Protein packing model with the universal geometric factor of volume and surface area ratio

Ming-Chya Wu (吳明佳)

Department of Physics, National Central University, Taoyuan, Taiwan
Institute of Physics, Academia Sinica, Taipei, Taiwan

E-mail: mcwu@ncu.edu.tw
Homepage: http://in.ncu.edu.tw/mcwu/

Plan of this talk

To introduce a *simple* and *universal parameter*, which is useful in characterizing *complicated* and *diverse protein conformations*.

Outline

Introduction

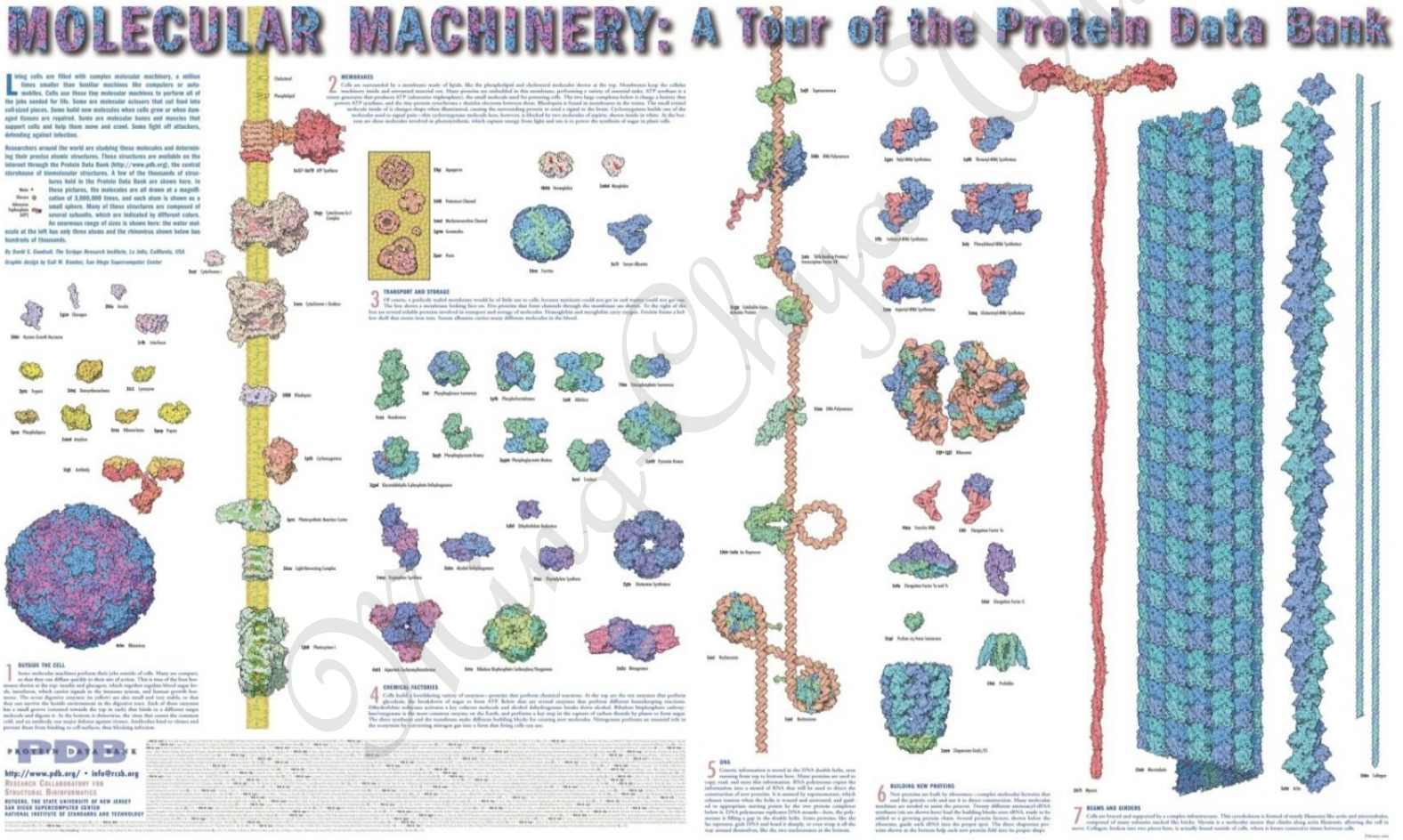
Universal geometrical factor for protein conformations

Simple packing models for the factor

Conclusion

Introduction

Proteins assume specified conformations from their chemical compositions or sequences to develop biological activity and functional specificity.

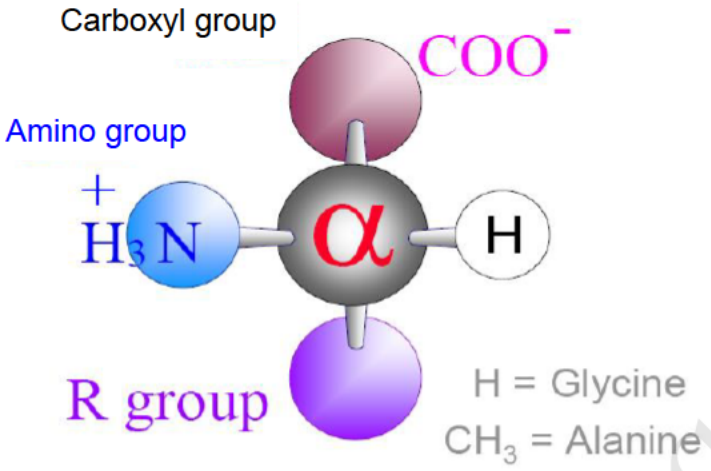


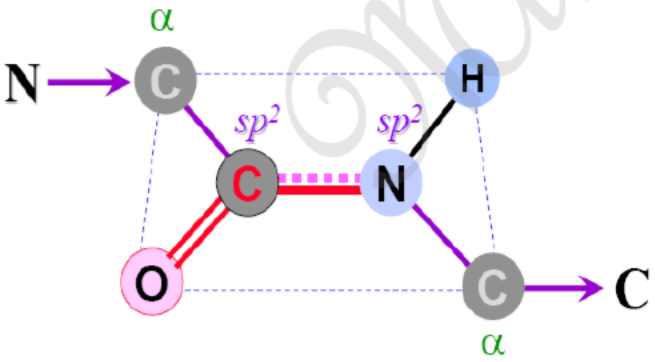
7 Classes of Proteins: proteins outside cells (I), proteins on membranes (II), proteins for transport and storage (III), proteins acting as chemical factories (IV), proteins working with DNA (V), proteins for building new proteins (VI), and proteins for building organizations (beams and girders) (VII). (Protein Data Bank, http://www.rcsb.org/pdb/index.html)

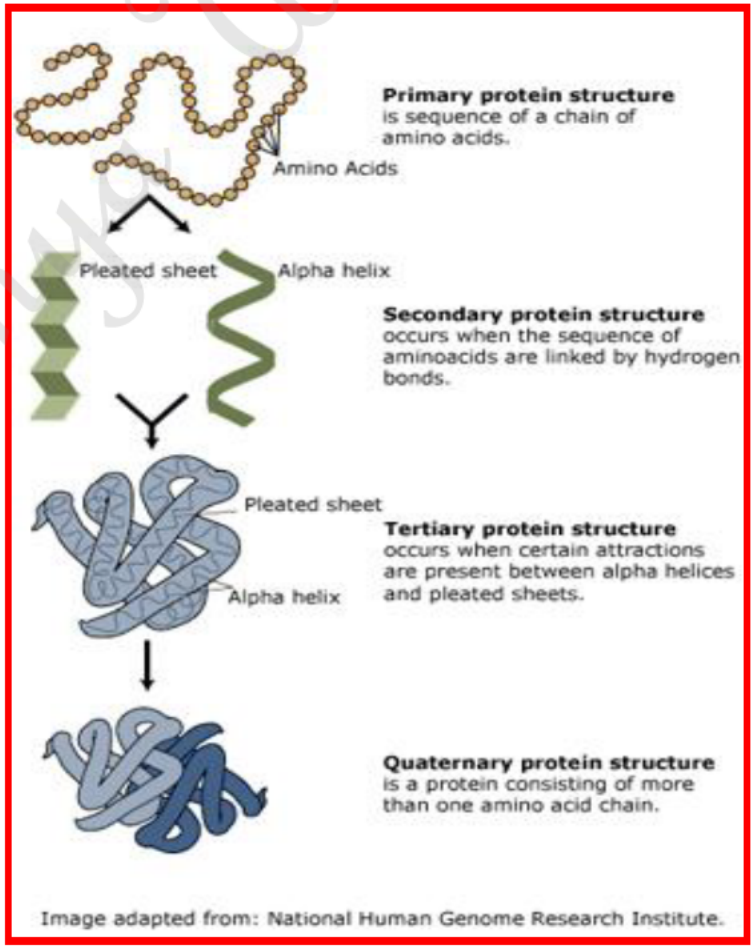
Introduction

Proteins are biomolecules, consisting of one or more long chains of amino acids residues.

Properties of the amino acid sequence of a protein determine its 3D structure.

