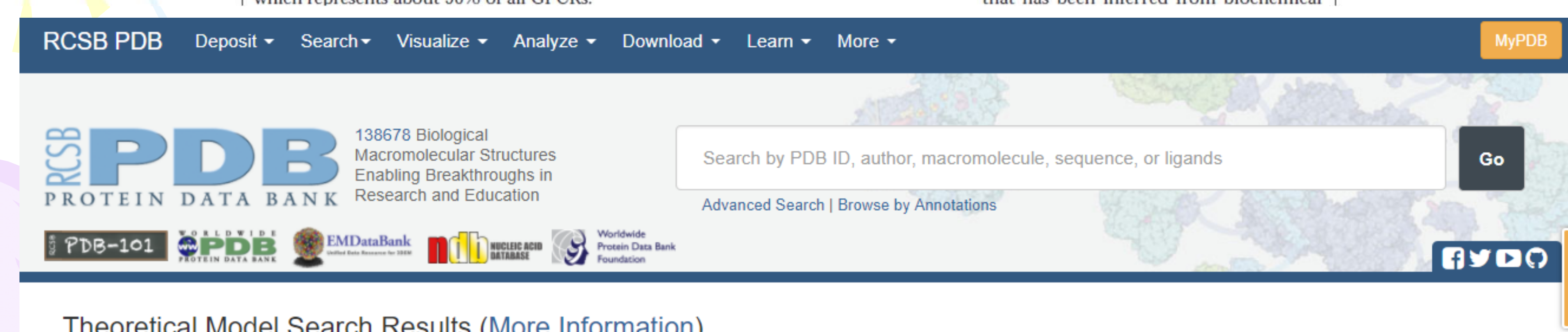


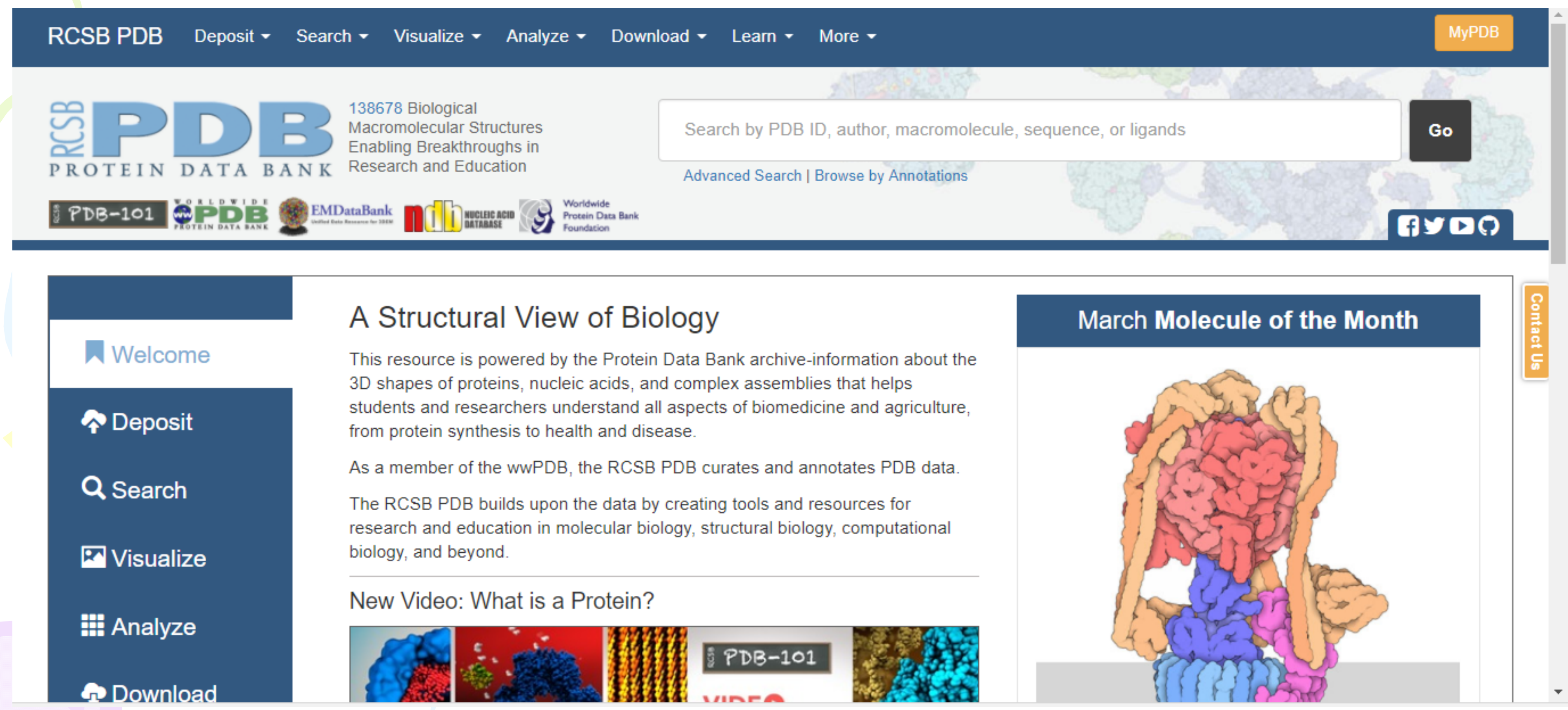
Figure 1 Deflection image of a native eye-disc membrane adsorbed on mica, visualized by atomic-force microscopy (Nanoscope Multimode, Digital Instruments). Three different surface types are evident: 1, the cytoplasmic side of the disc membrane; 2, lipid; and 3, mica. To avoid the formation of opsin, the chromophore-depleted form of rhodopsin[6], membrane samples were never exposed to light. After adsorption of osmotically shocked disc membranes onto mica, their topography was measured in buffer solution (20 mM Tris–HCl (pH 7.8), 150 mM KCl and 25 mM $MgCl_2$). Scale bar, 200 nm.

powder-diffraction pattern (Fig. 2a, inset, and b, arrows): the innermost arc peaks at $(8.4\ nm)^{-1}$, which results from regularly packed double rows of protrusions; the next ring, at $(4.2\ nm)^{-1}$, reflects the second-order of the double-row repeat, and the axial repeat of the paired rhodopsins that form these rows yields a third ring at $(3.8\ nm)^{-1}$. From real space measurements, we found the distance between the protrusions within each pair to be 3.8 ± 0.2 nm ($n=40$) and the angle between the lattice vectors to be $85 \pm 2°$ ($n=8$).

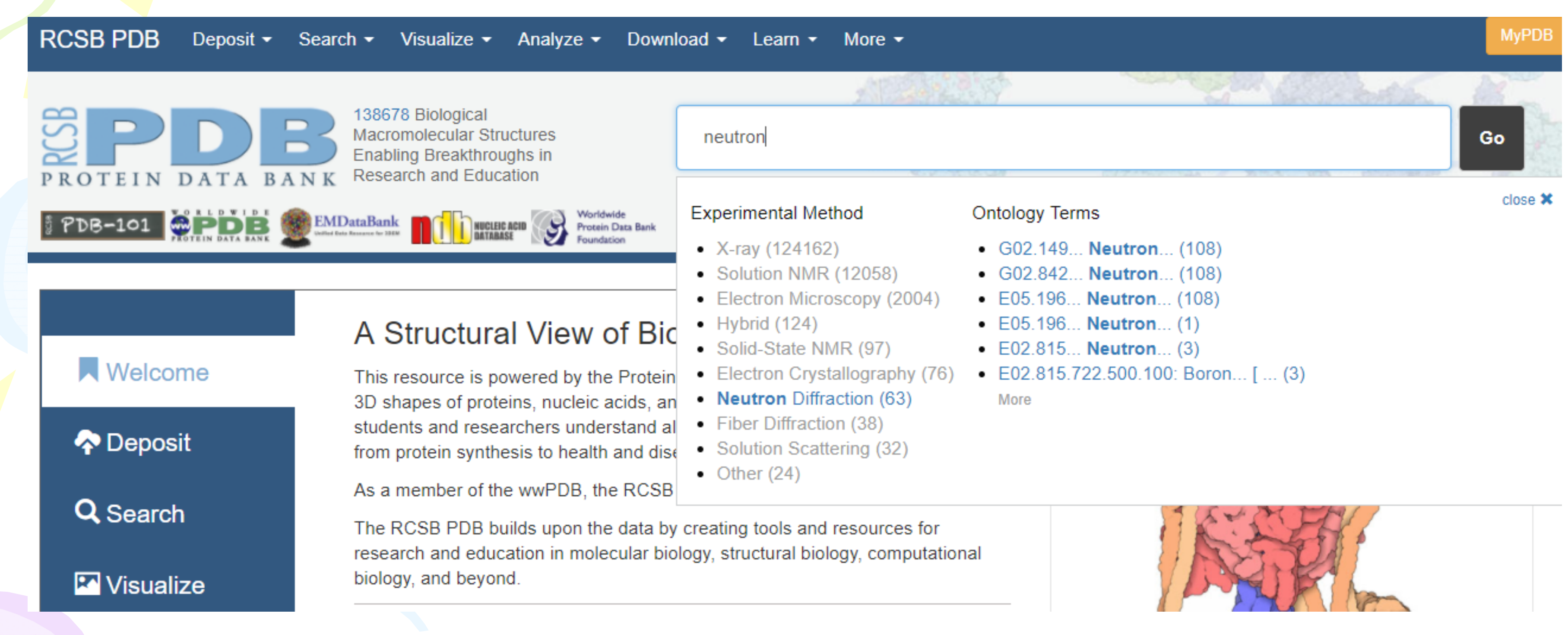
From these dimensions, the highest possible packing density is 62,900 rhodopsin monomers per μm^2. At higher magnification (Fig. 2c), it can be seen that almost all rhodopsin molecules are present in rows of dimers, with a few monomers and some single rhodopsin pairs that have broken away from the rows. This direct demonstration of distinct, densely packed rows of dimers at high resolution is consistent with the proposed dimeric form of native GPCRs that has been inferred from biochemical



Where are protein structures stored?

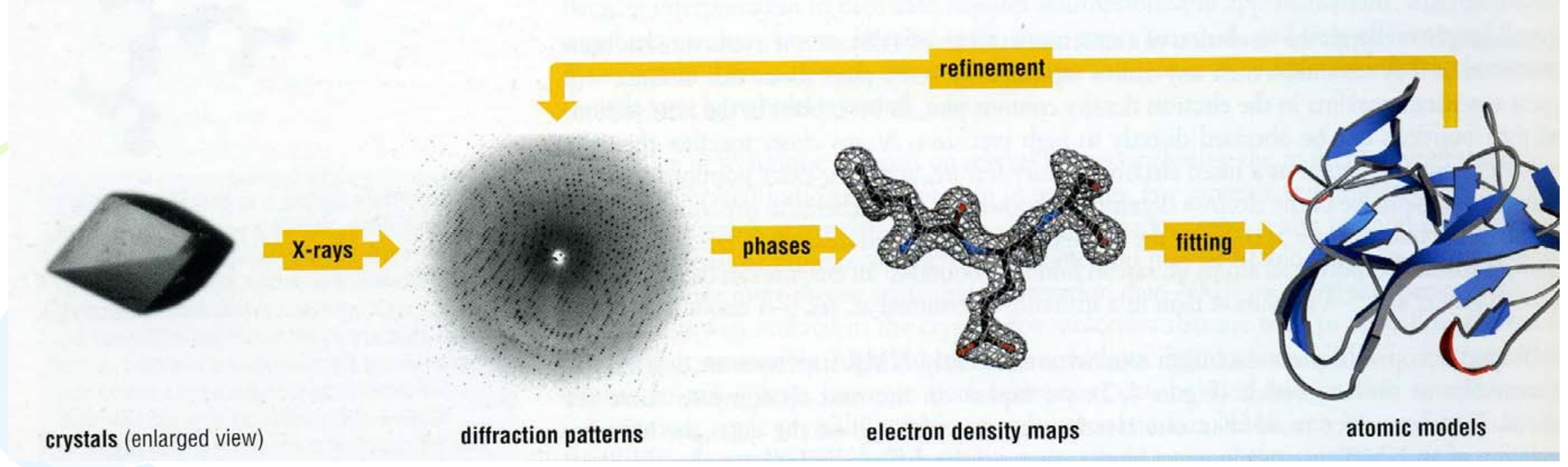


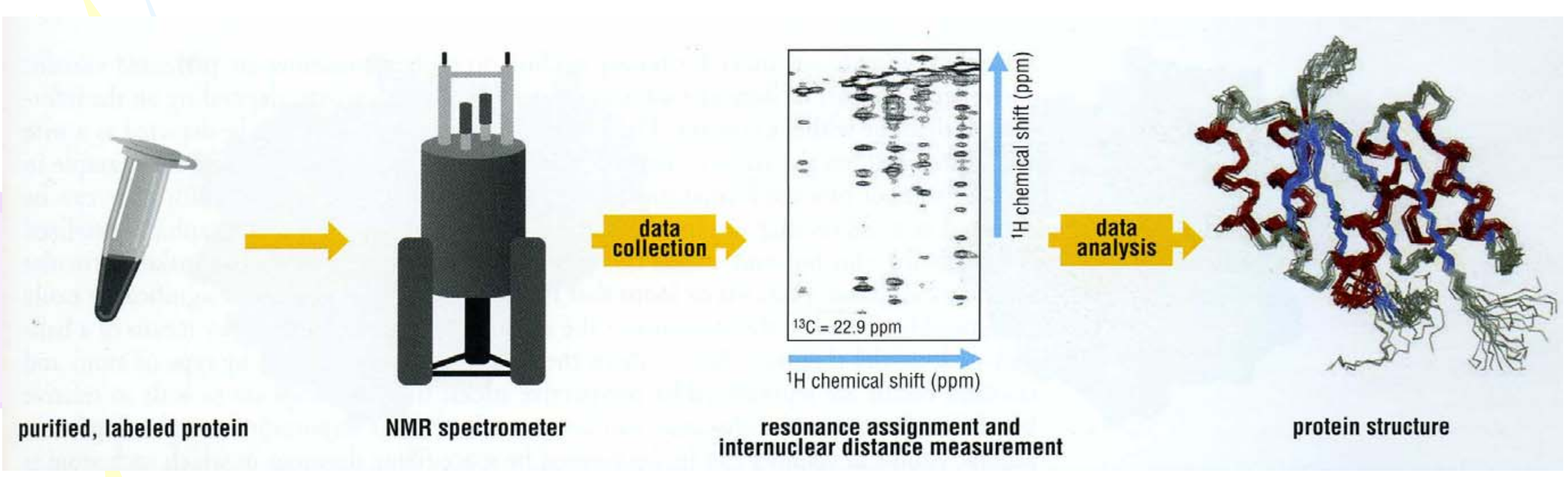
http://www.rcsb.org



Retrieved March 14, 2018.

X-ray crystallography and NMR are two major methods for protein structure determination





The application of NMR is limited by the size of the biomolecules

protein. Unfortunately, the method has historically been limited to proteins smaller than 25 kDa. The two major obstacles to a universal application of NMR are sensitivity losses and increased spectral complexity, both of which are pronounced in large proteins. The loss in sensitivity originates from NMR relaxation, the process by which the NMR spin systems return to their equilibrium state. In particular, transverse relaxation leads to concomitant losses in signal intensities and signal line-broadening. In a major breakthrough, so-called TROSY (transverse relaxation optimized spectroscopy) techniques have provided a means to combat these adverse effects when combined with isotope labeling [1,2]. Many NMR experiments were adapted to TROSY methods and structural studies of larger systems emerged. However, these studies have most often focused on multimeric proteins, as reflected by depositions in the PDB (Figure 1), highlighting the additional challenges to which large monomeric proteins are subject. As a consequence, this review

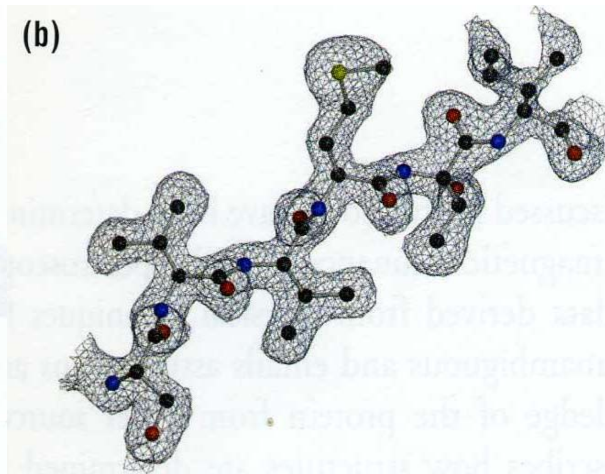
Available online at www.sciencedirect.com

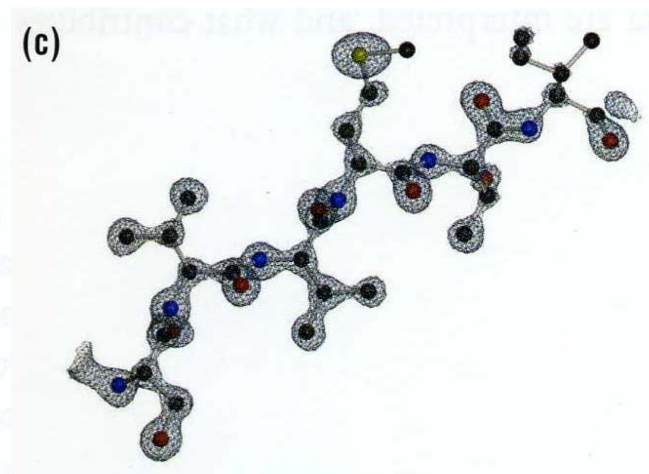
SciVerse ScienceDirect

NMR methods for structural studies of large monomeric and multimeric proteins

Dominique P Frueh, Andrew C Goodrich, Subrata H Mishra and Scott R Nichols

Good resolution versus bad resolution





At 3-Å resolution of fold of the polypeptide chain can be seen and approximate positions of side chains can be determined. Interatomic distances can only be measured to about 0.5 Å .

At 2-Å resolution side chains are well delineated and peptide carbonyls of backbone are discernible, allowing the chain to be oriented with precision. Interatomic distances can be measured to a precision of about 0.2 Å. Approximately three times as many data are required as were used at 3-Å resolution

At 1-Å resolution atoms are visible and resolved. Interatomic distances can be measured to a precision of a few hundredth of an Ångstrom. Almost 30 times more data are required for this resolution as were used at 3-Å resolution. In favorable cases, the position of hydrogens can be inferred at this resolution.

Positions of atoms with 0.001 Å resolution?

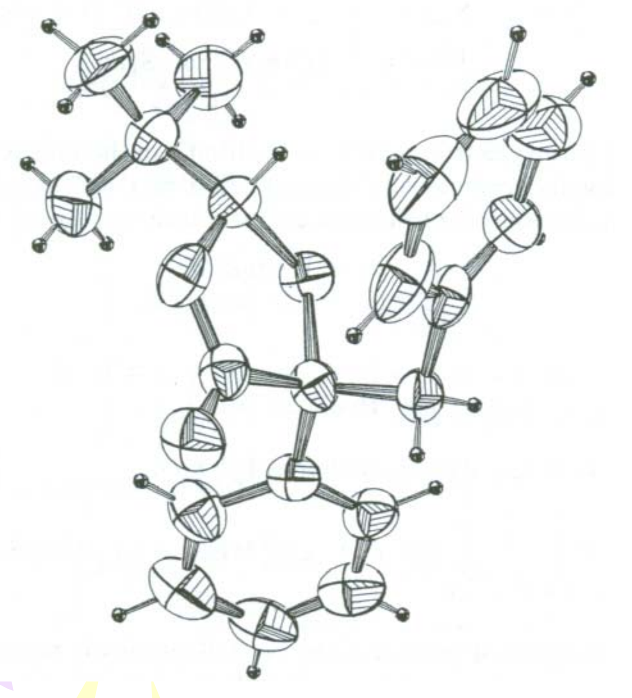
```
ATOM   1237  CA  PRO 1 187     61.209  11.195 102.027  1.00 43.25         C
ATOM   1238  C   PRO 1 187     61.917  12.397 101.399  1.00 47.87         C
ATOM   1239  O   PRO 1 187     61.528  13.539 101.726  1.00 49.71         O
ATOM   1240  CB  PRO 1 187     62.113  10.495 103.048  1.00 32.38         C
ATOM   1241  CG  PRO 1 187     61.665   9.101 103.004  1.00 37.10         C
ATOM   1242  CD  PRO 1 187     61.501   8.854 101.511  1.00 38.61         C
ATOM   1243  OXT PRO 1 187     62.872  12.183 100.608  1.00 57.89         O
TER    1244      PRO 1 187
ATOM   1245  N   HIS 2  12     32.324  25.971  37.670  1.00 46.84         N
ATOM   1246  CA  HIS 2  12     32.672  26.018  36.221  1.00 51.31         C
ATOM   1247  C   HIS 2  12     33.793  25.007  35.979  1.00 52.55         C
ATOM   1248  O   HIS 2  12     34.859  25.103  36.586  1.00 54.39         O
ATOM   1249  CB  HIS 2  12     33.143  27.436  35.840  1.00 58.43         C
ATOM   1250  CG  HIS 2  12     33.401  27.633  34.368  1.00 68.86         C
ATOM   1251  ND1 HIS 2  12     34.217  28.636  33.883  1.00 68.65         N
ATOM   1252  CD2 HIS 2  12     32.923  26.983  33.279  1.00 70.09         C
```

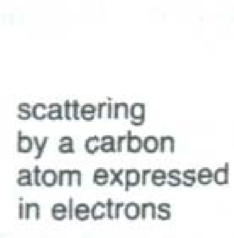
Temperature Factor

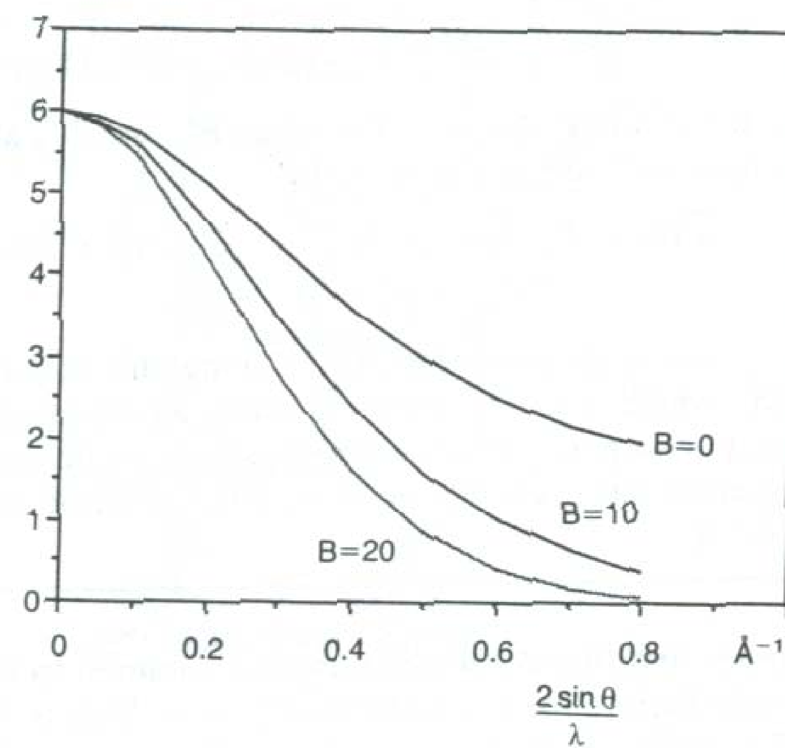
$$T_{iso} = \exp[-\frac{B}{4}\left(\frac{1}{d}\right)^2]$$

$$B = 8\pi^2 \times \overline{u^2}$$

u : mean square displacement



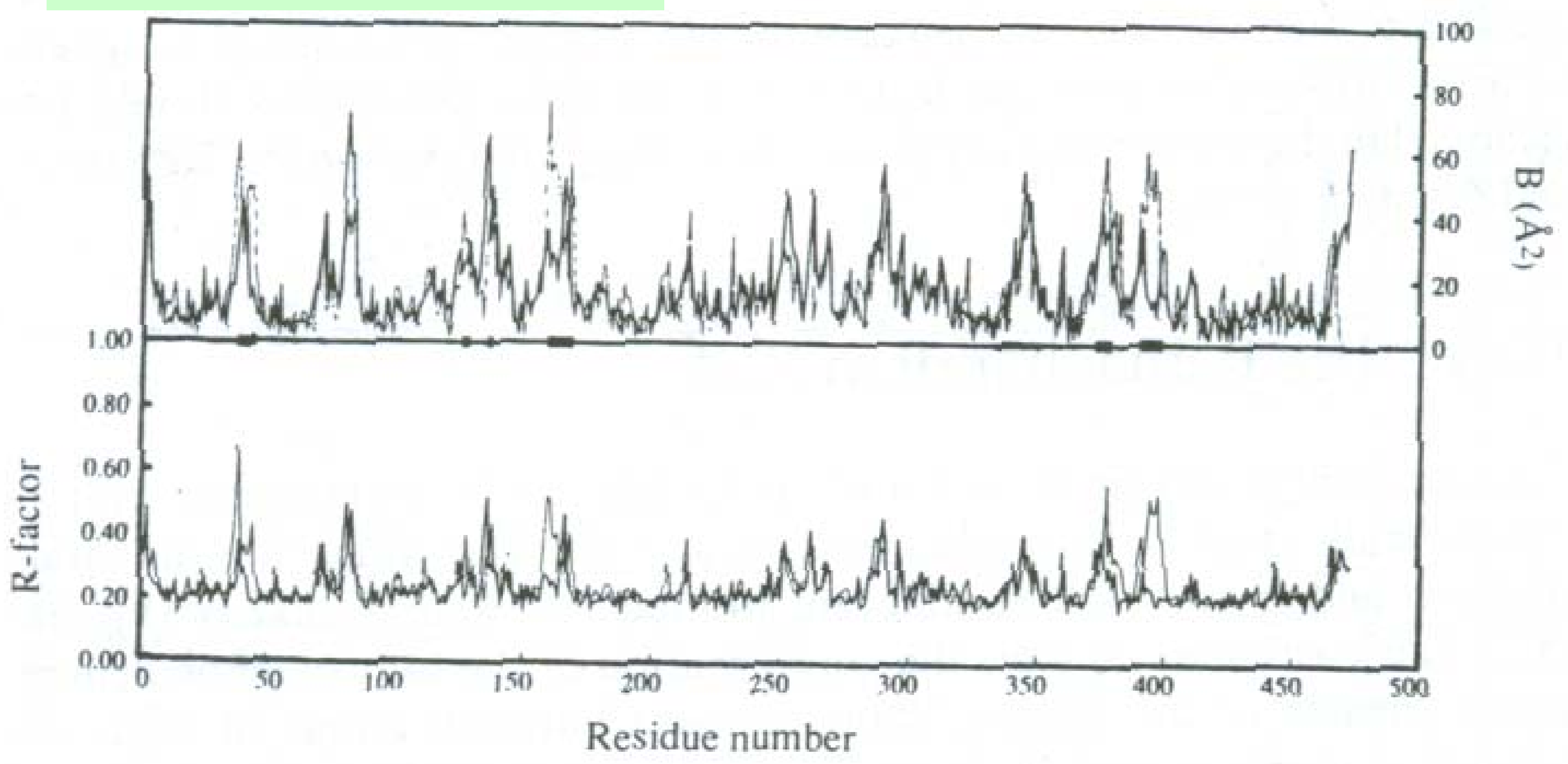




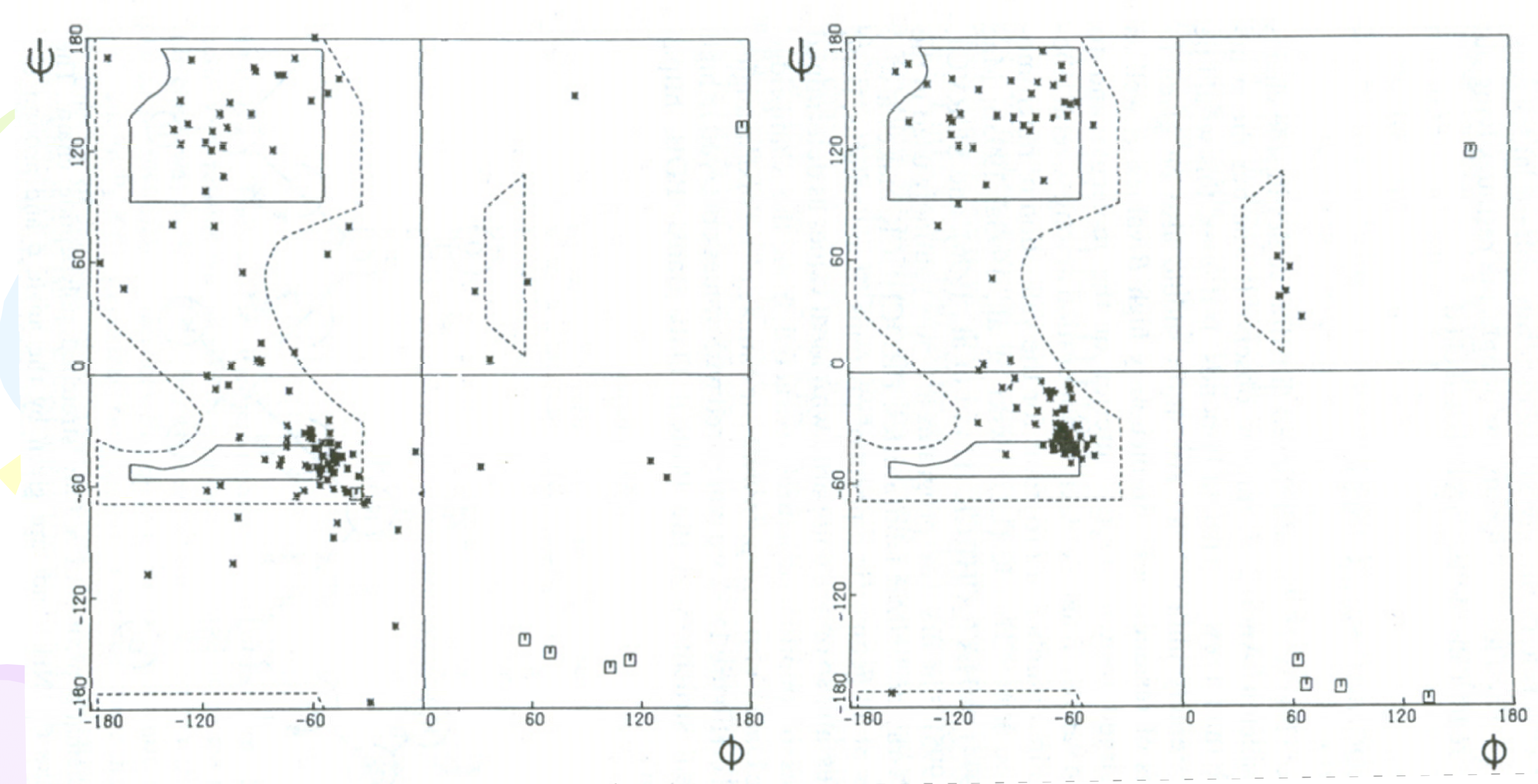
Checking the Gross Errors of the Structure Model

$$R = \frac{\sum_{hkl} \left\| F_{obs} \right| - k \left| F_{calc} \right\|}{\left| F_{obs} \right|}$$

- For structures refined to high resolution, for instance, 1.6 Å, the R-factor should not be much higher than 0.16.