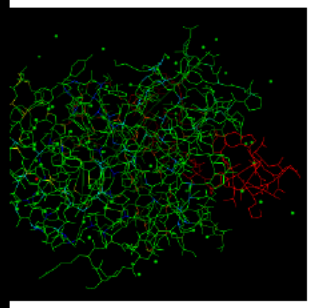




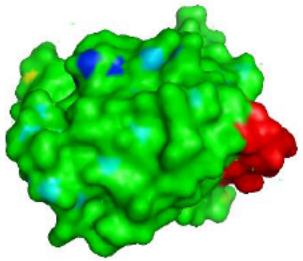


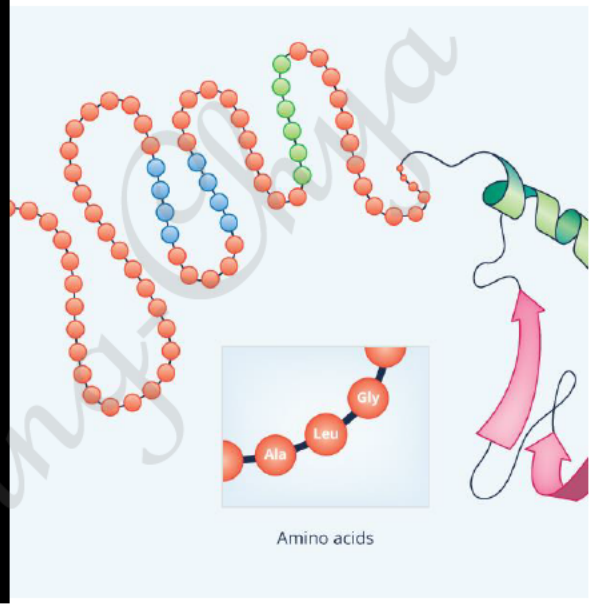
Introduction

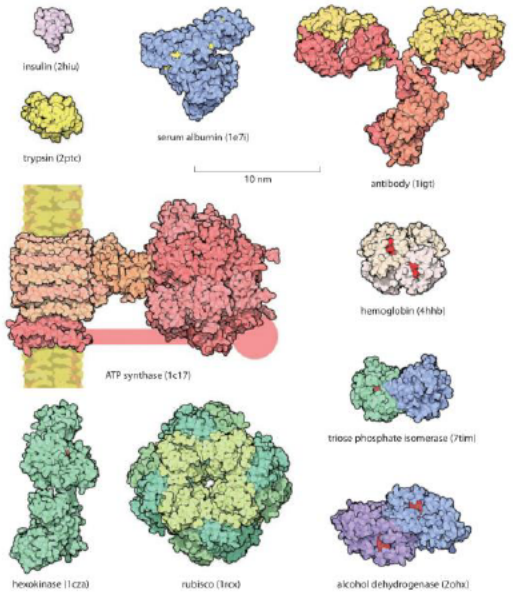
```
HEADER    HYDROLASE (SERINE PROTEINASE)              26-OCT-81   2PTN      2PTN   3
COMPND    TRYPSIN (ORTHORHOMBIC, 2.4 M AMMONIUM SULFATE)                   2PTN   4
COMPND   2 (E.C.3.4.21.4)                                                  2PTN   5
SOURCE    BOVINE (BOS TAURUS) PANCREAS                                     2PTN   6
AUTHOR    J.WALTER,W.STEIGEMANN,T.P.SINGH,H.BARTUNIK,W.BODE,R.HUBER        2PTN   7
REVDAT   6   15-APR-91 2PTNE   2       CONECT                             2PTNE  1
REVDAT   5   09-APR-85 2PTND   1       JRNL                               2PTND  1
REVDAT   4   14-MAR-85 2PTNC   3       SEQRES ATOM                        2PTNC  1
....
REMARK   7  RESIDUE 186 FROM GLN TO GLU, UPON DEPOSITORS INSTRUCTIONS.     2PTNC  4
REMARK   7  REVISE ATOM AND SEQRES RECORDS ACCORDINGLY.  14-MAR-85.        2PTNC  5
REMARK   8                                                                 2PTND  6
REMARK   8 CORRECTION. UPDATE JRNL REFERENCE TO REFLECT PUBLICATION.       2PTND  7
REMARK   8  09-APR-85.                                                     2PTND  8
REMARK   9                                                                 2PTNE  2
REMARK   9 CORRECTION. CORRECT CONECT RECORDS FOR HETATM CA.               2PTNE  3
REMARK   9  15-APR-91.                                                     2PTNE  4
SEQRES   1     223  ILE VAL GLY GLY TYR THR CYS GLY ALA ASN THR VAL PRO   2PTN  67
SEQRES   2     223  TYR GLN VAL SER LEU ASN SER GLY TYR HIS PHE CYS GLY   2PTN  68
SEQRES   3     223  GLY SER LEU ILE ASN SER GLN TRP VAL VAL SER ALA ALA   2PTN  69
SEQRES   4     223  HIS CYS TYR LYS SER GLY ILE GLN VAL ARG LEU GLY GLU   2PTN  70
SEQRES   5     223  ASP ASN ILE ASN VAL VAL GLU GLY ASN GLU GLN PHE ILE   2PTN  71
....
HELIX    1   H1 SER     164  ILE     176  1 SNGL ALPHA TURN,REST IRREG.    2PTN  90
HELIX    2   H2 LYS     230  VAL     235  5 CONTIGUOUS WITH H3             2PTN  91
HELIX    3   H3 SER     236  ASN     245  1 CONTIGUOUS WITH H2             2PTN  92
SSBOND   1 CYS      22    CYS     157                                      2PTN  93
SSBOND   2 CYS      42    CYS      58                                      2PTN  94
SSBOND   3 CYS     128    CYS     232                                      2PTN  95
SSBOND   4 CYS     136    CYS     201                                      2PTN  96
SSBOND   5 CYS     168    CYS     182                                      2PTN  97
SSBOND   6 CYS     191    CYS     220                                      2PTN  98
CRYST1  54.090   58.520   67.630  90.00  90.00  90.00 P 21 21 21    4      2PTN  99
ORIGX1      1.000000  0.000000  0.000000        0.00000                    2PTN 100
ORIGX2      0.000000  1.000000  0.000000        0.00000                    2PTN 101
ORIGX3      0.000000  0.000000  1.000000        0.00000                    2PTN 102
SCALE1      .018218  0.000000  0.000000        0.00000                    2PTN 103
SCALE2      0.000000  .017088  0.000000        0.00000                    2PTN 104
SCALE3      0.000000  0.000000  .014786        0.00000                    2PTN 105
ATOM      1  N   ILE    16      -8.155   9.648  20.365  1.00 10.60        2PTN 106
ATOM      2  CA  ILE    16      -8.150   8.766  19.179  1.00 10.60        2PTN 107
ATOM      3  C   ILE    16      -9.405   9.018  18.348  1.00 10.60        2PTN 108
ATOM      4  O   ILE    16     -10.533   8.888  18.870  1.00 10.60        2PTN 109
ATOM      5  CB  ILE    16      -8.091   7.261  19.602  1.00 10.60        2PTN 110
ATOM      6  CG1 ILE    16      -6.898   6.882  20.508  1.00  7.42        2PTN 111
ATOM      7  CG2 ILE    16      -8.178   6.281  18.408  1.00  7.42        2PTN 112
ATOM      8  CD1 ILE    16      -5.555   6.893  19.773  1.00  7.42        2PTN 113
ATOM      9  N   VAL    17      -9.224   9.305  17.090  1.00  9.63        2PTN 114
ATOM     10  CA  VAL    17     -10.351   9.448  16.157  1.00  9.63        2PTN 115
ATOM     11  C   VAL    17     -10.500   8.184  15.315  1.00  9.63        2PTN 116
ATOM     12  O   VAL    17      -9.496   7.688  14.748  1.00  9.63        2PTN 117
```



Solve structures by
X-ray, NMR, …







SMMP utilizes the following energy function

$$E_{\text{tot}} = E_{\text{LJ}} + E_{\text{el}} + E_{\text{hb}} + E_{\text{tors}},$$

where

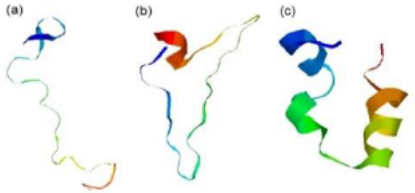
$$E_{\text{LJ}} = \sum_{j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right),$$

$$E_{\text{el}} = 332 \sum_{j>i} \frac{q_i q_j}{\varepsilon r_{ij}},$$

$$E_{\text{hb}} = \sum_{j>i} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right),$$

$$E_{\text{tors}} = \sum_{n} U_n (1 \pm \cos(k_n \varphi_n)).$$

Folding simulations



Protein data bank

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

RCSB PDB PROTEIN DATA BANK

157145 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 wwPDB EMDataResource NDB Worldwide Protein Data Bank Foundation

Welcome
Deposit
Search
Visualize
Analyze
Download
Learn

A Structural View of Biology

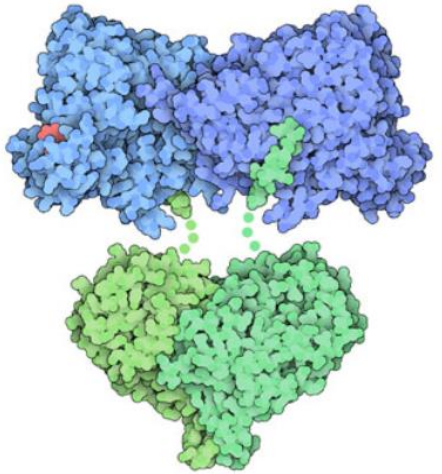
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

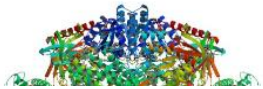
Job Opportunities for Biocurators and Developers

October **Molecule of the Month**



Ribonucleotide Reductase

Latest Entries *As of Tuesday Oct 22 2019*

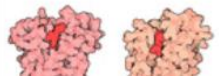


Features & Highlights

Improved resolution of DOIs for PDB entries

Access new wwPDB summary pages

News Publications



PDB Turns 48

Today is the anniversary of the 1971 announcement of

Protein data bank

PDB-101 | Molecule of the Month | Browse | Learn | Global Health | Teach | SciArt | Events | More

PDB-101

Molecular explorations through biology and medicine

Search Molecule of the Month articles and more | Go

Educational portal of RCSB PDB

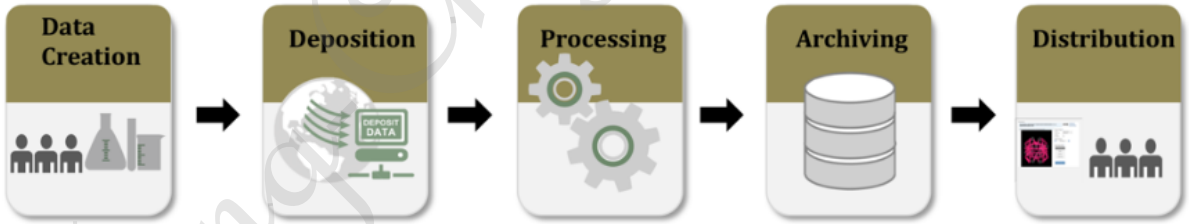
PDB & Data Archiving Modules | Modules | About | Objectives | Contributors | Index

Modules

- Introduction
- Requirements
- Infrastructure
- Deposition
- Curation
- Data Consistency
- Archiving
- Distribution

About the PDB & Data Archiving Curriculum

This curriculum introduces best practices recommendations for anyone interested in developing and managing a data pipeline for a curated public archive of biological experimental data, leveraging the extensive experience of the RCSB PDB team. View this short overview (pdf) to learn more.



Development of this resource has been supported by National Institutes of Health National Library of Medicine R25LM012286 as part of the BD2K Initiative.

About PDB-101

PDB-101 helps teachers, students, and the general public explore the 3D world of proteins and nucleic acids. Learning about their diverse shapes and functions helps to understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease to biological energy.

Why PDB-101? Researchers around the globe make these 3D structures freely available at the Protein Data Bank (PDB) archive. PDB-101 builds introductory materials to help beginners get started in the subject ("101", as in an entry level course) as well as resources for extended learning.

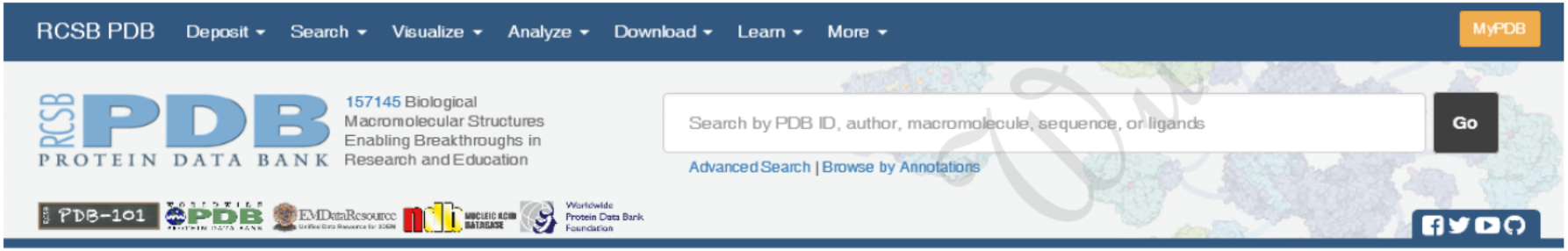
RCSB PDB (citation) is hosted by

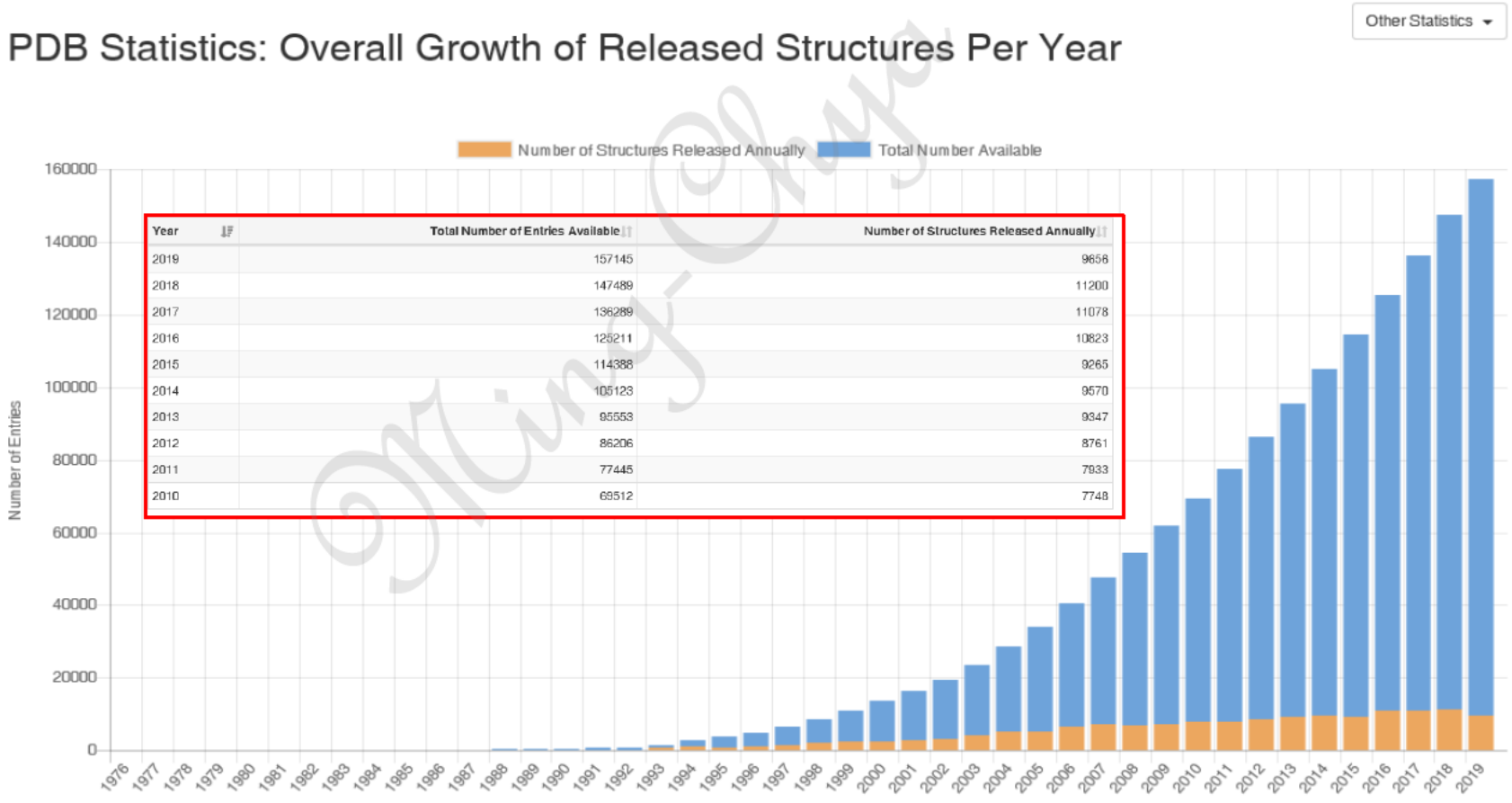
Rutgers | UC San Diego | SDSC | UCSF

RCSB PDB is a member of worldwide PDB | EMDataResource

RCSB PDB is funded by a grant from the National Science Foundation (DBI-1338415), the National Institutes of Health, and the US Department of Energy (DE-SC0019749).

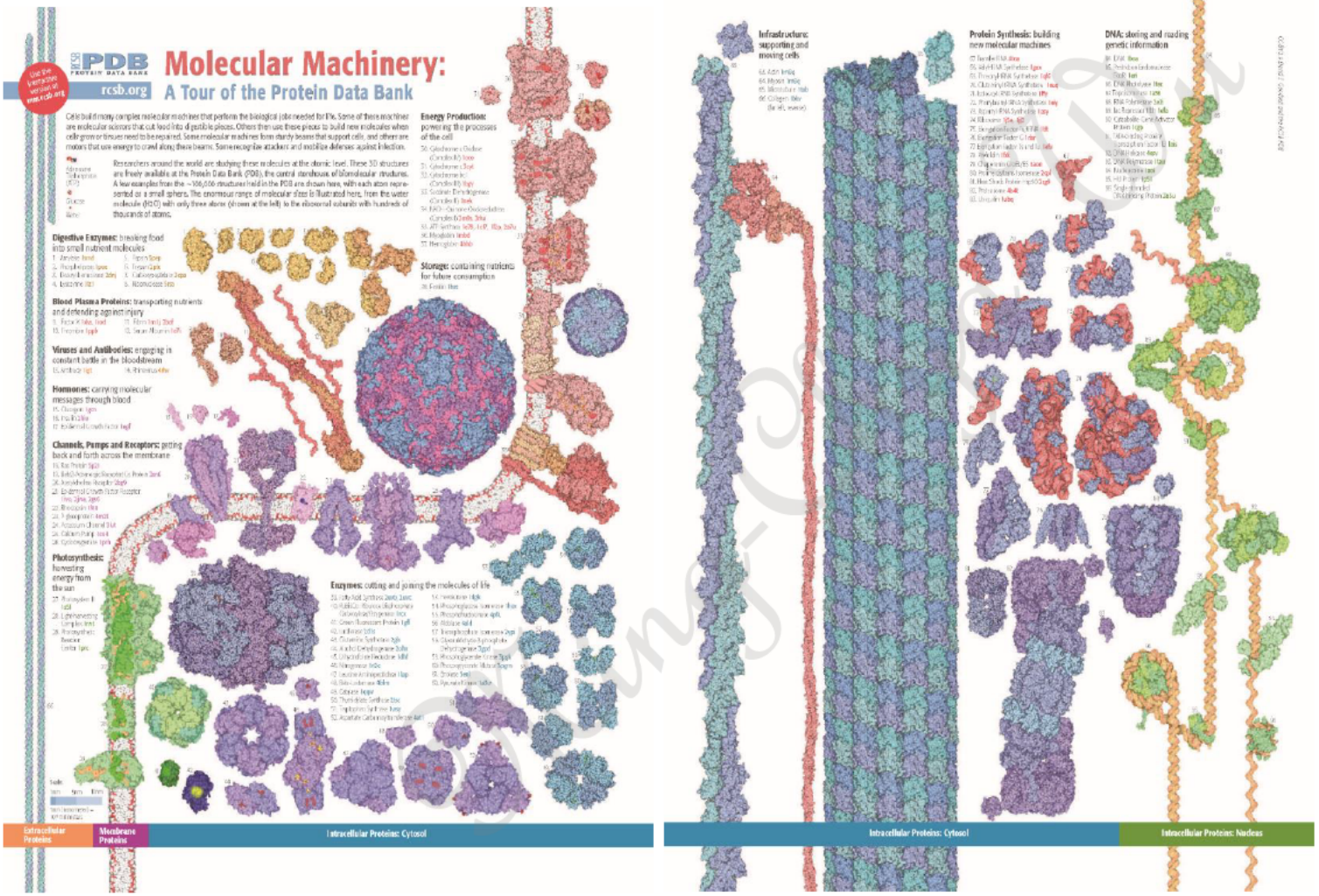
Protein data bank

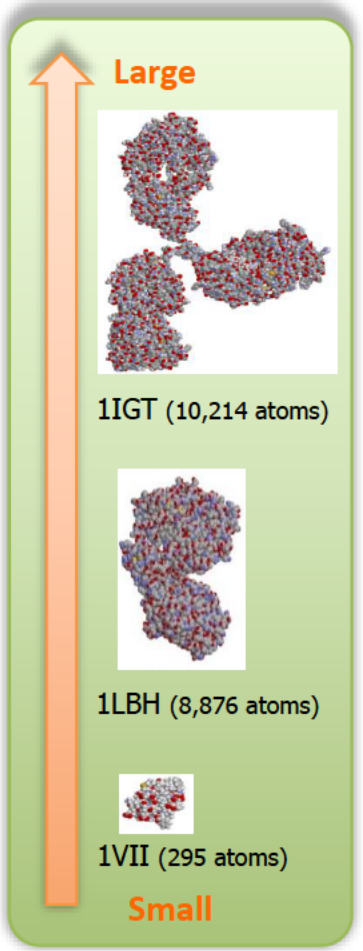


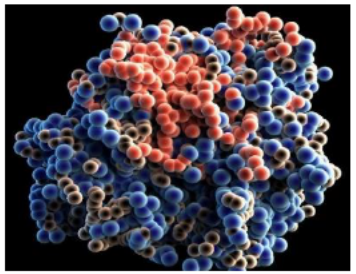


Year	Total Number of Entries Available	Number of Structures Released Annually
2019	157145	9658
2018	147489	11200
2017	136289	11078
2016	125211	10823
2015	114388	9265
2014	105123	9570
2013	95553	9347
2012	86206	8761
2011	77445	7933
2010	69512	7748