Checking the Ramachandran Plot



- Phospholipase A2 at 1.7 Å resolution. Open squares: glycine; asterisks: all other residues. Left: before refinement; right: after refinement.

Protein Backbone Structures and Ramachandran Plot

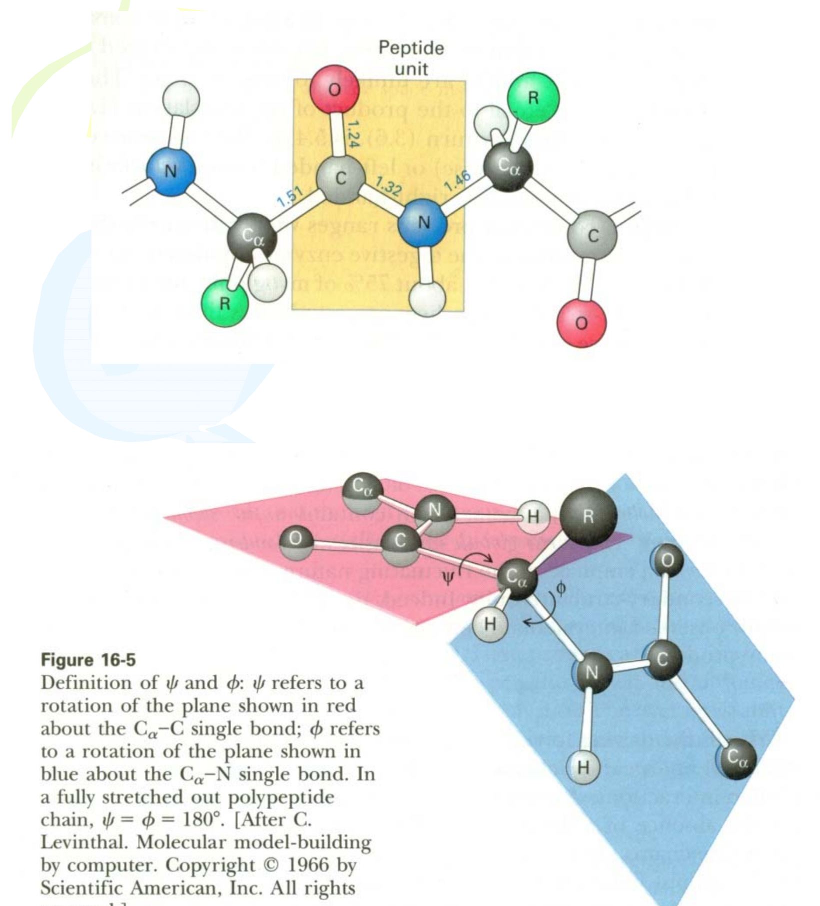


Figure 16-5
Definition of ψ and ϕ: ψ refers to a rotation of the plane shown in red about the C_α–C single bond; ϕ refers to a rotation of the plane shown in blue about the C_α–N single bond. In a fully stretched out polypeptide chain, $\psi = \phi = 180°$. [After C. Levinthal. Molecular model-building by computer. Copyright © 1966 by Scientific American, Inc. All rights reserved.]

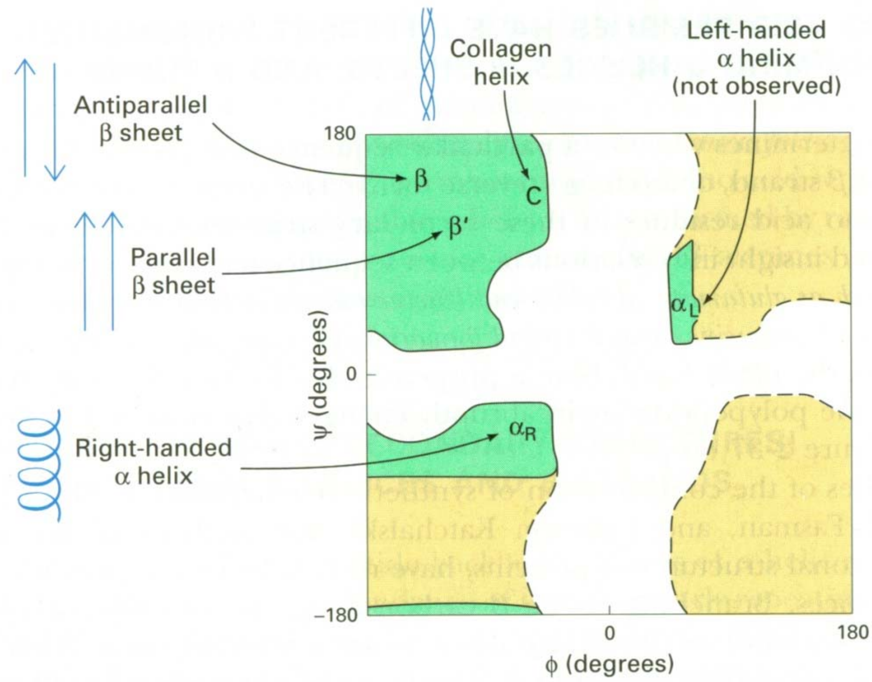


Figure 16-6
Ramachandran plot showing allowed values of ϕ and ψ for L-alanine residues (green regions). Additional conformations are accessible to glycine (yellow regions) because it has a very small side chain.

PyMOL

PyMOL | www.pymol.or ×

www.pymol.org

Home | Products | Buy | News | Support | Contact

PyMOL

*A molecular visualization system on an **open source** foundation, maintained and distributed by **Schrödinger***

DOWNLOAD

PyMOL runs on

VIEW	RENDER	ANIMATE	EXPORT	PRESENT	DEVELOP
3D Molecular Structures	Figures Artistically	Molecules Dynamically	Geometry Data	3D Data in PowerPoint	Customized Visualizations

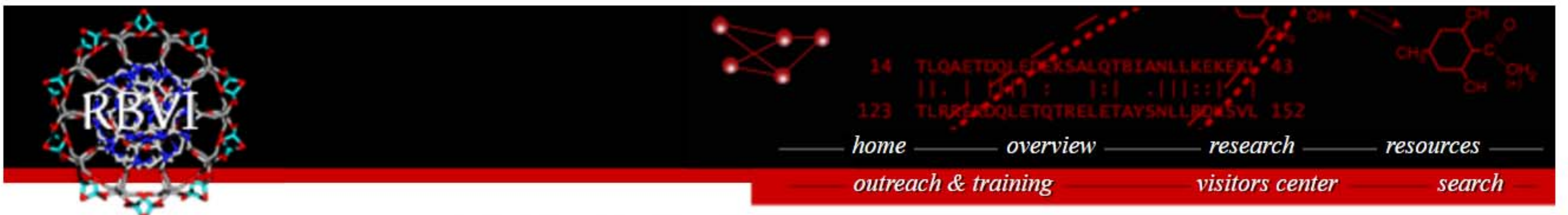
News

March 9, 2017: PyMOL v1.8.6.0 released. Review the release notes, and download the binaries.

More news...

PyMOL is a **user-sponsored** molecular visualization system on an **open-source** foundation. Please support development of this open, effective, and affordable software by purchasing an incentive copy, which is pre-built and comes with maintenance and support

CHIMERA



UCSF CHIMERA

an Extensible Molecular Modeling System

Quick Links

Documentation
- Getting Started
- User's Guide
- Command Index
- Tutorials and Videos
- Guide to Volume Data
- Release Notes

Download
- What's New in Daily Builds
- Map of Download Locations

UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. Chimera includes complete documentation and several tutorials, and can be downloaded free of charge for academic, government, nonprofit, and personal use. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics (RBVI), funded by the National Institutes of Health (NIGMS P41-GM103311).

Chimera Search

Go

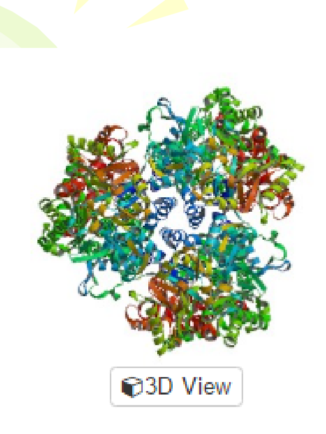
Google™ Search

News

December 2, 2016

Chimera production release 1.11.2 is now available. This version has been updated to work with changes in NCBI Blast and to avoid crashes on Mac

http://www.cgl.ucsf.edu/chimera/



1IWG

Contact Us

Crystal structure of Bacterial Multidrug Efflux transporter AcrB

Murakami, S., Nakashima, R., Yamashita, E., Yamaguchi, A.

(2002) Nature **419** 587-593

Released: 10/23/2002
Method: X-ray Diffraction
Resolution: 3.5 Å
Residue Count: 1053

Macromolecule:
AcrB (protein)
Unique Ligands: --

RCSB PDB Deposit ▾ Search ▾ Visualize ▾ Analyze ▾ Download ▾ Learn ▾ More ▾ MyPDB Login

Transmembrane View

transmembrane regions OPM

Display Files ▾ Download Files ▾

FASTA Sequence
PDB Format
PDB Format (gz)
PDBx/mmCIF Format
PDBx/mmCIF Format (gz)
PDBML/XML Format (gz)
Biological Assembly (PDB format - gz)

1IWG

Crystal structure of Bacterial Multidrug Efflux tran

DOI: 10.2210/pdb1iwg/pdb

Classification: MEMBRANE PROTEIN
Deposited: 2002-05-15 **Released**: 2002-10-23
Deposition author(s): Murakami, S., Nakashima, R., Yamasl
Organism: Escherichia coli
Expression System: Escherichia coli K12

Structural Biology Knowledgebase: 1IWG (>17 annotations)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 3.5 Å
R-Value Free: 0.355
R-Value Work: 0.290

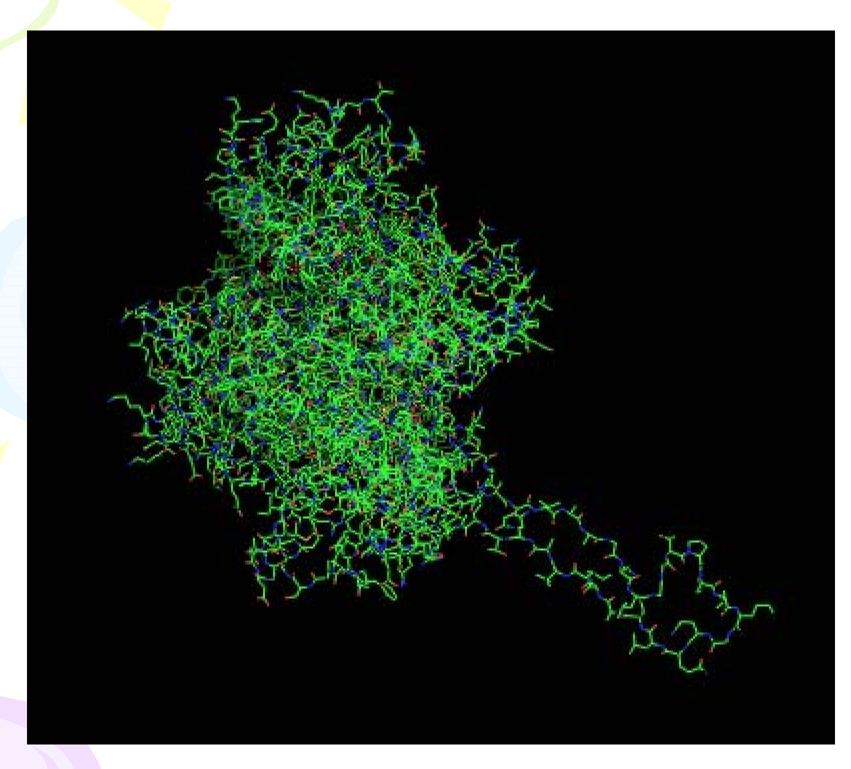
wwPDB Validation

Metric	Percentile Ranks	Value
Clashscore		52
Ramachandran outliers		7.5%
Sidechain outliers		4.5%

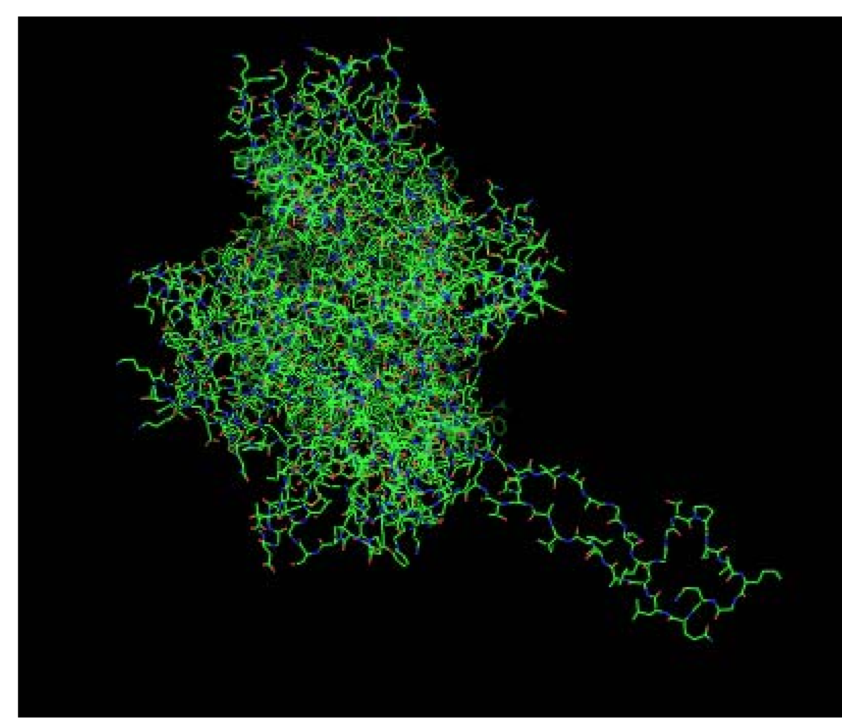
Worse Better

Percentile relative to all X-ray structures
Percentile relative to X-ray structures of similar resolution

PyMOL



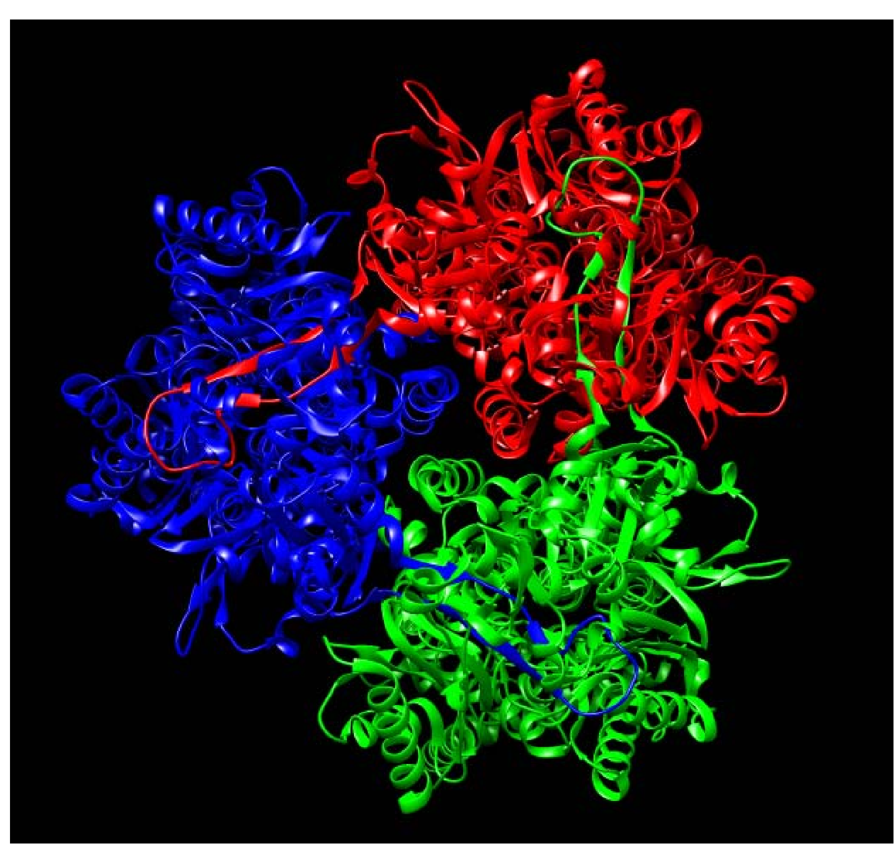
1iwq.pdb



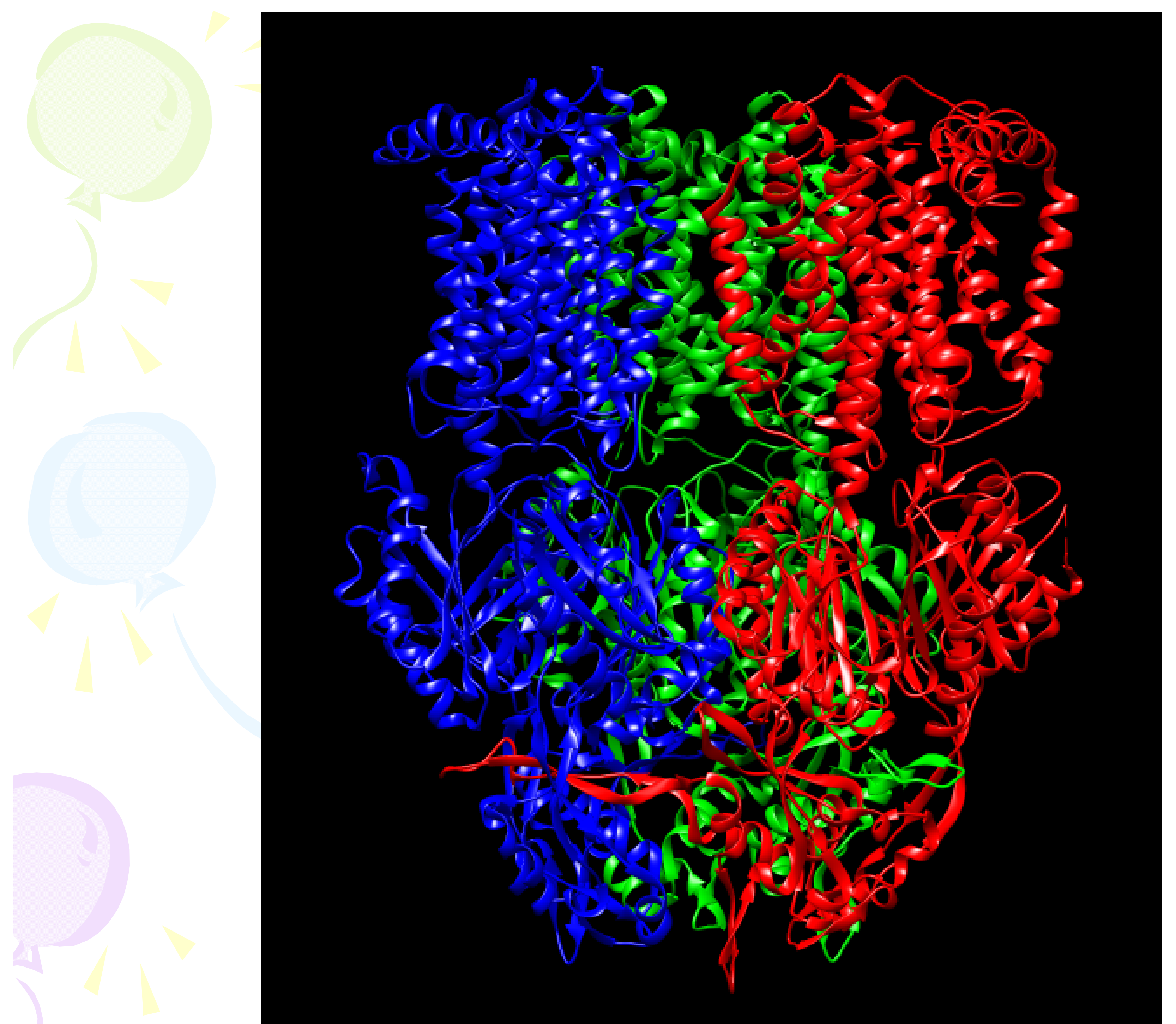
1iwq.pdb1

CHIMERA

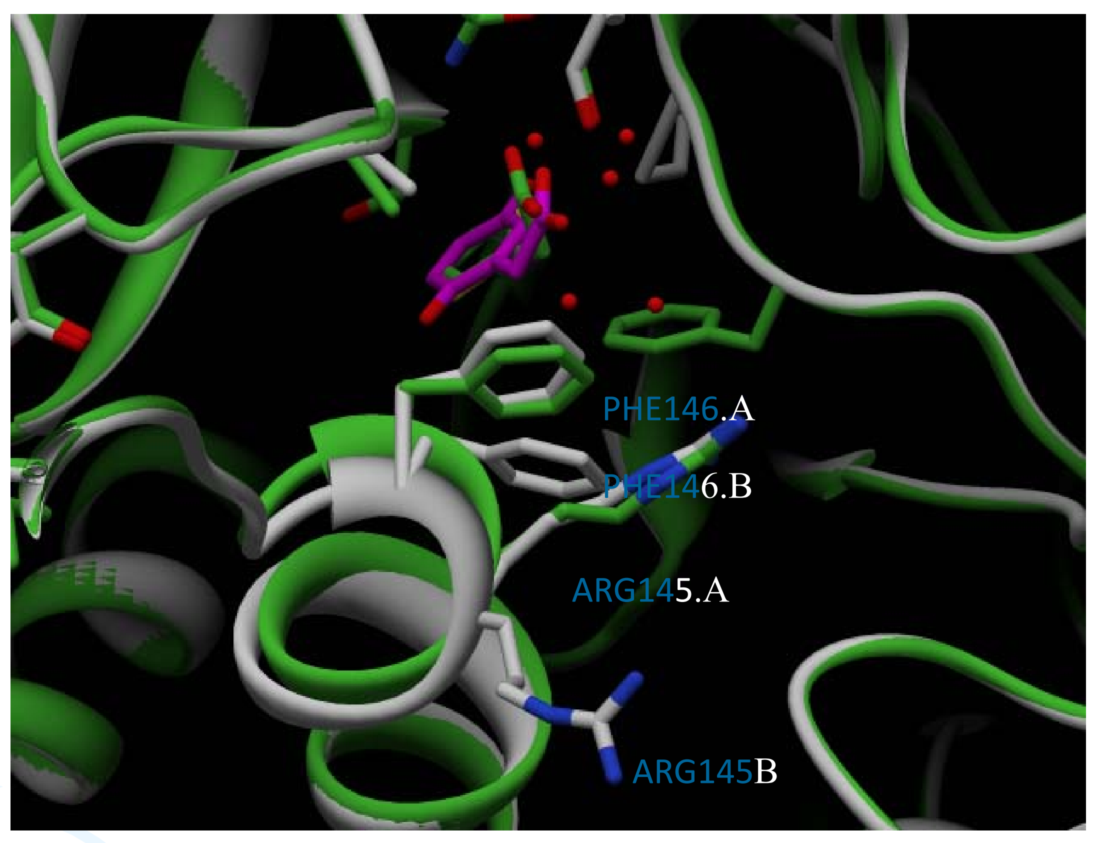
1iwq.pdb



1iwq.pdb1

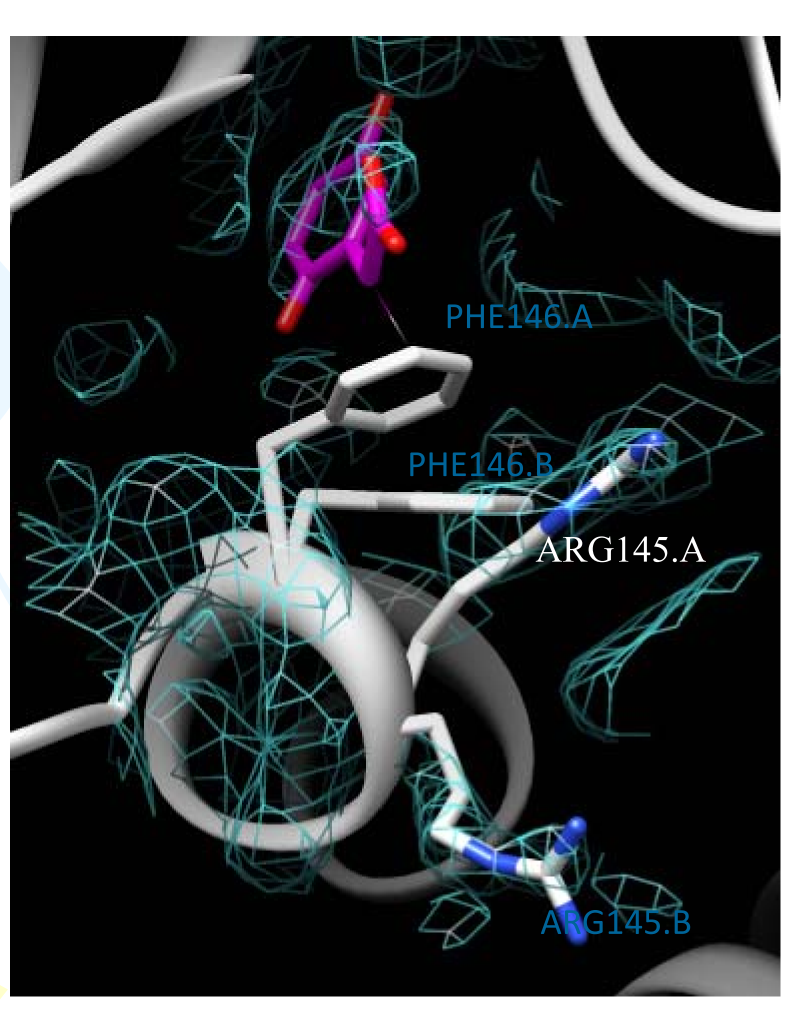


Partial atomic occupancy in 1ajp



1ajp (grey) has two conformations of side chains on PHE146 and ARG145, whose atomic occupancy is 0.5. If the alternative conformations that are similar to 1ajq (green) were chosen, the estimated energy is only -1.87 kcal/mol due to the appeared clash (magenta ligand and PHE146.A).

Poor fitting quality of the model structure to the electron density map for 1ajp



- The electron density was presented by Chimera with the setting of “Level = 0.426”.
- Apparently, the side chains of PHE146 dost not map to the electron density shown in this figure.
- A close contact exists between the ligand and PHE146.A

Can you see what are missing?

```
ATOM   1237  CA  PRO 1 187      61.209  11.195 102.027  1.00 43.25           C
ATOM   1238  C   PRO 1 187      61.917  12.397 101.399  1.00 47.87           C
ATOM   1239  O   PRO 1 187      61.528  13.539 101.726  1.00 49.71           O
ATOM   1240  CB  PRO 1 187      62.113  10.495 103.048  1.00 32.38           C
ATOM   1241  CG  PRO 1 187      61.665   9.101 103.004  1.00 37.10           C
ATOM   1242  CD  PRO 1 187      61.501   8.854 101.511  1.00 38.61           C
ATOM   1243  OXT PRO 1 187      62.872  12.183 100.608  1.00 57.89           O
TER    1244      PRO 1 187
ATOM   1245  N   HIS 2  12      32.324  25.971  37.670  1.00 46.84           N
ATOM   1246  CA  HIS 2  12      32.672  26.018  36.221  1.00 51.31           C
ATOM   1247  C   HIS 2  12      33.793  25.007  35.979  1.00 52.55           C
ATOM   1248  O   HIS 2  12      34.859  25.103  36.586  1.00 54.39           O
ATOM   1249  CB  HIS 2  12      33.143  27.436  35.840  1.00 58.43           C
ATOM   1250  CG  HIS 2  12      33.401  27.633  34.368  1.00 68.86           C
ATOM   1251  ND1 HIS 2  12      34.217  28.636  33.883  1.00 68.65           N
ATOM   1252  CD2 HIS 2  12      32.923  26.983  33.279  1.00 70.09           C
```

In principle, you should not see hydrogen atoms in a PDB file unless .....

- The structure is determined by nucleic magnetic resonance (NMR) or ultra-high resolution X-ray crystallography.
- The positions of the hydrogen atoms are determined by **neutron diffraction.**

Protonation

International Edition: DOI: 10.1002/anie.201701038
German Edition: DOI: 10.1002/ange.201701038

Charges Shift Protonation: Neutron Diffraction Reveals that Aniline and 2-Aminopyridine Become Protonated Upon Binding to Trypsin

*Johannes Schiebel, Roberto Gaspari, Anna Sandner, Khang Ngo, Hans-Dieter Gerber, Andrea Cavalli, Andreas Ostermann, Andreas Heine, and Gerhard Klebe**

Abstract: *Hydrogen atoms play a key role in protein–ligand recognition. They determine the quality of established H-bonding networks and define the protonation of bound ligands. Structural visualization of H atoms by X-ray crystallography is rarely possible. We used neutron diffraction to determine the positions of the hydrogen atoms in the ligands aniline and 2-aminopyridine bound to the archetypical serine protease trypsin. The resulting structures show the best resolution so far achieved for proteins larger than 100 residues and allow an accurate description of the protonation states and interactions with nearby water molecules. Despite its low pK_a of 4.6 and a large distance of 3.6 Å to the charged Asp189 at the bottom of the S1 pocket, the amino group of aniline becomes protonated, whereas in 2-aminopyridine, the pyridine nitrogen picks up the proton although its amino group is 1.6 Å closer to Asp189. Therefore, apart from charge–charge distances, tautomer stability is decisive for the resulting binding poses, an aspect that is pivotal for predicting correct binding.*

Angew. Chem. Int. Ed. **2017**, *56*, 4887–4890

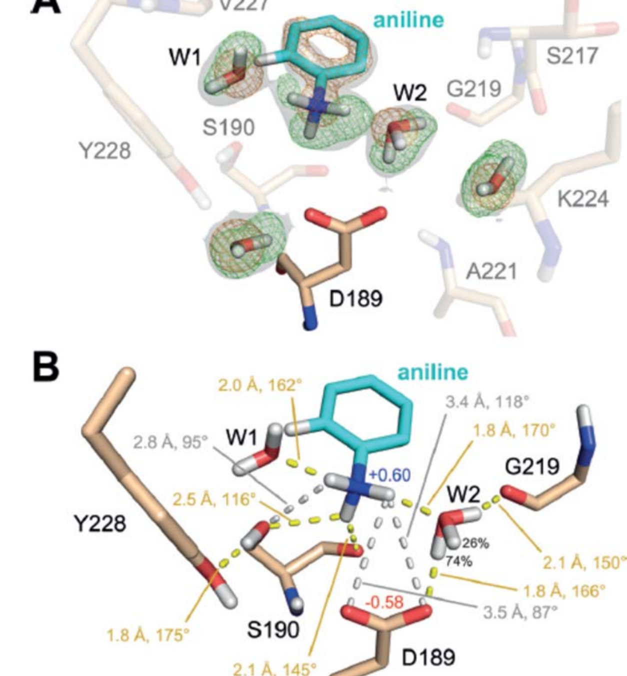


Figure 1. Protonation state and interaction pattern of aniline within the S1 pocket of trypsin. The XN structure of the trypsin-aniline complex is shown in stick representation. Protein residues are colored light brown while the ligand is depicted in cyan. For clarity, only selected protein hydrogen atoms are shown. All ligand hydrogen atoms that were visible in the difference nuclear density during model building are

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The determination of protonation states in proteins

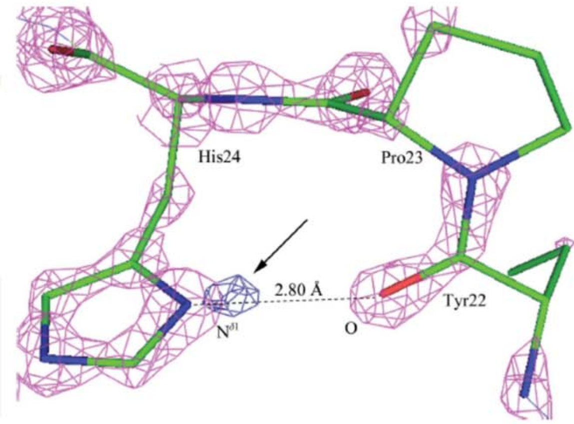


Figure 1
$2F_o - F_c$ electron-density map section (magenta, 4σ) showing His24 $N^{\delta 1}$ forming a hydrogen bond with Tyr22 main-chain O. The $F_o - F_c$ electron-density map (blue, 3σ, marked with an arrow) indicates the H atom on the $N^{\delta 1}$ atom.