Protein conformation and function



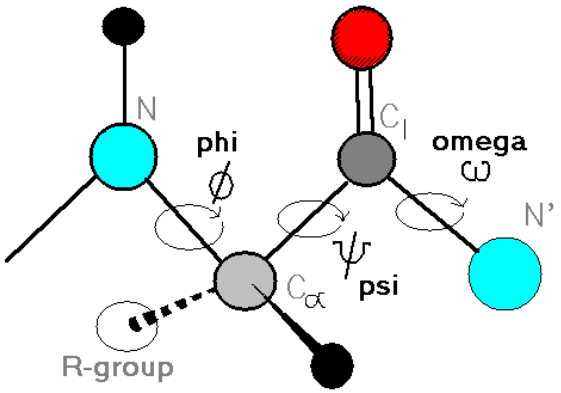


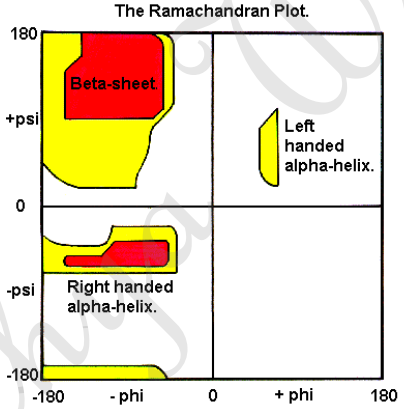


Basically, a protein can be considered as a number of atoms (spheres) in a certain arrangement.

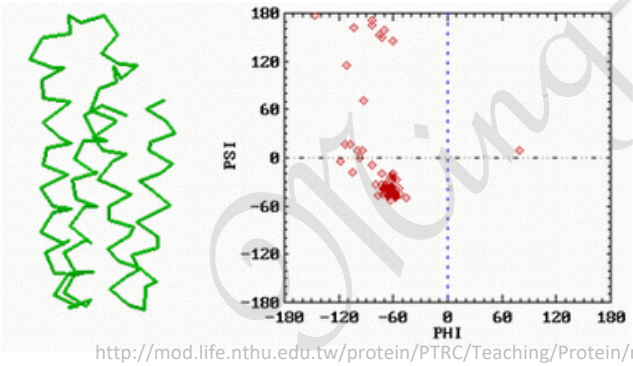
Ramachandran plot

The main chain N-C_α and C_α-C bonds are relatively free to rotate, and can be respectively represented by two torsion angles ϕ and ψ. These angles can only appear in certain combinations due to steric hindrances, which define the allowed regions of the torsion angles for secondary structures in the plot

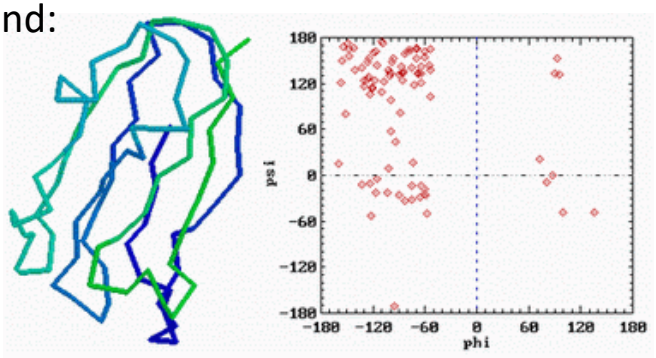




α-helix:



β-strand:



http://mod.life.nthu.edu.tw/protein/PTRC/Teaching/Protein/rp.htm

G. Ramachandran, C. Venkatachalam, and S. Krimm, Biophys. J. 6, 849-872 (1966).

- The Ramachandran plot is a practical criterion for improving the quality of NMR or crystallographic protein structures.
- However, correct torsion angle distributions do not imply the validity of 3D structures.

Protein packing density

Biophysical Journal Volume 81 August 2001 751–766

Are Proteins Well-Packed?

Jie Liang* and Ken A. Dill†
*Department of Bioengineering, University of Illinois at Chicago, Chicago, Illinois 60607-7052 and †Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-1204

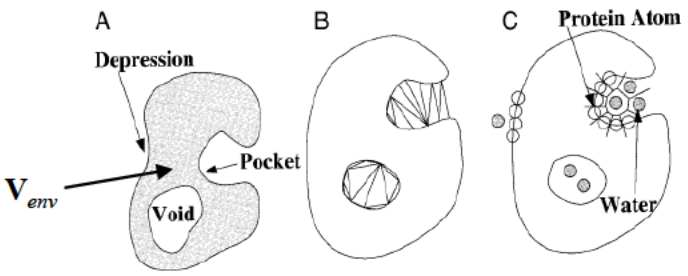


FIGURE 1 The concave regions and surface packing of proteins. (*a*) There are three types of concave regions on protein surfaces: Fully enclosed voids with no outlet, pockets accessible from the outside but with constriction at mouths, and shallow depressions. (*b*) In the analysis of protein self packing, the volume of interior voids and surface pockets are calculated based on Delaunay triangulation and alpha shape described later. All water molecules need to be removed first. (*c*) Protein–water packing has been studied by measuring the volume of Voronoi polyhedra of water molecules and neighboring protein atoms (Harpaz et al., 1994; Gerstein et al., 1995; Gerstein and Chothia, 1996). Here water molecules define the Voronoi volume of protein surface atoms.

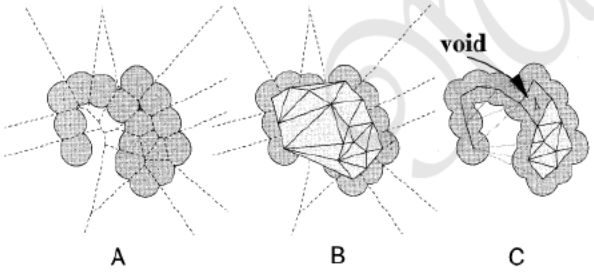
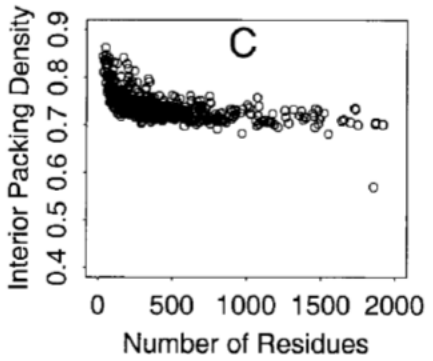


FIGURE 2 Geometry of a simplified 2D model molecule to illustrate the procedure mapping the Voronoi diagram to the Delaunay triangulation. (*a*) The molecule formed by the union of atom disks of uniform size. Voronoi diagram is in dashed lines. (*b*) The shape enclosed by the boundary polygon is the convex hull. It is tessellated by the Delaunay triangulation. (*c*) The alpha shape of the molecule is formed by removing those Delaunay edges and triangles that are not completely contained within the molecule. A molecular void is represented in the alpha shape by two empty triangles.

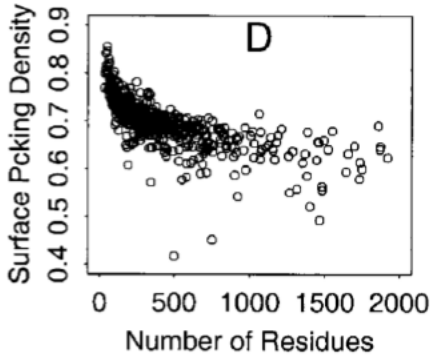
Internal packing density

$$P_i = \frac{V_{vdw}}{V_{env}} = \frac{V_{vdw}}{V_{void} + V_{ms}}$$



Surface packing density

$$P_s = \frac{V_{vdw}}{V_{pocket} + V_{ms}}$$



Larger proteins are packed more loosely than smaller proteins.

Criteria for 3D structures

- Liang and Dill have reported that the protein packing is heterogeneous, and in terms of packing density, protein molecules may be either well-packed or loosely packed.
  (J. Liang and K. A. Dill, Biophys. J. 81, 751 (2001).)

- Zhang et al. showed that the packing density of single domain proteins decreases with chain length, which shares a generic feature of random polymers satisfying loose constraint in compactness.
  (J. Zhang, et al., J. Chem. Phys. 118, 6102 (2003).)

- Protein packing density is not a universal parameter!

- **Question:** Is there a universal parameter for protein conformations?

ARVO, CAVE, ARVO-CL, CAVE-CL packages

These packages are designed to calculate the solvent accessible surface area and volume of overlapping spheres via analytic equations.


Available online at www.sciencedirect.com
SCIENCE DIRECT
Computer Physics Communications 165 (2005) 59–96
Computer Physics Communications
www.elsevier.com/locate/cpc

ARVO: A Fortran package for computing the solvent accessible surface area and the excluded volume of overlapping spheres via analytic equations ☆

Ján Buša [a,b,c], Jozef Džurina [b,c], Edik Hayryan [a,c], Shura Hayryan [a], Chin-Kun Hu [a,*], Ján Plavka [b,c], Imrich Pokorný [b,c], Jaroslav Skřivánek [b,c], Ming-Chya Wu [a]

[a] Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
[b] Department of Applied Mathematics, Technical University in Košice, 040 01 Košice, Slovak Republic
[c] Laboratory of Computing Technology and Automation, Joint Institute for Nuclear Research, Dubna, Russia

Comp. Phys. Comm. 165, 59-96 (2005).