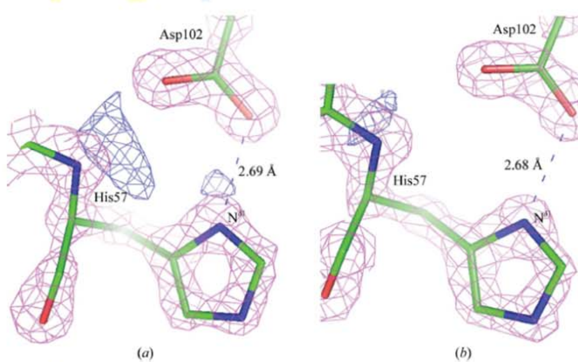
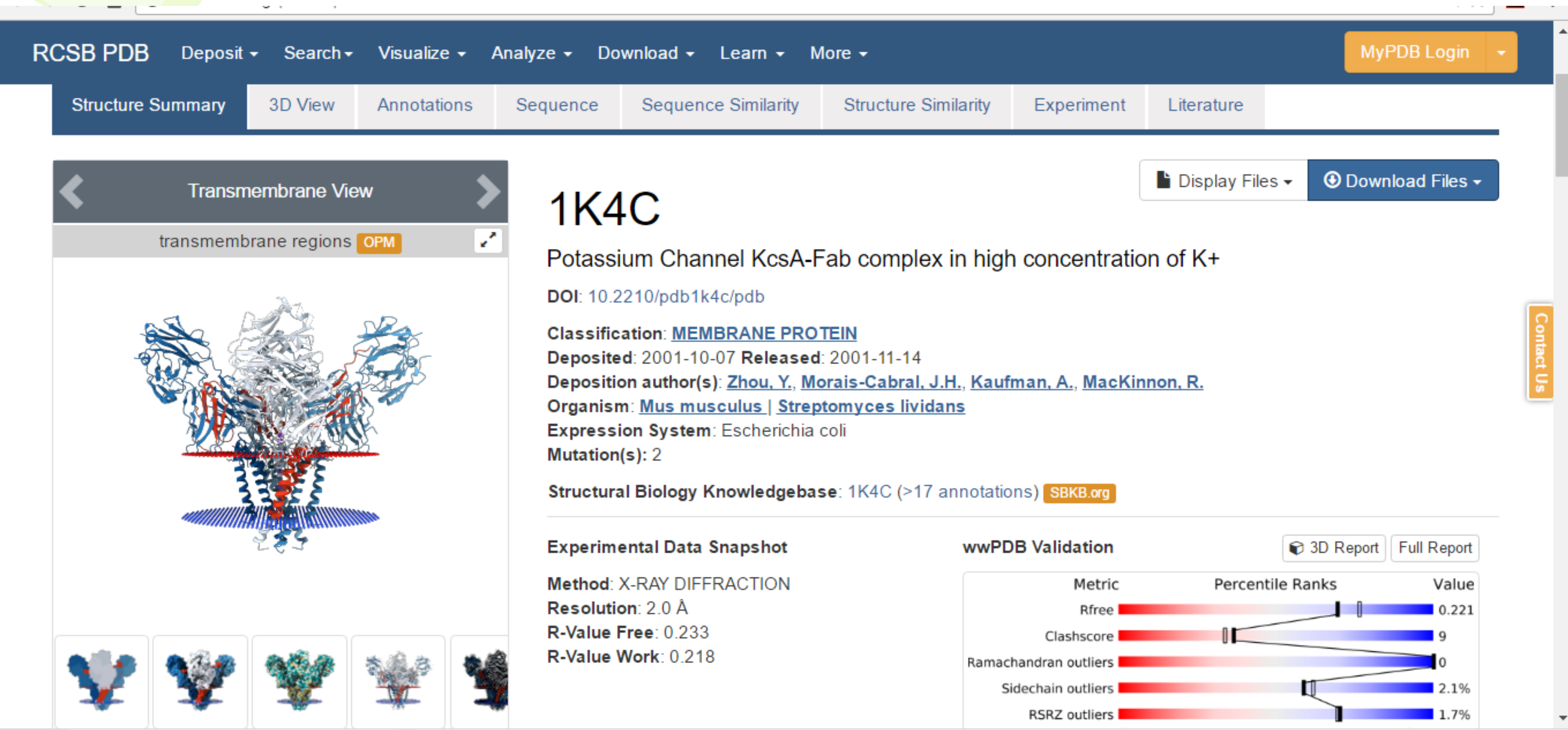


Figure 2
(*a*) The 1.26 Å thrombin–hirugen binary complex triad active site with His57 and Asp102. The $2F_o - F_c$ electron-density map is shown in magenta (3.5 r.m.s.) and the $F_o - F_c$ map (2.5 r.m.s.) is visible on the amide N and $N^{\delta 1}$ of His57. However, when contoured at 3 r.m.s. the $F_o - F_c$ map peak is no longer visible on $N^{\delta 1}$. (*b*) The 1.32 Å thrombin–hirugen–gw473178 ternary complex triad active site with His57 and Asp102. The $2F_o - F_c$ electron-density map is shown in magenta (3.5 r.m.s.) and the $F_o - F_c$ map peak in blue (2.5 r.m.s.) is visible below the amide N (slightly offset) only.

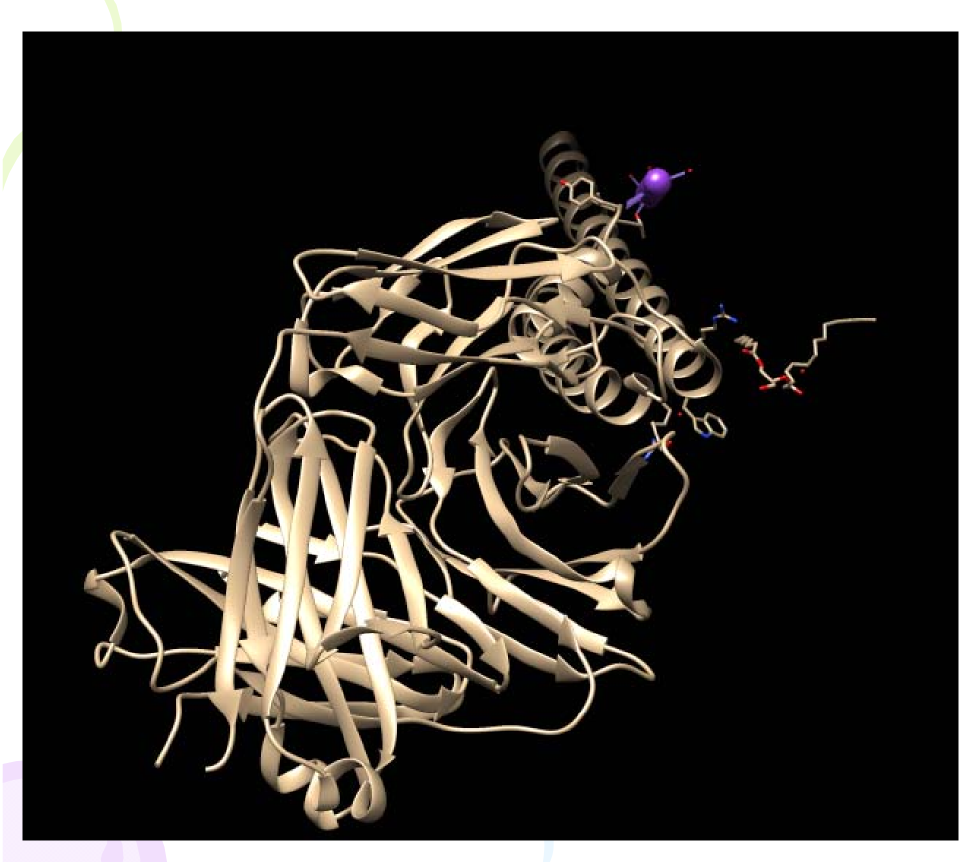
The number of H atoms exceeds the total number of C, N, O and S atoms in an enzyme. In high-resolution structures, the detailed structural information improves our understanding of enzyme mechanisms. Ultrahigh-resolution X-ray structures (typically ~0.9–1.0 Å) have allowed the determination of atom positions with very high accuracy, including the positions of H atoms. In the X-ray structure of human aldose reductase solved at an exceptional 0.66 Å resolution, 54% of the possible H atoms in the whole protein were observed in $F_o - F_c$ maps contoured at 2.5σ (Howard *et al.*, 2004). The X-ray structure of the serine protease subtilisin was solved at 0.78 Å resolution with 65% of the H atoms modelled in the final $F_o - F_c$ map contoured at 2σ; 14% of them appeared in the map contoured at 3σ (Kuhn *et al.*, 1998). H atoms have also been observed at 1.10 Å resolution: in elastase, His57 was singly protonated on

Preprocessing the PDB files ……

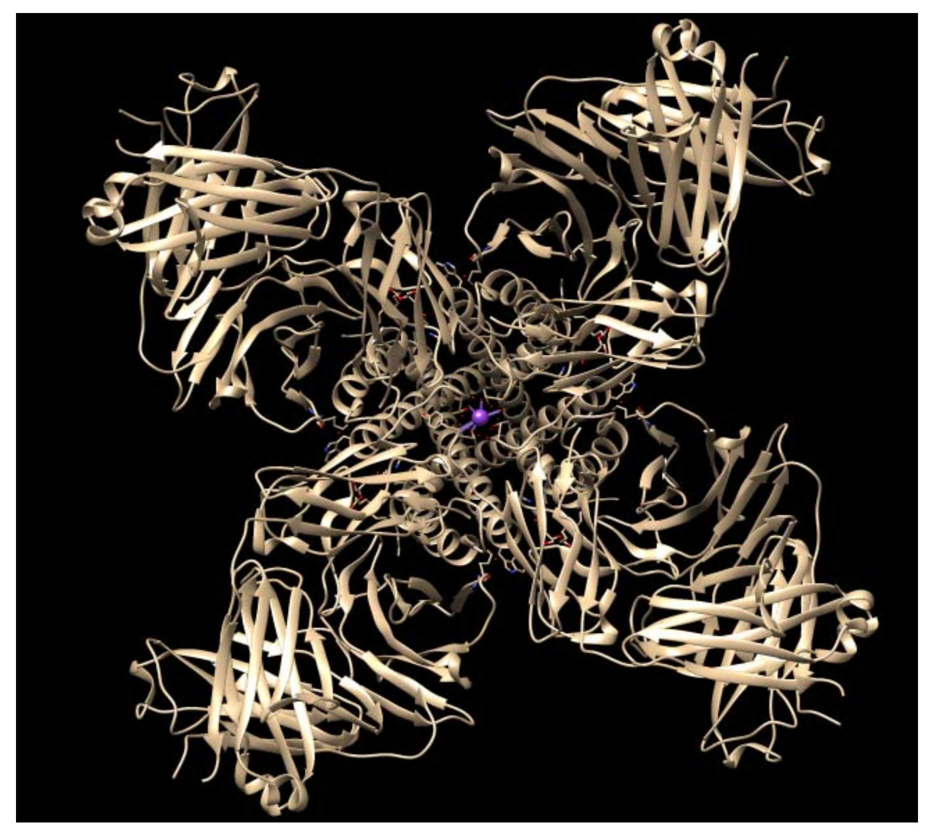
- Construct the functional unit of the biomolecule(s) of interest.
- Make sure if there are missing residues in the PDB file.
- Check if there are post-translational modifications of amino acids.
- Decide whether the crystal water molecules should be kept.
- Decide whether the metal ion should be loaded, and decide the configuration of the ions.
- Check whether there should be disulfide bonds between cysteines. (CYS:cysteine; CYX: cystine)
- Determine the pH condition for the simulations.
- Predict the protonation states of titratable residues according to the given pH condition.
- Make sure whether the N- and C- terminal should be “capped” or not.


RCSB PDB
Deposit
Search
Visualize
Analyze
Download
Learn
More
MyPDB Login
Structure Summary
3D View
Annotations
Sequence
Sequence Similarity
Structure Similarity
Experiment
Literature
Transmembrane View
transmembrane regions
OPM
Display Files
Download Files
1K4C
Potassium Channel KcsA-Fab complex in high concentration of K+
DOI: 10.2210/pdb1k4c/pdb
Classification: MEMBRANE PROTEIN
Deposited: 2001-10-07 Released: 2001-11-14
Deposition author(s): Zhou, Y., Morais-Cabral, J.H., Kaufman, A., MacKinnon, R.
Organism: Mus musculus | Streptomyces lividans
Expression System: Escherichia coli
Mutation(s): 2
Structural Biology Knowledgebase: 1K4C (>17 annotations) SBKB.org
Experimental Data Snapshot
Method: X-RAY DIFFRACTION
Resolution: 2.0 Å
R-Value Free: 0.233
R-Value Work: 0.218
wwPDB Validation
3D Report
Full Report
Metric
Percentile Ranks
Value
Rfree
0.221
Clashscore
9
Ramachandran outliers
0
Sidechain outliers
2.1%
RSRZ outliers
1.7%
Contact Us

CHIMERA: K^+ channel

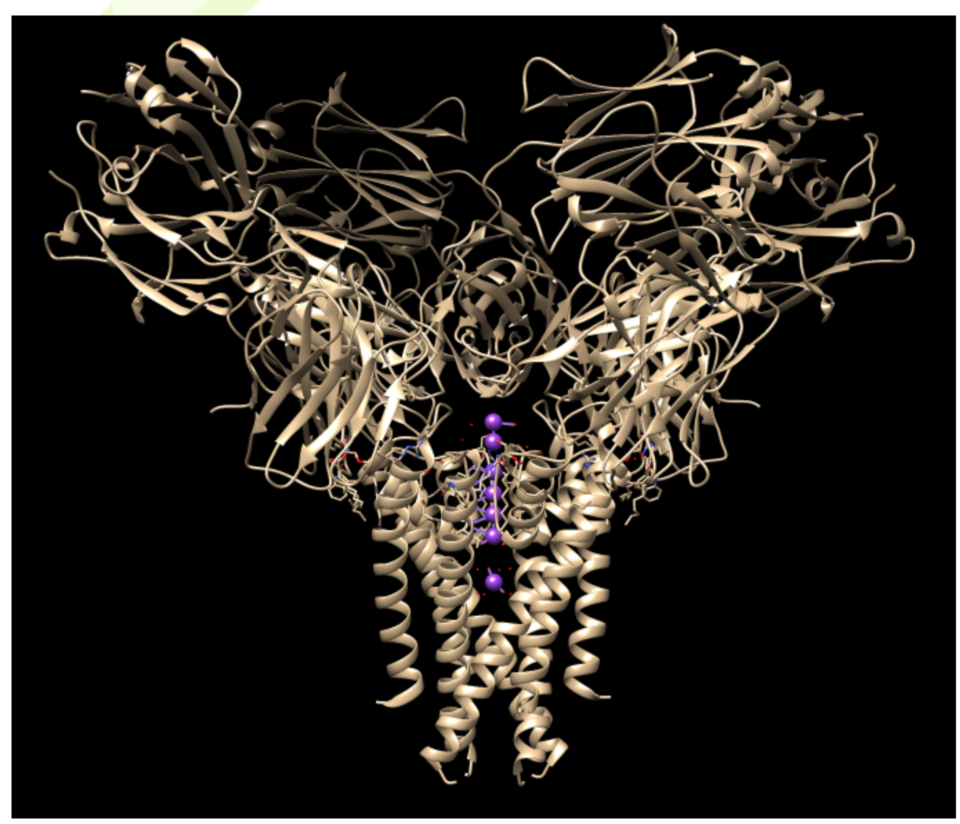


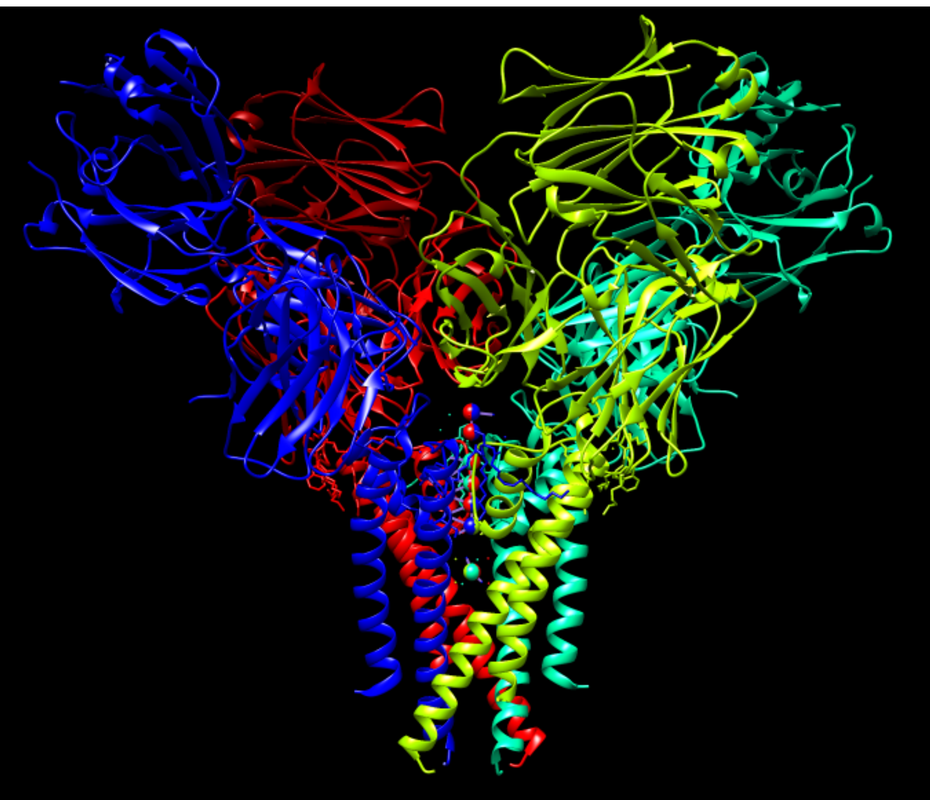
1k4c.pdb

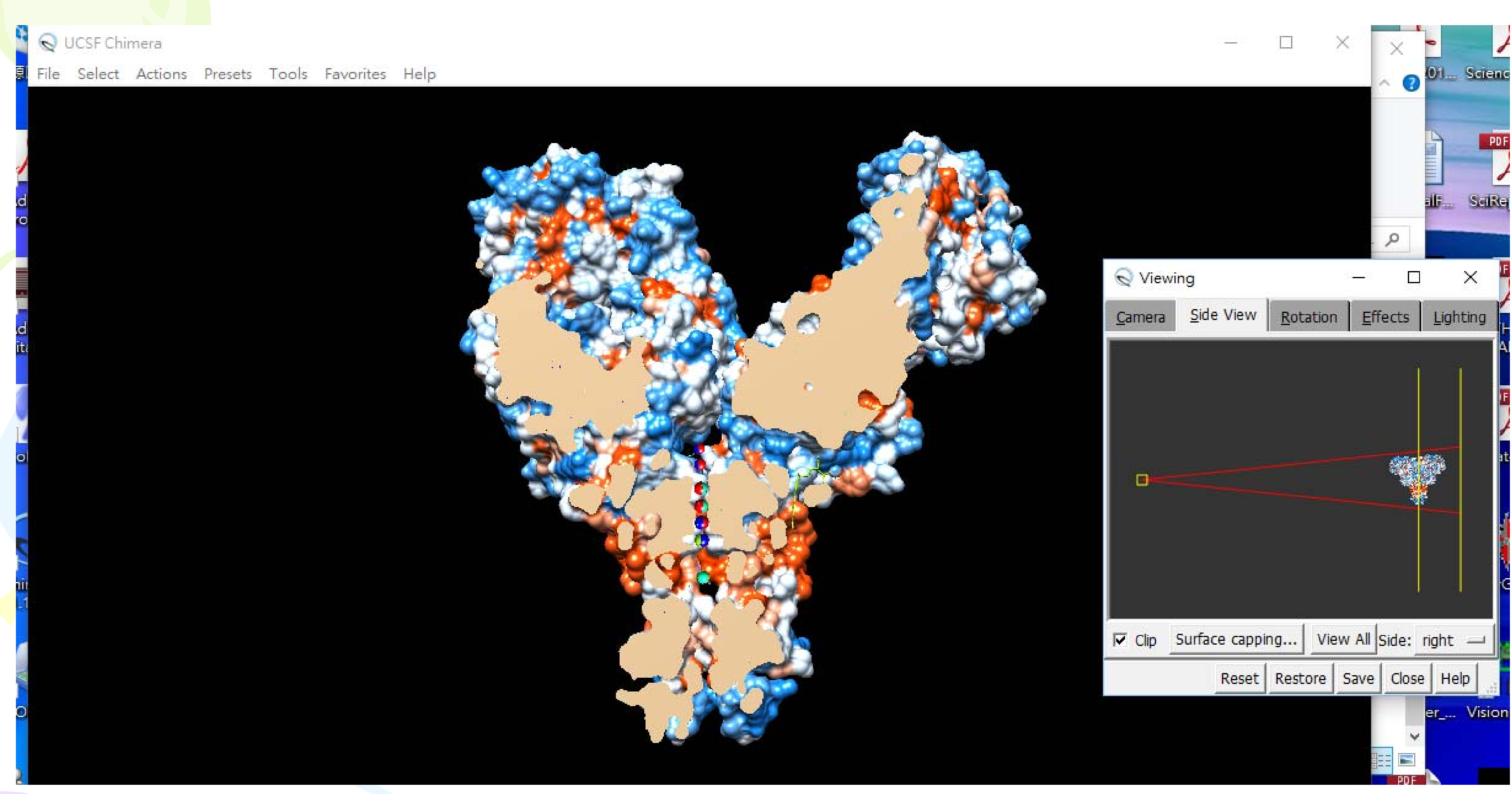


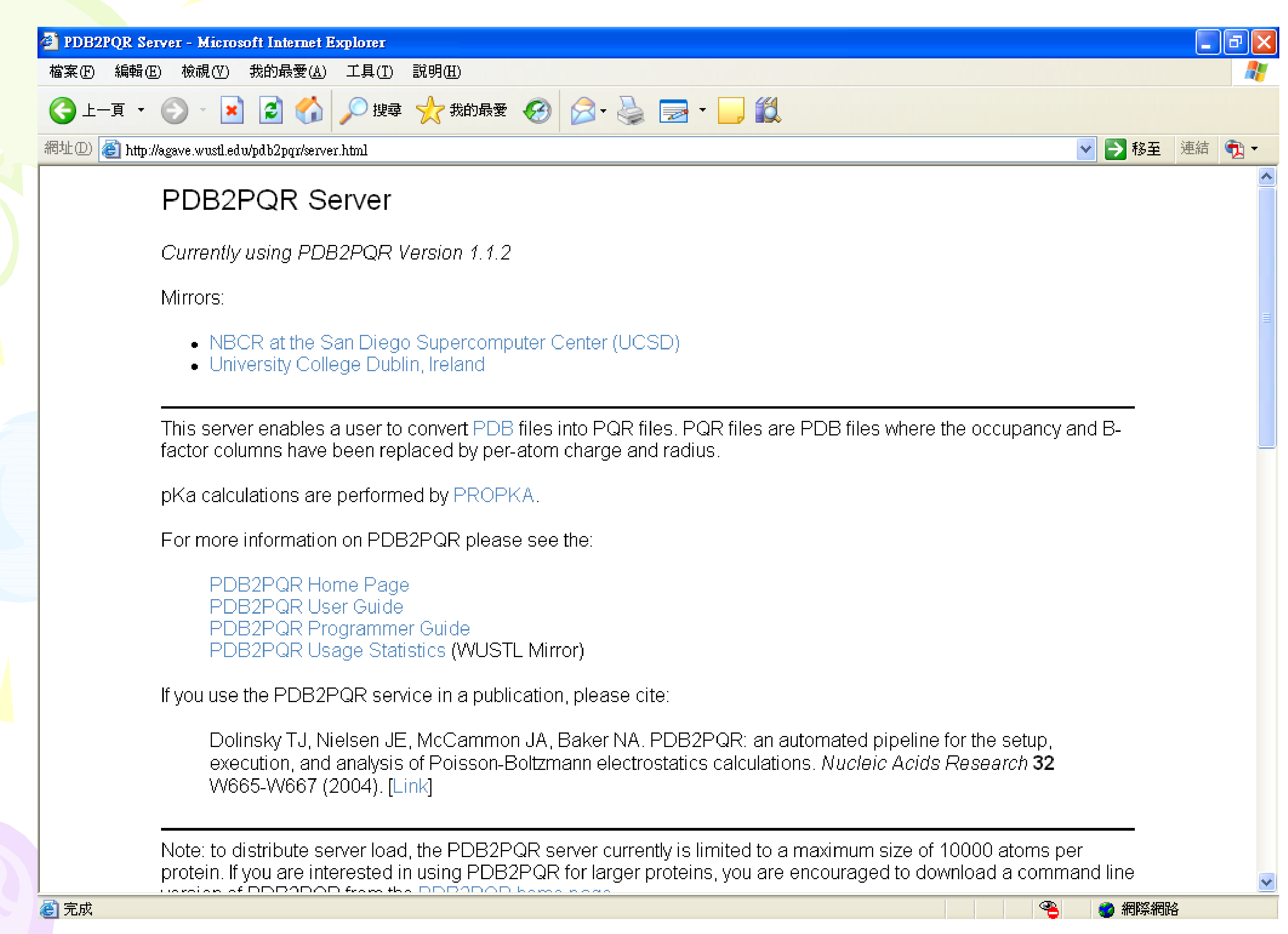
1k4c.pdb1

```
HETATM 4022  C2  F09 A2001     157.182 133.064 -37.120  1.00 61.72           C
HETATM 4023  C3  F09 A2001     156.625 131.660 -36.846  1.00 61.62           C
HETATM 4024  C4  F09 A2001     155.246 131.683 -36.169  1.00 61.61           C
HETATM 4025  C5  F09 A2001     155.319 131.799 -34.639  1.00 61.51           C
HETATM 4026  C6  F09 A2001     155.716 130.474 -33.993  1.00 61.74           C
HETATM 4027  C7  F09 A2001     155.669 130.525 -32.473  1.00 61.86           C
HETATM 4028  C8  F09 A2001     156.103 129.184 -31.889  1.00 62.11           C
HETATM 4029  C9  F09 A2001     156.083 129.170 -30.359  1.00 62.49           C
HETATM 4030  OXT F09 A2001     156.850 129.882 -29.697  1.00 62.52           O
HETATM 4031  K      K C3001     155.336 155.342 -30.553  0.25 14.49           K
HETATM 4032  K      K C3002     155.331 155.331 -33.953  0.25 15.51           K
HETATM 4033  K      K C3003     155.341 155.323 -37.162  0.25 15.69           K
HETATM 4034  K      K C3004     155.330 155.324 -40.505  0.25 16.84           K
HETATM 4035  K      K C3005     155.327 155.335 -47.577  0.25 24.94           K
HETATM 4036  K      K C3006     155.339 155.327 -22.975  0.25 47.51           K
HETATM 4037  K      K C3007     155.343 155.330 -26.017  0.25 65.44           K
HETATM 4038  CA1 DGA C1001     165.096 138.737 -29.005  1.00 69.56           C
HETATM 4039  CA2 DGA C1001     165.422 139.686 -30.114  1.00 68.93           C
HETATM 4040  CA3 DGA C1001     164.534 139.447 -31.305  1.00 67.90           C
HETATM 4041  CA4 DGA C1001     165.110 140.093 -32.543  1.00 67.30           C
```