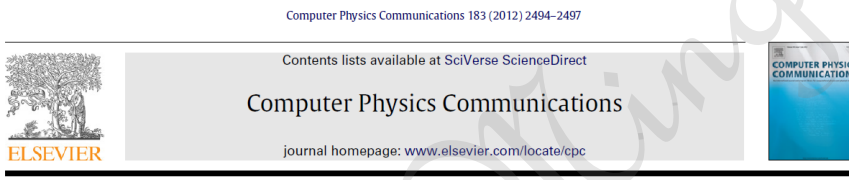

Computer Physics Communications 181 (2010) 2116–2125
Contents lists available at ScienceDirect
Computer Physics Communications
www.elsevier.com/locate/cpc

CAVE: A package for detection and quantitative analysis of internal cavities in a system of overlapping balls: Application to proteins ☆

Ján Buša [a,b,c], Shura Hayryan [a], Chin-Kun Hu [a,d,*], Jaroslav Skřivánek [a,e], Ming-Chya Wu [a,f,g]

[a] Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
[b] Technical University in Košice, 040 01 Košice, Slovak Republic
[c] Joint Institute for Nuclear Research, Dubna, Russia
[d] Center for Nonlinear and Complex Systems and Department of Physics, Chung Yuan Christian University, Chungli 32023, Taiwan
[e] SORS Research, 04001 Košice, Slovak Republic
[f] Research Center for Adaptive Data Analysis, National Central University, Chungli 32001, Taiwan
[g] Department of Physics, National Central University, Chungli 32001, Taiwan

J Comp. Chem. 30, 346-357 (2009);
Comp. Phys. Comm. 181, 2116-2125 (2010).


Computer Physics Communications 183 (2012) 2494–2497
Contents lists available at SciVerse ScienceDirect
Computer Physics Communications
journal homepage: www.elsevier.com/locate/cpc

ARVO-CL: The OpenCL version of the ARVO package — An efficient tool for computing the accessible surface area and the excluded volume of proteins via analytical equations ☆

Ján Buša Jr. [a,b,c], Shura Hayryan [a], Ming-Chya Wu [a,d,e], Ján Buša [b], Chin-Kun Hu [a,*]

[a] Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
[b] Faculty of Electrical Engineering and Informatics, Technical University of Košice, 040 01 Košice, Slovak Republic
[c] FURT Solutions, s.r.o., Strojárenská 3, 040 01 Košice, Slovak Republic
[d] Research Center for Adaptive Data Analysis, National Central University, Chungli 32001, Taiwan
[e] Department of Physics, National Central University, Chungli 32001, Taiwan

Comp. Phys. Comm. 183, 2494-2497 (2012).


Computer Physics Communications 190 (2015) 224–227
Contents lists available at ScienceDirect
Computer Physics Communications
journal homepage: www.elsevier.com/locate/cpc

CAVE-CL: An OpenCL version of the package for detection and quantitative analysis of internal cavities in a system of overlapping balls: Application to proteins

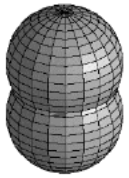
Ján Buša Jr. [a,b], Ján Buša [b], Shura Hayryan [a], Chin-Kun Hu [a,*], Ming-Chya Wu [a,c]

[a] Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
[b] Faculty of Electrical Engineering and Informatics, Technical University of Košice, 040 01 Košice, Slovak Republic
[c] Research Center for Adaptive Data Analysis, National Central University, Chungli 32001, Taiwan

Comp. Phys. Comm. 190, 224-227 (2015).

ARVO

The file input.dat consists of two lines

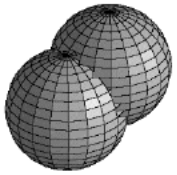


0 0 0 2 - 1st sphere: center at (0,0,0) and radius=2
0 0 2 2 - 2nd sphere: center at (0,0,2) and radius=2

The output.dat file read as:

Volume: 56.5486678 Area: 75.3982237 Spheres num: 2

The file input.dat contains two lines



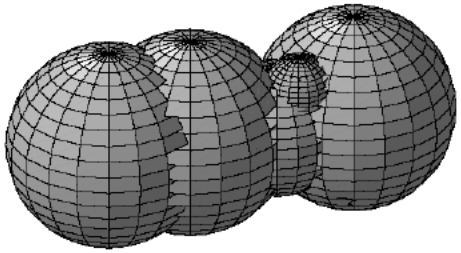
0 0 0 2
0 1.73205080756888 1 2

The screen output af er th command: "mpirun -np 2 parvo" was

Rotation after bad Nord Pole test!
Processor 0 results: V: 26.1654833 A: 37.6991118
Processor 1 results: V: 30.3831845 A: 37.6991118
Volume: 56.5486678 Area: 75.3982237 Spheres num: 2

The input.dat file with 8 spheres:

0 0 0 2
-5 0 0 6
5 0 0 6
0 1 0 4
-1 2 3 2
1 1 1 1
0 0 1 2
10 0 0 6

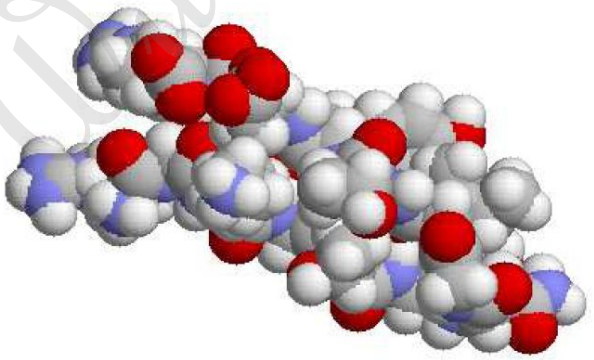


The screen output after the command: "mpirun -np 5 parvo" was

Processor 0 results: V: 852.520233 A: 389.756652
Processor 3 results: V: 0. A: 0.
Processor 4 results: V: 718.770219 A: 320.442451
Processor 2 results: V: 36.6247159 A: 13.114286
Processor 1 results: V: 722.019662 A: 288.559143
Volume: 2329.93483 Area: 1011.87253 Spheres num: 8

ARVO

ARVO can calculate solvent inaccessible volume and solvent accessible surface area in a very accurate way.



14-residue peptide rgkwtyngityegr

Input file is obtained from the PDB code 1j4m.pdb

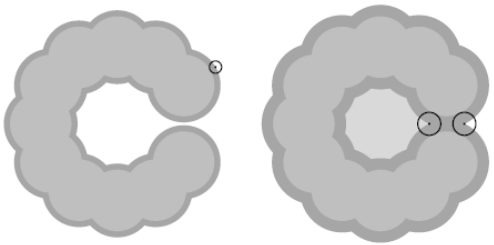
The screen output after the command: "mpirun -np 8 parvo" reads as:

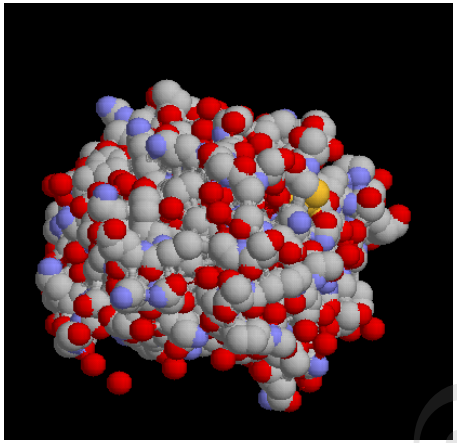
```
Processor 6  results: V:   138.191267 A:   170.402091
Processor 7  results: V:   139.91866  A:   165.475971
Processor 0  results: V:   127.595148 A:   169.080939
Processor 1  results: V:   115.924238 A:   171.063623
Processor 5  results: V:   111.994168 A:   176.278727
Processor 4  results: V:   140.388721 A:   183.673615
Processor 2  results: V:   157.553423 A:   173.478737
Processor 3  results: V:   110.807138 A:   151.68257
```

Volume: 1042.37276 Area: 1361.13627 Spheres num: 121

CAVE

CAVE can detect the cavities in a molecule and calculate solvent inaccessible volume and solvent accessible surface area.





```
The probe radius 1.20000005
The protein name 2ptn
        Total number of atoms: 1629 allocated: 20000
        Rashin set
Total number of neighbors: 62054
Neighbors total: 62054 allocated: 2000000
Maximum: 66 allocated: 200
Number of the preliminary triangles: 2236
Number of vertices: 2278
2197 active triangles before lugs cutting. After lugs cutting: 2197
                1.  south: 1443 1444 1453
```

```
340 internal points
Clusters: 26 Re ation : 11
1. segment conta  s 17   vities.
11 triangles were d leted from the list IT.
326 internal triangles were deleted from the list IT.
340 internal triangles vertices!
1. segment contains cavity!
Triangles number after 1. segment: 0
Number of s gments:
Total: 1 Deleted: 0 Cavity enveloping: 1
1. ca ity segment contains 340 triangles.
To al n mber of cavities is 17.
Total number of triangles is 340.
Total number of spheres is 203.

Number of atoms: 1629
There exists cavity!
Inaccessible volume: 37436.2361
Accessible surface area: 9315.41184
Cavities volume: 32.4669432
Surface area of cavities: 181.732348
1. boundary volume: 563.196503
1. boundary surface area: 389.741749
1. cavity volume: 0.0590741273
1. cavity surface area: 1.10512934
2. boundary volume: 515.125257
2. boundary surface area: 375.817426
2. cavity volume: 0.00464992906
2. cavity surface area: 0.194346908
...
16. boundary volume: 1628.63036
16. boundary surface area: 862.982504
16. cavity volume: 6.12086952
16. cavity surface area: 36.522405
17. boundary volume: 614.397572
17. boundary surface area: 423.464562
17. cavity volume: 0.0684193718
17. cavity surface area: 1.43204525
STOP End statement executed
```

Scaling in volume and surface area

- **Definition**

  $V(r_p, N)$: van der Waals volume (unit: Å^3)

  $A(r_p, N)$: solvent accessible surface area (unit: Å^2)

  r_p : radius of the probe sphere (unit: Å)