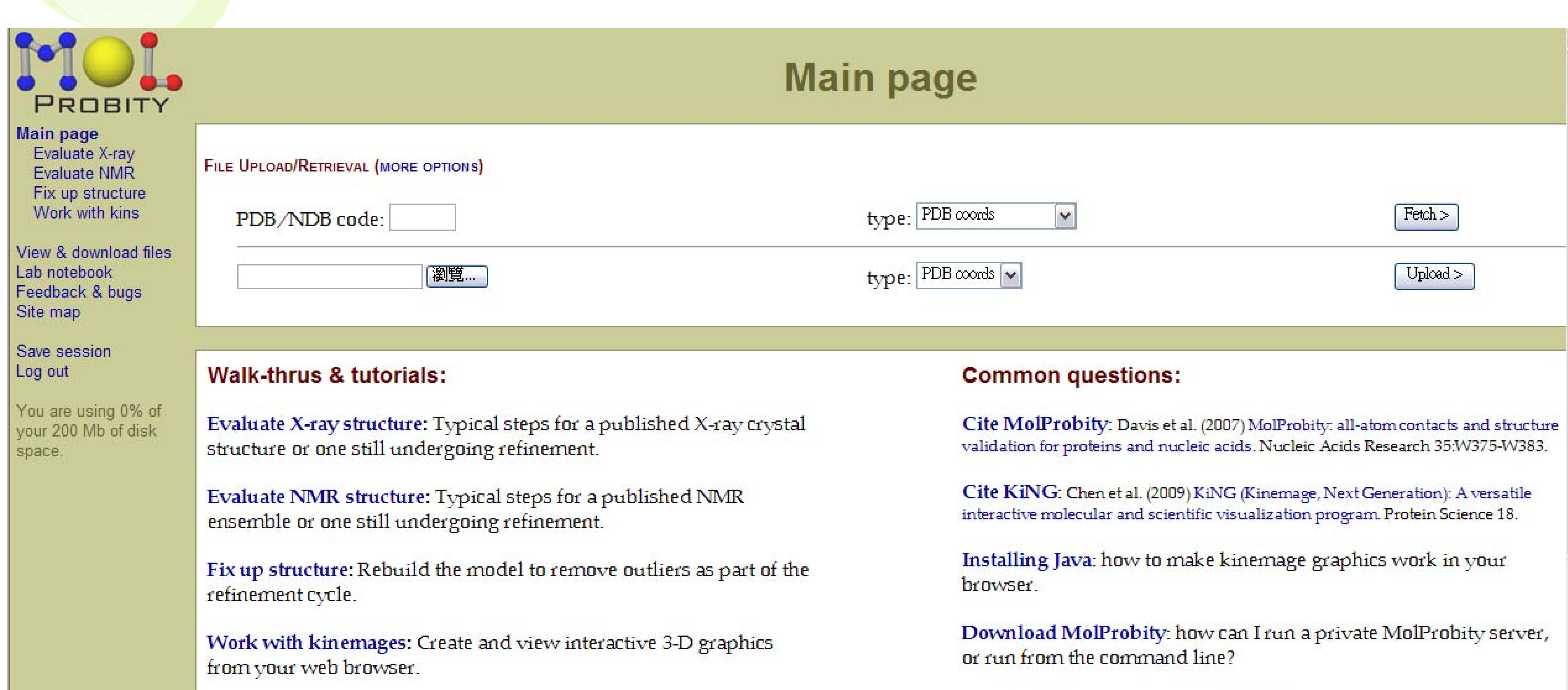





UCSF Chimera
File Select Actions Presets Tools Favorites Help
Viewing
Camera Side View Rotation Effects Lighting
Clip Surface capping... View All Side: right
Reset Restore Save Close Help



http://agave.wustl.edu/pdb2pqr/server.html



http://molprobity.biochem.duke.edu/

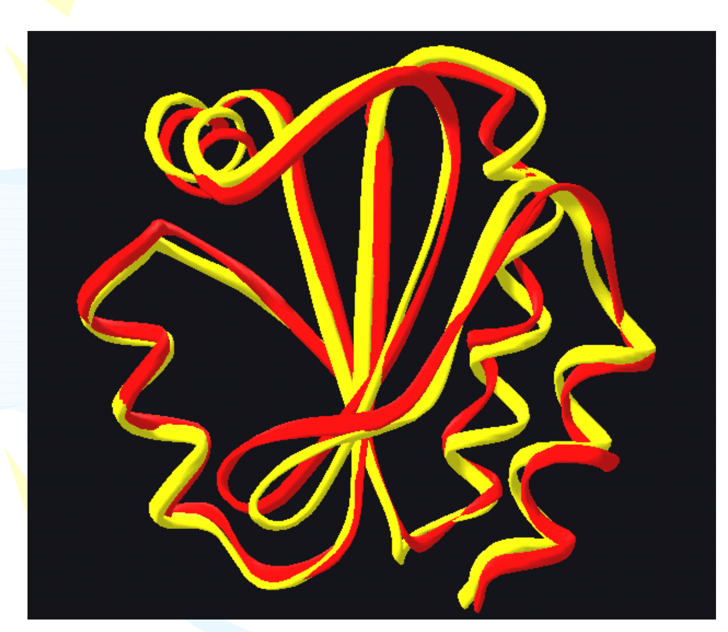
Comparative Modeling Procedure

1) Aligning of the target sequence on the backbone of the parent structures
2) Building a structure framework
3) Constructing of core side chains
4) Building the loops
5) Refining the models
6) Estimating the reliability of models

Superposition of two protein structures

- Translate the second protein structure so that its center of masses matches the center of masses of the first protein.
- Rotate the second protein structure in the way that the root-mean-square deviation of two structures are minimal.

Quantifying structural difference: root mean square deviation (RMSD)



$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r'}_i)^2}{N}}$$

Acta Cryst. (1976). A**32**, 922

A solution for the best rotation to relate two sets of vectors. By WOLFGANG KABSCH, *Max–Planck-Institut für Medizinische Forschung*, 6900 *Heidelberg*, *Jahnstrasse* 29, *Germany* (*BRD*)

(*Received* 23 *February* 1976; *accepted* 12 *April* 1976)

A simple procedure is derived which determines a best rotation of a given vector set into a second vector set by minimizing the weighted sum of squared deviations. The method is generalized for any given metric constraint on the transformation.

Acta Cryst. (1989). A**45**, 208–210

On the orthogonal transformation used for structural comparisons. By SIMON K. KEARSLEY, *Department of Chemistry*, *Yale University*, 225 *Prospect Street*, *New Haven*, *CT* 06511, *USA*

(*Received* 5 *May* 1988; *accepted* 11 *August* 1988)

Using Quaternions to Calculate RMSD

EVANGELOS A. COUTSIAS,[1] CHAOK SEOK,[2] KEN A. DILL[3]
[1]*Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131*
[2]*School of Chemistry, College of Natural Sciences, Seoul National University, Gwanak-gu, Shillim-dong, San 56-1, Seoul 151-747, Republic of Korea*
[3]*Department of Pharmaceutical Chemistry, University of California in San Francisco, San Francisco, California 94143-2240*

Received 29 December 2003; Accepted 13 July 2004
DOI 10.1002/jcc.20110
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: A widely used way to compare the structures of biomolecules or solid bodies is to translate and rotate one structure with respect to the other to minimize the root-mean-square deviation (RMSD). We present a simple derivation, based on quaternions, for the optimal solid body transformation (rotation-translation) that minimizes the RMSD between two sets of vectors. We prove that the quaternion method is equivalent to the well-known formula due to Kabsch. We analyze the various cases that may arise, and give a complete enumeration of the special cases in terms of the arrangement of the eigenvalues of a traceless, 4×4 symmetric matrix. A key result here is an expression for the gradient of the RMSD as a function of model parameters. This can be useful, for example, in finding the minimum energy path of a reaction using the elastic band methods or in optimizing model parameters to best fit a target structure.

© 2004 Wiley Periodicals, Inc. J Comput Chem 25: 1849–1857, 2004

Kearsley's approach

Acta Cryst. ***A.* 45:** 208-210 (1989)

$$x_m \equiv (x'-x), \quad x_p \equiv (x'+x) \qquad y_m \equiv (y'-y), \quad y_p \equiv (y'+y) \qquad z_m \equiv (z'-z), \quad z_p \equiv (z'+z)$$

$$\begin{pmatrix} \sum(x_m^2+y_m^2+z_m^2) & \sum(y_p z_m - y_m z_p) & \sum(x_m z_p - x_p z_m) & \sum(x_p y_m - x_m y_p) \\ \sum(y_p z_m - y_m z_p) & \sum(y_p^2+z_p^2+x_m^2) & \sum(x_m y_m - x_p y_p) & \sum(x_m z_m - x_p z_p) \\ \sum(x_m z_p - x_p z_m) & \sum(x_m y_m - x_p y_p) & \sum(x_p^2+z_p^2+y_m^2) & \sum(y_m z_m - y_p z_p) \\ \sum(x_p y_m - x_m y_p) & \sum(x_p y_m - x_m y_p) & \sum(y_m z_m - y_p z_p) & \sum(x_p^2+y_p^2+z_m^2) \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}$$

$$= \lambda \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}$$

$$RMSD = \sqrt{\frac{\lambda}{N}}$$

Normalization of Quaternions

$$e_0 = q_0 / \left(q_0^2 + q_1^2 + q_2^2 + q_3^2 \right)$$

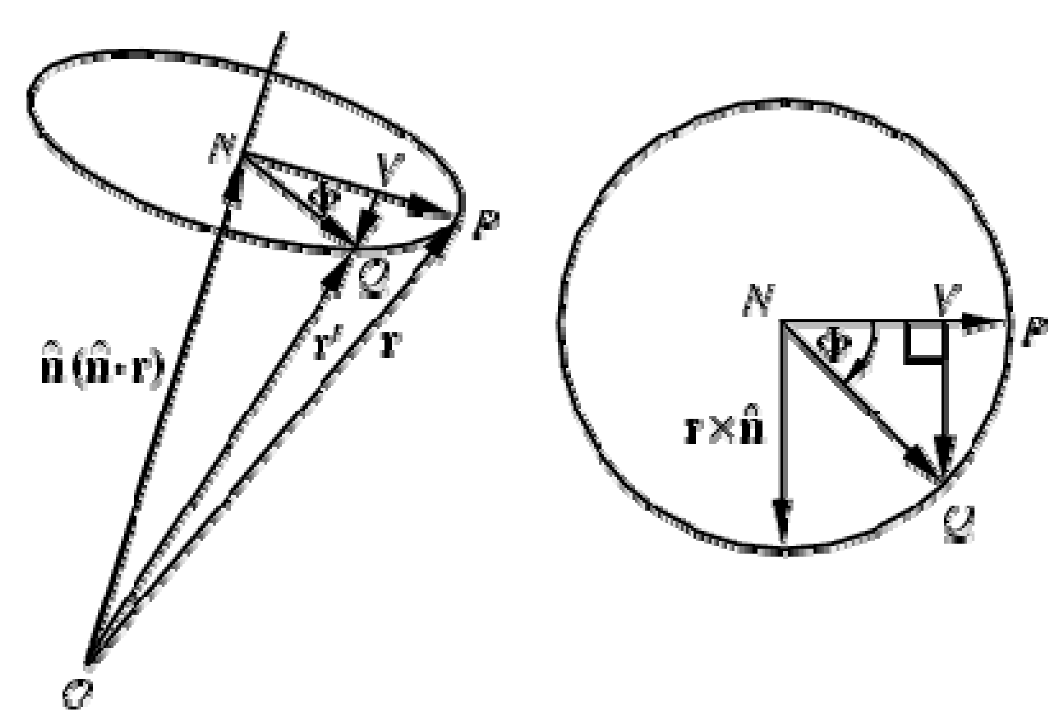
$$e_1 = q_1 / \left(q_0^2 + q_1^2 + q_2^2 + q_3^2 \right)$$

$$e_2 = q_2 / \left(q_0^2 + q_1^2 + q_2^2 + q_3^2 \right)$$

$$e_3 = q_3 / \left(q_0^2 + q_1^2 + q_2^2 + q_3^2 \right)$$

$$e_0^2 + e_1^2 + e_2^2 + e_3^2 = 1$$

Gibbs Rotation Formula



$$\mathbf{r'} = \mathbf{r}\cos\Phi + \hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{r})(1-\cos\Phi) + (\mathbf{r}\times\hat{\mathbf{n}})\sin\Phi$$

$$= \mathbf{r}(e_0^2 - e_1^2 - e_2^2 - e_3^2) + 2\mathbf{e}(\mathbf{e}\cdot\mathbf{r}) + 2(\mathbf{r}\times\mathbf{e})e_0$$

Mapping Quaternions to Elements of the Rotation Matrix

$$e_0 \equiv \cos(\frac{\phi}{2})$$

$$\mathbf{e} \equiv \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \hat{\mathbf{n}} \sin(\frac{\phi}{2})$$

$$\alpha_{11} = e_0^2 + e_1^2 - e_2^2 - e_3^2$$

$$\alpha_{12} = 2(e_1 e_2 + e_0 e_3)$$

$$\alpha_{13} = 2(e_1 e_3 - e_0 e_2)$$

$$\alpha_{21} = 2(e_1 e_2 - e_0 e_3)$$

$$\alpha_{22} = e_0^2 - e_1^2 + e_2^2 - e_3^2$$

$$\alpha_{23} = 2(e_2 e_3 + e_0 e_1)$$

$$\alpha_{31} = 2(e_1 e_3 + e_0 e_2)$$

$$\alpha_{32} = 2(e_2 e_3 - e_0 e_1)$$

$$\alpha_{33} = e_0^2 - e_1^2 - e_2^2 + e_3^2$$

$$\mathbf{D} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\alpha_{11} = \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi$$

$$\alpha_{12} = \cos\psi\sin\phi - \cos\theta\cos\phi\sin\psi$$

$$\alpha_{13} = \sin\psi\sin\phi$$

$$\alpha_{21} = -\sin\psi\cos\phi - \cos\theta\sin\phi\sin\psi$$

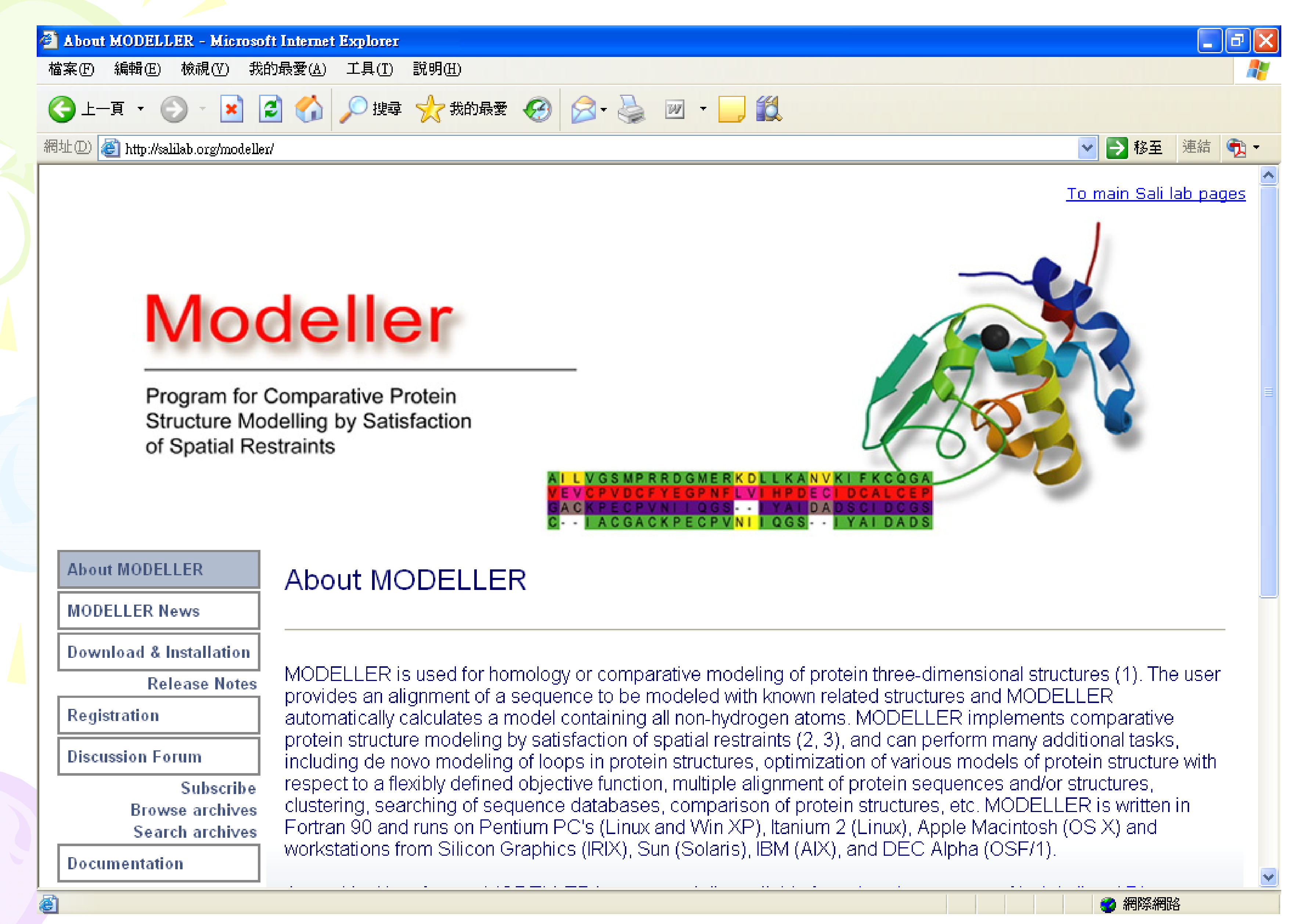
$$\alpha_{22} = -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi$$