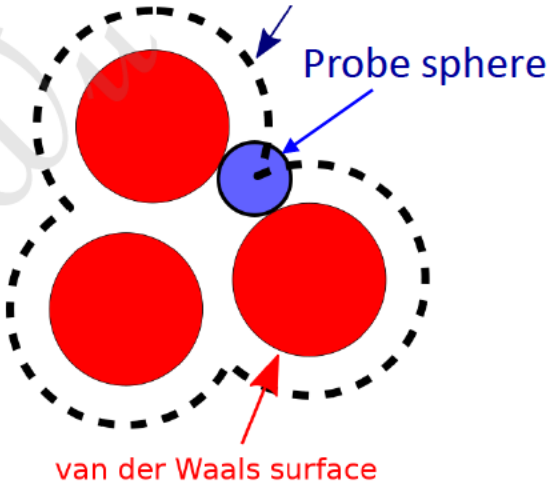
  N: number of atoms in the protein

- $r_N = r_C = 1.55$Å; $r_S = 2.00$Å; $r_O = 1.40$Å; $r_p = 0.0$Å

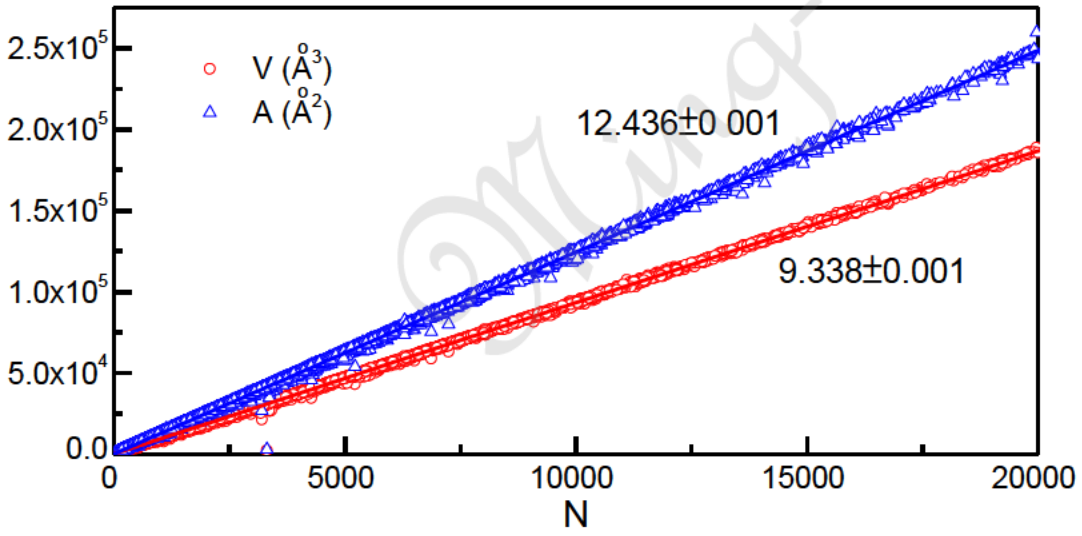
- Using ARVO package for the calculation.

  [ARVO package: J. Busa, et al. Comp. Phys. Comm. 165, 59-96 (2005).]

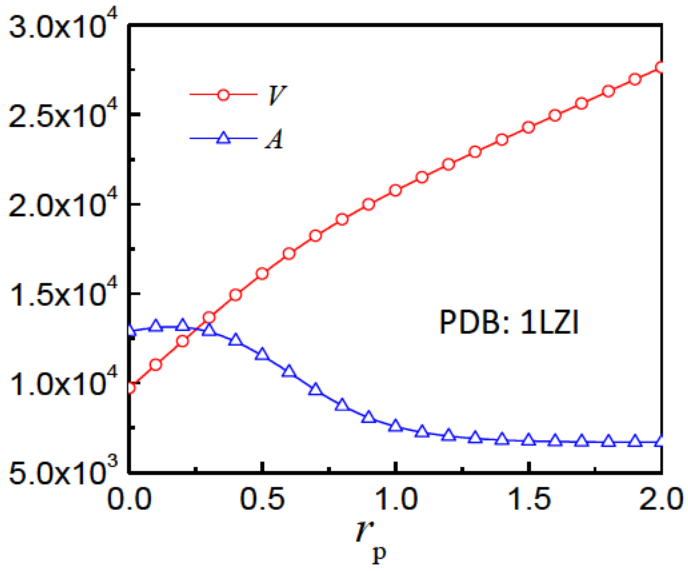
Solvent accessible surface



http://en.wikipedia.org/wiki/Accessible_surface_area



Dependence of V and A on N for 28,664 PDB protein structures. The numbers indicate the slopes.



Dependence of V and A on r_p

V/A<r> Ratio

■ **Definition:** $$R(r_p) = \frac{V(r_p, N)}{A(r_p, N) \cdot \langle r \rangle}$$

V: van der Waals volume (unit: Å^3)

A: solvent accessible surface area (unit: Å^2)

r_p: radius of the probe sphere (unit: Å)

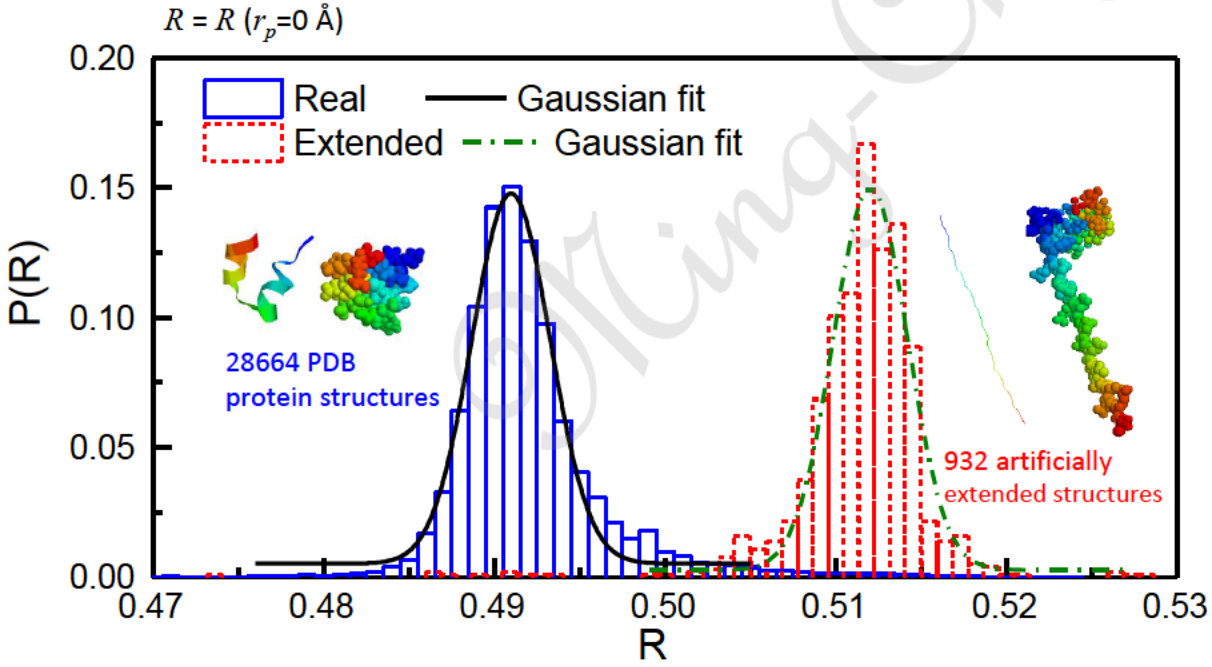
N: number of atoms in the protein

<r>: average radius of atoms in the protein (unit: Å)

$r_N = r_C = 1.55$Å; $r_S = 2.00$Å;

$r_O = 1.40$Å; $r_p = 1.40$Å

$R = R\ (r_p{=}0$ Å) $R_s = R\ (r_p = 1.4$ Å)



Gaussian fit:

$$y = y_0 + \frac{S}{w\sqrt{\pi/2}} exp\left[-\frac{2(x - x_c)^2}{w^2}\right]$$

y_0	x_c	w	S	P
0.0053	0.4910	0.0046	0.0008	0.9840

Maxima locate at:

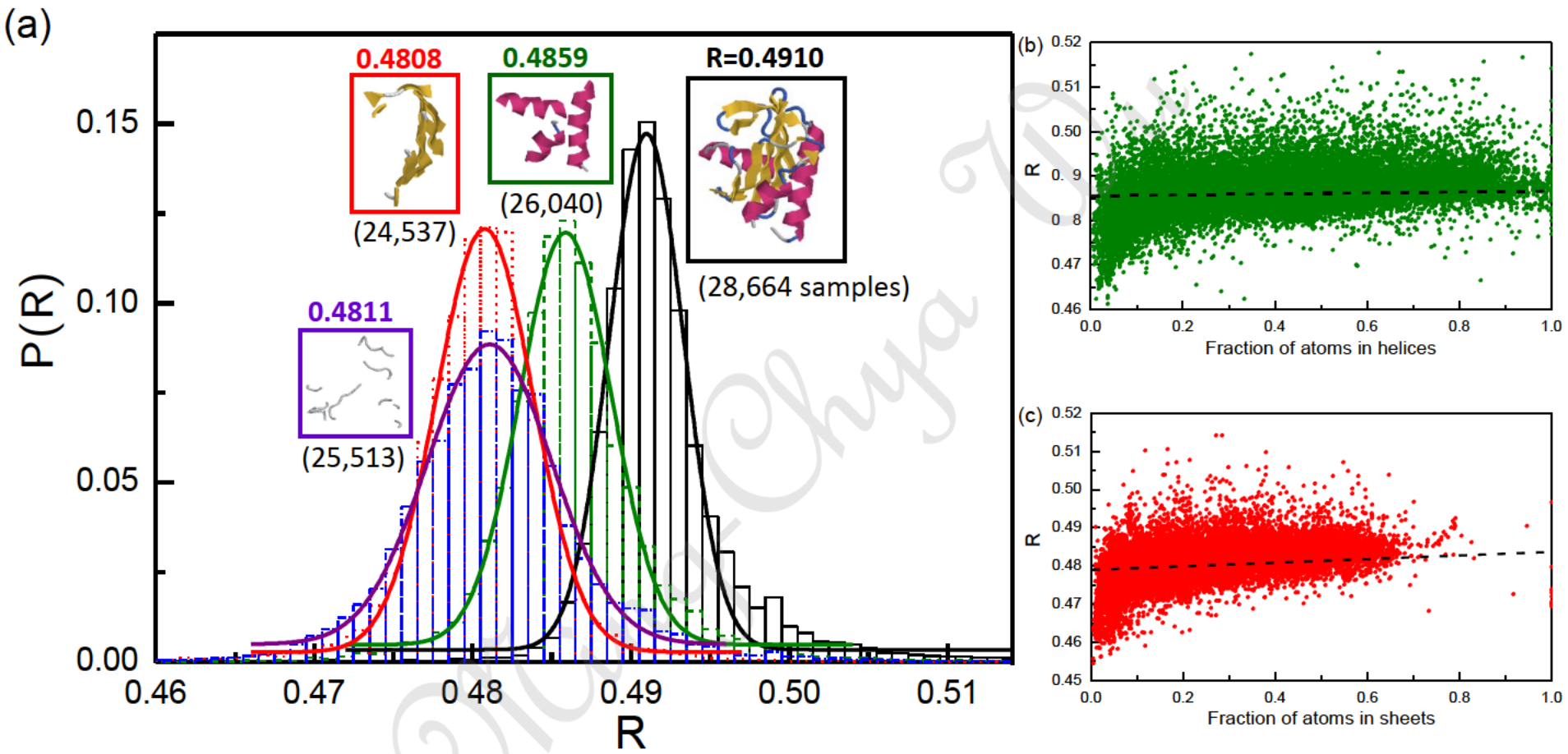
$R \cong 0.4910$ *(real structure)*

$R \cong 0.5120$ *(extended structure)*

Remark: $R = 0.333$ *for distinct spheres*

M.-C. Wu, M. S. Li, W.-J. Ma, M. Kouza, and C.-K. Hu, EPL 96, 68005 (2011).