$$\alpha_{23} = \cos\psi\sin\theta$$

$$\alpha_{31} = \sin\theta\sin\phi$$

$$\alpha_{32} = -\sin\theta\cos\phi$$

$$\alpha_{33} = \cos\theta$$



http://salilab.org/modeller/

J. Mol. Biol. (1993) **234**, 779–815

Comparative Protein Modelling by Satisfaction of Spatial Restraints

Andrej Šali† and Tom L. Blundell

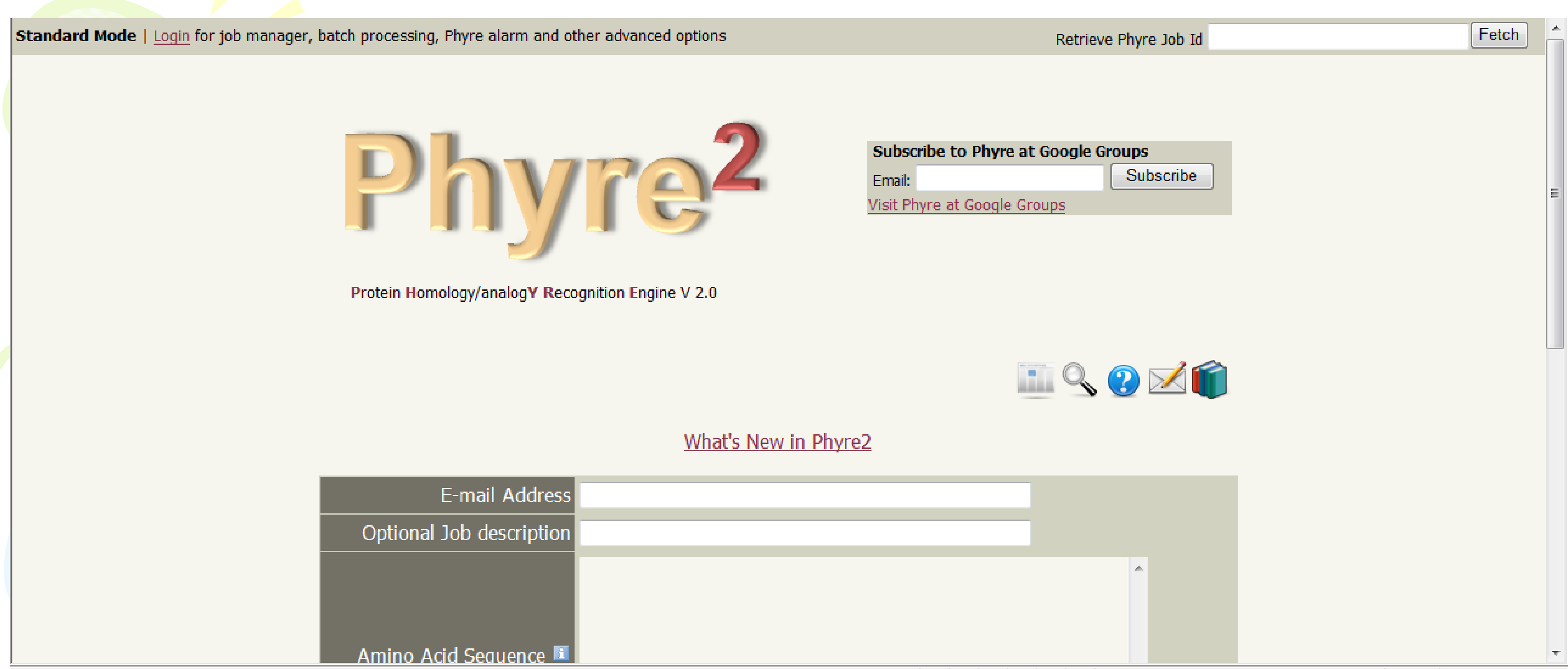
ICRF Unit of Structural Molecular Biology
Department of Crystallography
Birkbeck College, London WC1E 7HX, England

(Received 30 March 1993; accepted 13 July 1993)

We describe a comparative protein modelling method designed to find the most probable structure for a sequence given its alignment with related structures. The three-dimensional (3D) model is obtained by optimally satisfying spatial restraints derived from the alignment and expressed as probability density functions (pdfs) for the features restrained. For example, the probabilities for main-chain conformations of a modelled residue may be restrained by its residue type, main-chain conformation of an equivalent residue in a related protein, and the local similarity between the two sequences. Several such pdfs are obtained from the correlations between structural features in 17 families of homologous proteins which have been aligned on the basis of their 3D structures. The pdfs restrain C^{α}–C^{α} distances, main-chain N–O distances, main-chain and side-chain dihedral angles. A smoothing procedure is used in the derivation of these relationships to minimize the problem of a sparse database. The 3D model of a protein is obtained by optimization of the molecular pdf such that the model violates the input restraints as little as possible. The molecular pdf is derived as a combination of pdfs restraining individual spatial features of the whole molecule. The optimization procedure is a variable target function method that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms. The method is automated and is illustrated by the modelling of trypsin from two other serine proteinases.

Keywords: comparative protein modelling; restraints; optimization; protein database; serine proteinases

Times cited: 10022 (2018/3/15)



http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

Nature Protocols **4**, 363 (2009) Times cited: 4176 (2018/3/15)

PROTOCOL

The Phyre2 web portal for protein modeling, prediction and analysis

Times cited: 1721 (2018/3/15)

Lawrence A Kelley[1], Stefans Mezulis[1], Christopher M Yates[1,2], Mark N Wass[1,2] & Michael J E Sternberg[1]

[1]Structural Bioinformatics Group, Imperial College London, London, UK. [2]Present addresses: University College London (UCL) Cancer Institute, London, UK (C.M.Y.); Centre for Molecular Processing, School of Biosciences, University of Kent, Kent, UK (M.N.W.). Correspondence should be addressed to L.A.K. (l.a.kelley@imperial.ac.uk).

Nature Protocols **10**, 845 (2015)

PROTOCOL

Protein structure prediction on the Web: a case study using the Phyre server

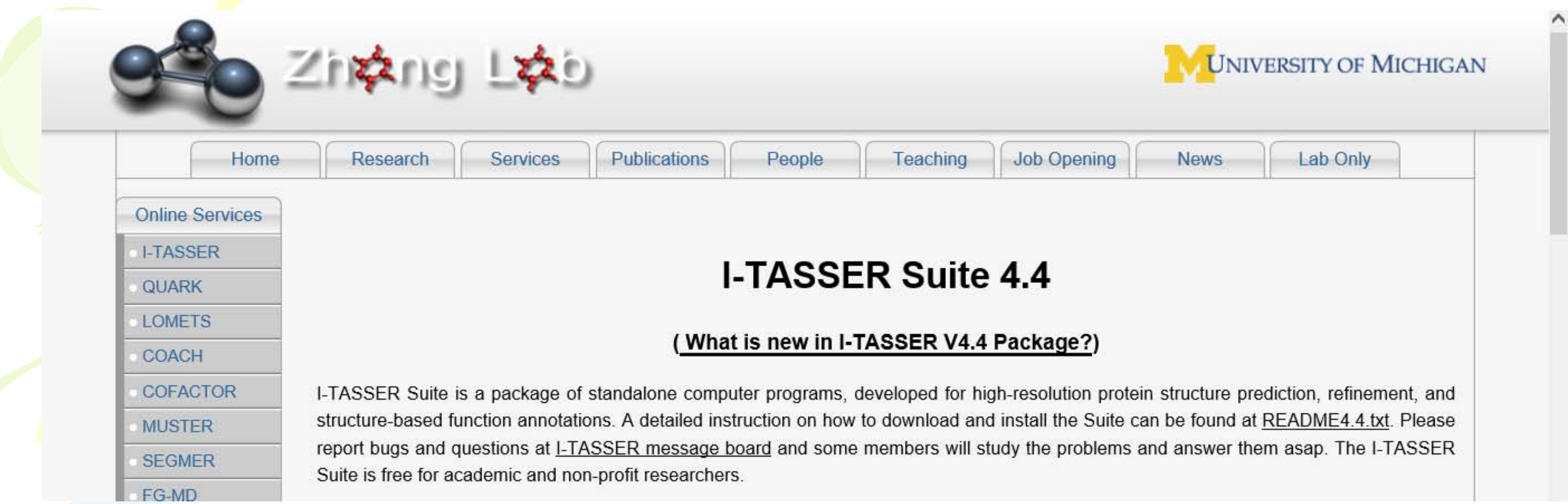
Lawrence A Kelley & Michael J E Sternberg

Structural Bioinformatics Group, Division of Molecular Biosciences, Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. Correspondence should be addressed to L.A.K (l.a.kelley@imperial.ac.uk).

Published online 26 February 2009; doi:10.1038/nprot.2009.2

Determining the structure and function of a novel protein is a cornerstone of many aspects of modern biology. Over the past decades, a number of computational tools for structure prediction have been developed. It is critical that the biological community is aware of such tools and is able to interpret their results in an informed way. This protocol provides a guide to interpreting the output of structure prediction servers in general and one such tool in particular, the protein homology/analogy recognition engine (Phyre). New profile–profile matching algorithms have improved structure prediction considerably in recent years. Although the performance of Phyre is typical of many structure prediction systems using such algorithms, all these systems can reliably detect up to twice as many remote homologies as standard sequence-profile searching. Phyre is widely used by the biological community, with >150 submissions per day, and provides a simple interface to results. Phyre takes 30 min to predict the structure of a 250-residue protein.

Times cited: 4176 (2018/3/15)



http://zhanglab.ccmb.med.umich.edu/I-TASSER/download/

The I-TASSER Suite: protein structure and function prediction

To the Editor: Assignment of structure and function to all genes and gene products (such as proteins) of all organisms represents a major challenge in this postgenomic era. Here we present the I-TASSER Suite (http://zhanglab.ccmb.med.umich.edu/I-TASSER/download/), a stand-alone software package for protein structure and function modeling.

NATURE METHODS | VOL.12 NO.1 | JANUARY 2015 | 7

Times cited: 1190 (2018/3/15)

PROTOCOL

I-TASSER: a unified platform for automated protein structure and function prediction

Ambrish Roy[1,2], Alper Kucukural[2] & Yang Zhang[1,2]

[1]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [2]Center for Bioinformatics and Department of Molecular Bioscience, University of Kansas, Lawrence, Kansas, USA. Correspondence should be addressed to Y.Z. (zhng@umich.edu).

Published online 25 March 2010; doi:10.1038/nprot.2010.5

NATURE PROTOCOLS | VOL.5 NO.4 | 2010 | 725

Times cited: 3326 (2018/3/15)

PROTOCOL

Protein structure prediction on the Web: a case study using the Phyre server

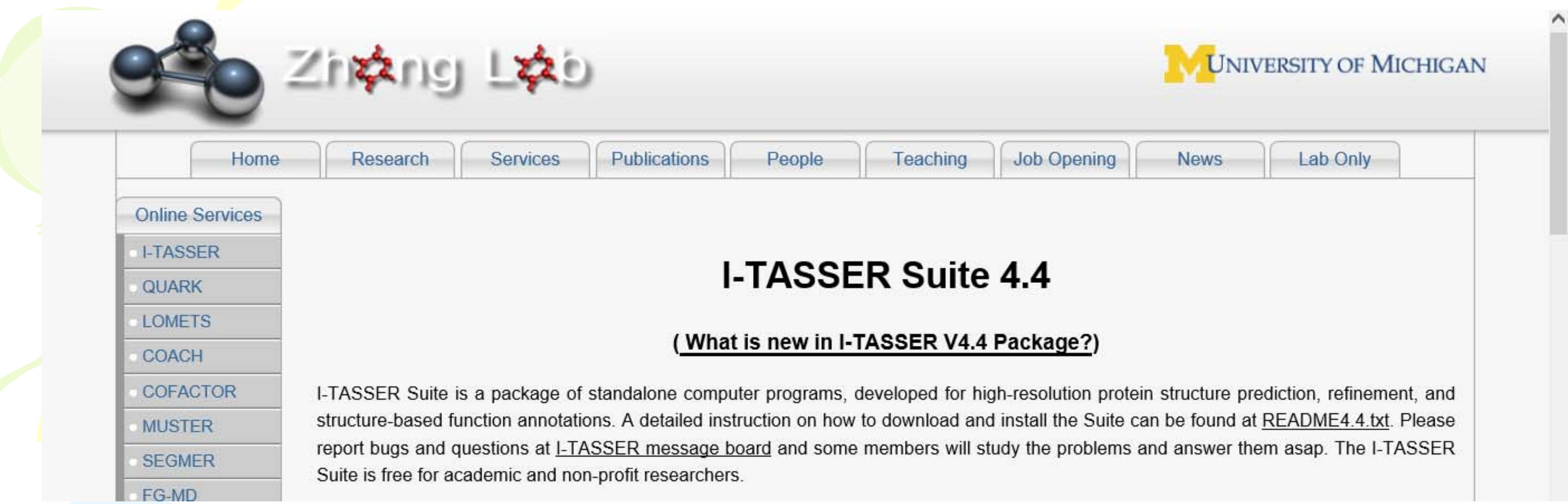
Lawrence A Kelley & Michael J E Sternberg

Structural Bioinformatics Group, Division of Molecular Biosciences, Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. Correspondence should be addressed to L.A.K (l.a.kelley@imperial.ac.uk).

Published online 26 February 2009; doi:10.1038/nprot.2009.2

Determining the structure and function of a novel protein is a cornerstone of many aspects of modern biology. Over the past decades, a number of computational tools for structure prediction have been developed. It is critical that the biological community is aware of such tools and is able to interpret their results in an informed way. This protocol provides a guide to interpreting the output of structure prediction servers in general and one such tool in particular, the protein homology/analogy recognition engine (Phyre). New profile–profile matching algorithms have improved structure prediction considerably in recent years. Although the performance of Phyre is typical of many structure prediction systems using such algorithms, all these systems can reliably detect up to twice as many remote homologies as standard sequence-profile searching. Phyre is widely used by the biological community, with >150 submissions per day, and provides a simple interface to results. Phyre takes 30 min to predict the structure of a 250-residue protein.

Times cited: 3353 (2016/5/30)



http://zhanglab.ccmb.med.umich.edu/I-TASSER/download/

The I-TASSER Suite: protein structure and function prediction

To the Editor: Assignment of structure and function to all genes and gene products (such as proteins) of all organisms represents a major challenge in this postgenomic era. Here we present the I-TASSER Suite (http://zhanglab.ccmb.med.umich.edu/I-TASSER/download/), a stand-alone software package for protein structure and function modeling.

NATURE METHODS | VOL.12 NO.1 | JANUARY 2015 | 7

Times cited: 254 (2016/5/30)

PROTOCOL

I-TASSER: a unified platform for automated protein structure and function prediction

Ambrish Roy[1,2], Alper Kucukural[2] & Yang Zhang[1,2]

[1]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. [2]Center for Bioinformatics and Department of Molecular Bioscience, University of Kansas, Lawrence, Kansas, USA. Correspondence should be addressed to Y.Z. (zhng@umich.edu).

Published online 25 March 2010; doi:10.1038/nprot.2010.5

NATURE PROTOCOLS | VOL.5 NO.4 | 2010 | 725

Times cited: 2300 (2016/5/30)