Statistics on secondary structures



(a) Probability density function of R statistics for whole protein molecules, helix, sheet, and other structures.

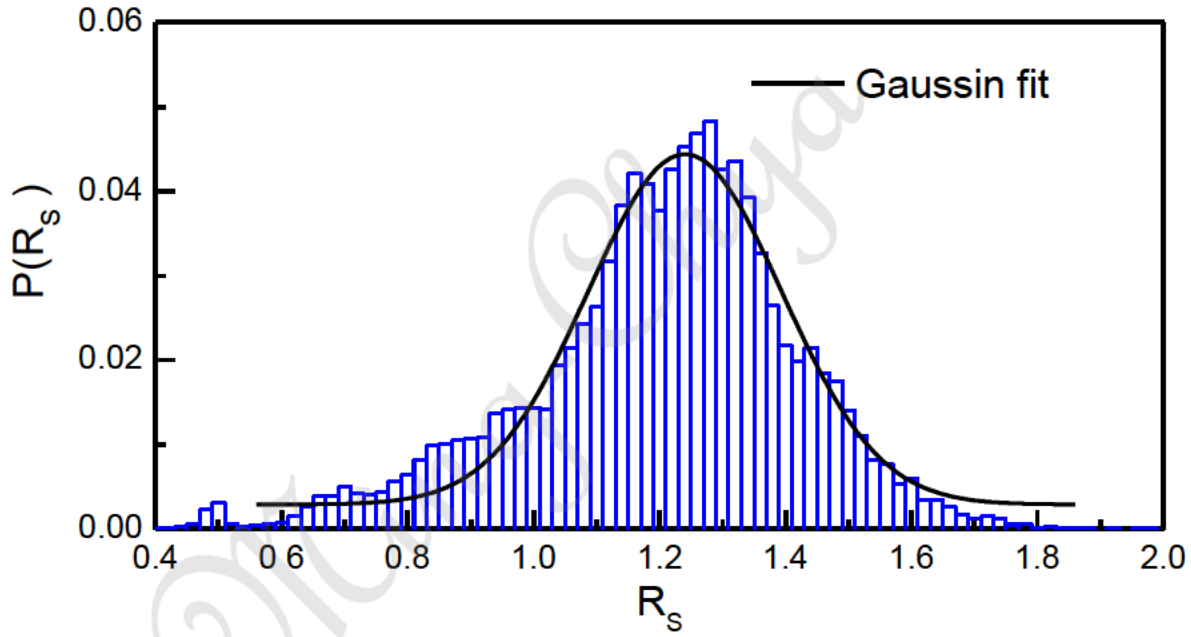
(b) R as a function of the fraction of atoms in helix structures. The slope of the linear fit is 0.0001, and the correlation level is 0.005 (very low).

(c) R as a function of the fraction of atoms in sheet structures. The slope of the linear is 0.0002, and the correlation level is 0.005 (very low).

V/A<r> Ratio for $r_p = 1.4$Å

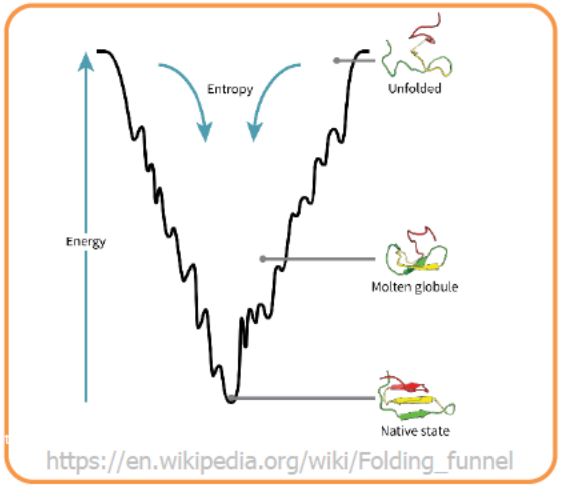
$$R_S = R(r_p = 1.4) = \frac{V}{A \cdot \langle r \rangle}$$

V: solvent inaccessible volume (unit: Å^3)
A: solvent accessible surface area (unit: Å^2)
r_p: radius of the probe sphere (unit: Å)
<r>: average radius of atoms in the protein (unit: Å)



- The distribution of R_s for 28,236 proteins using the probe sphere with radius of 1.4Å. The maximum of the Gaussian fit is located at $R_s \approx 1.2402$.
- As a universal parameter, R is more useful than R_s.

R and Energy Minimization



https://en.wikipedia.org/wiki/Folding_funnel

Folding simulation using the SMMP package

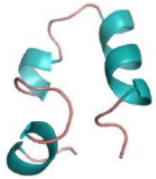
F. Eisenmenger, et.al., Comput. Phys. Commun. 138, 192-212(2001).

Randomly chosen configuration

with simulated annealing

with canonical Monte Carlo

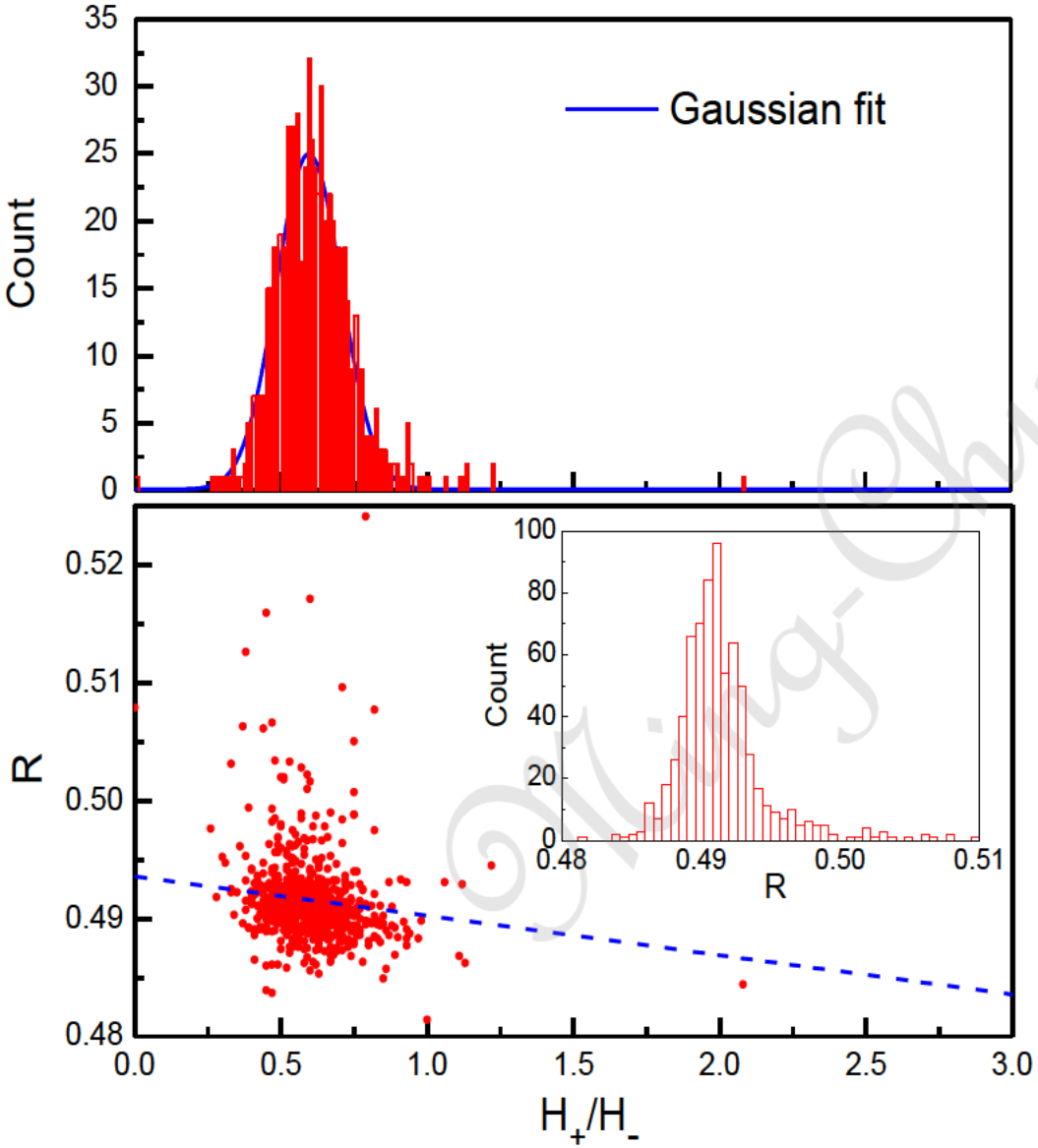
PDB structure



PDB	N	$R_{(a)}'$	$R_{(a)}''$	$R_{(c)}''$	R
1HP9	183	0.5694	0.4912	0.4875	0.5057
1KDL	193	0.5364	0.4867	0.4863	0.4998
1GCN	246	0.5480	0.4917	0.4875	0.4946
1VII	295	0.5560	0.4876	0.4878	0.5058
2PLH	330	0.6280	0.4864	0.4857	0.4954
2OVO	418	0.5549	0.4981	0.4891	0.4933
1PGB	436	0.5385	0.4866	0.4878	0.4891
1HPT	440	0.5726	0.4879	0.4871	0.4928
1UOY	452	0.5414	0.4896	0.4872	0.4945
1UTG	548	0.5785	0.4780	0.4858	0.4896

R = 0.491 should be somewhat related to the energy global minimum of protein molecules.

V/A<r> and hydrophobicity of amino acids



The distribution of the ratio of hydrophilic (H_+) and hydrophobic (H_-) amino acids in a molecule for 723 protein structures (from Protein Sequence Culling Server), based on the Kyte-Doolittle scale.

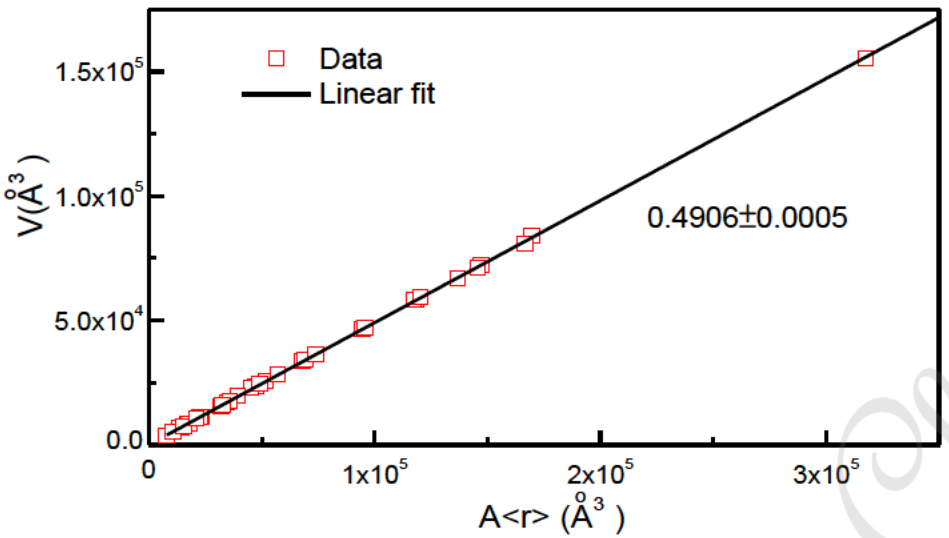
- The Gaussian fit is centered at 0.594.

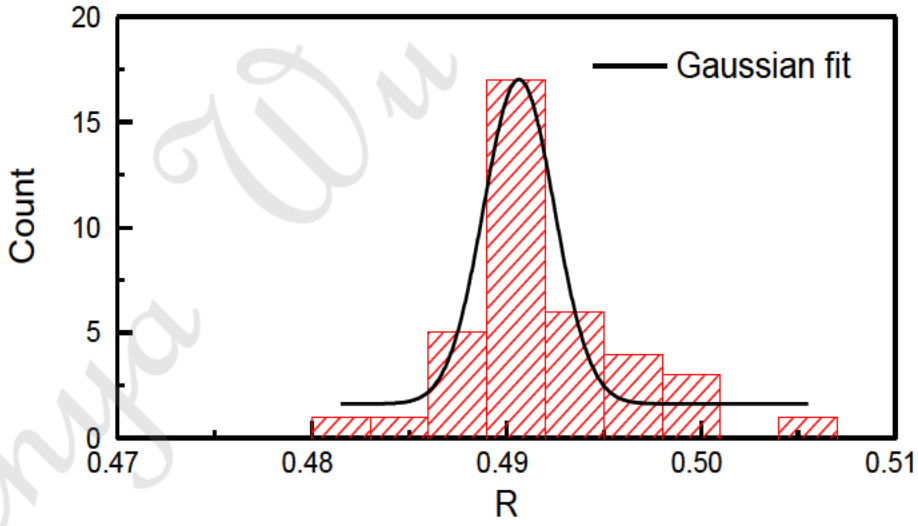
R as a function of H_+/H_-.

- The slope of the linear fit (blue dashed line) is −0.0033, and the correlation level is 0.2 (notably lower than 0.5, indicating no correlation between them).
- The inset shows the distribution of R for the 723 protein structures.

M.-C. Wu, M. S. Li, W.-J. Ma, M. Kouza, and C.-K. Hu, EPL 96, 68005 (2011).

V/A<r> for intrinsically disordered proteins





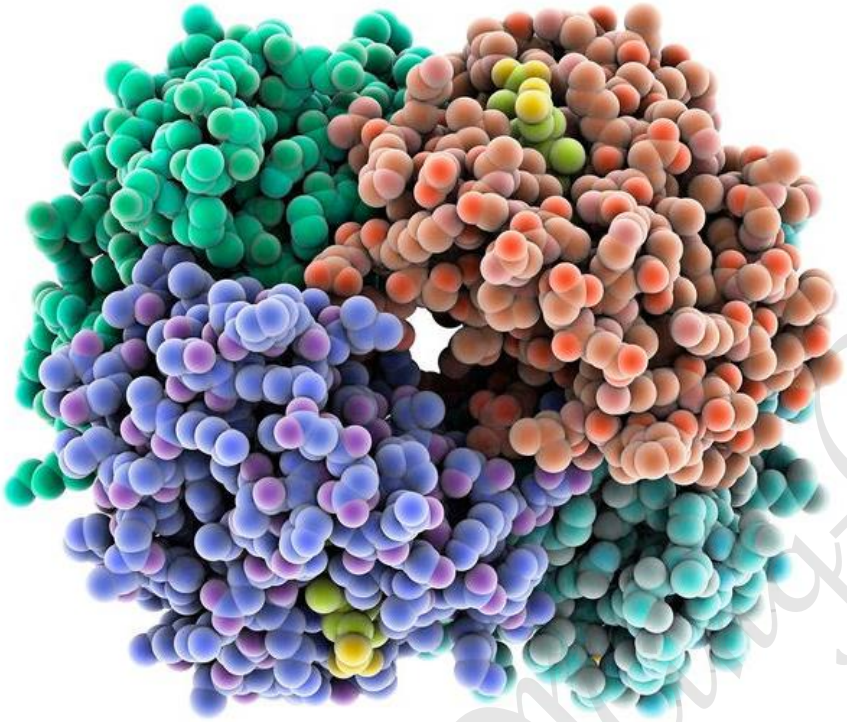
- Statistics of R for 38 intrinsically disordered protein structures shows that the ratio is 0.4906 ± 0.005, which is within the tolerance determined by the ensemble of 28,664 PDB protein structures.

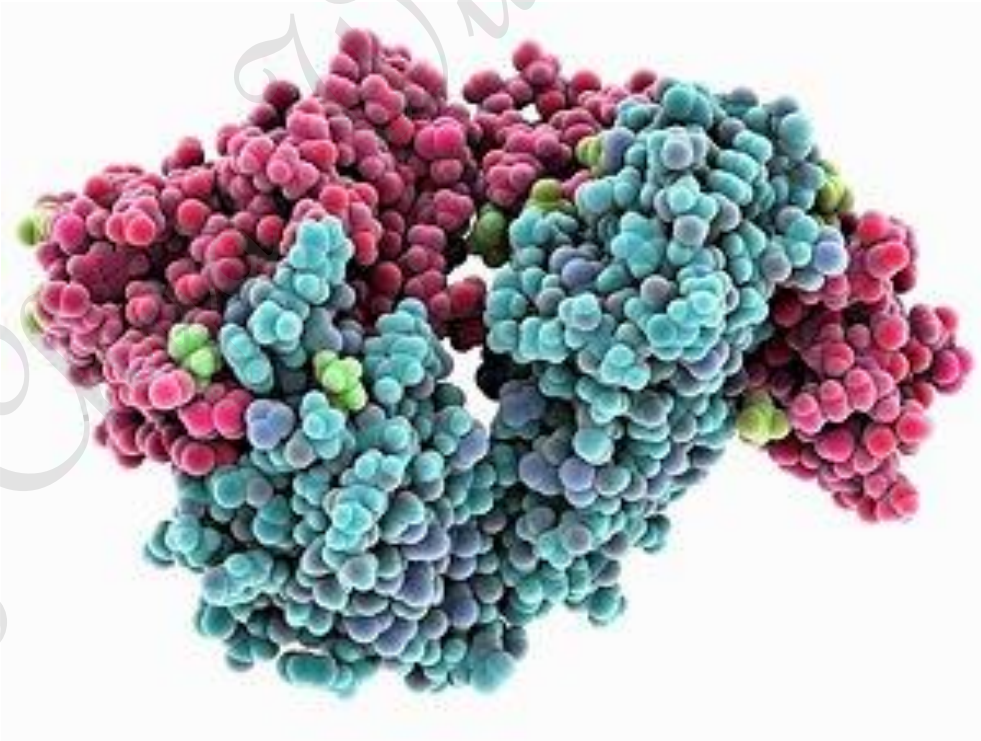
The 38 intrinsically disordered protein structures are available in B. Mészáros, et al., J. Mol. Biol. 372, 549 (2007).

- The ratio holds once a polypeptide folds to a compact structure no matter of its species.

M.-C. Wu, M. S. Li, W.-J. Ma, M. Kouza, and C.-K. Hu, EPL 96, 68005 (2011).

Proteins with multiple chains

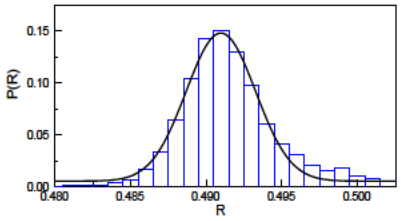




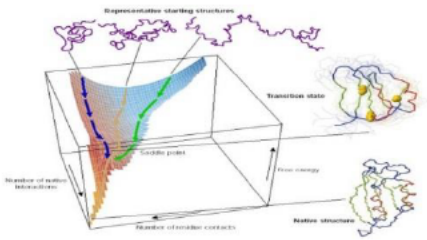
https://www.thoughtco.com/protein-function-373550

- $R = 0.491$ holds for proteins with multiple chains or complexes.

Conclusions



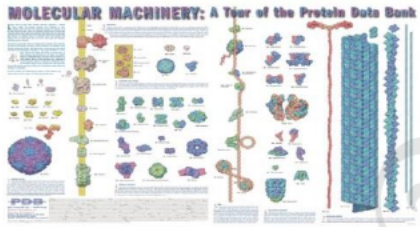
- $R(r_p = 0) = 0.491 \pm 0.005$ is universal for native protein structures, including intrinsically disordered proteins.



http://www.nature.com/nature/journal/v426/n6968/fig_tab/nature02261_F1.html

- $R = 0.491$ should be somewhat related to the energy global minimum of protein molecules.

- The tube model and sheet model seem potential models for the ratio: $R = 0.491$.



Protein Data Bank, http://www.rcsb.org/pdb/home/home.do

- $R = 0.491$ can serve as a necessary condition for checking the validity of PDB data and designing protein-like sequences.

http://h10010.www1.hp.com/wwpc/tw/zh/
http://www.princeton.edu/cbe/news/archive/?id=2360

- $R = 0.491$ does not correspond to a unique conformation, but it confines molecular conformations in a folding simulation from vast possibilities to a smaller space. It may be useful in protein folding simulations or in structure identification experiments.

Thanks for your attention.