



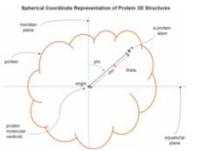
Representing Protein 3D Structures in Spherical Coordinates: Two Applications:

(1.) Detection of Invaginations and Protrusions on the Protein Surface; and

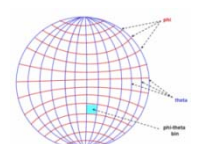
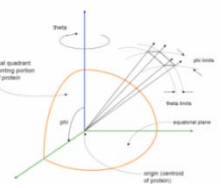
(2.) Separating the Protein Hydrophobic Inner core from the Hydrophilic Outer layer

Vicente M. Reyes, Ph.D., Assistant Professor [E-mail: vmrsbi@rit.edu]
 Department of Biological Sciences, School of Life Sciences, College of Science,
 Rochester Institute of Technology, Rochester, NY 14623-5603

(1.) Introducing the Spherical Coordinate System

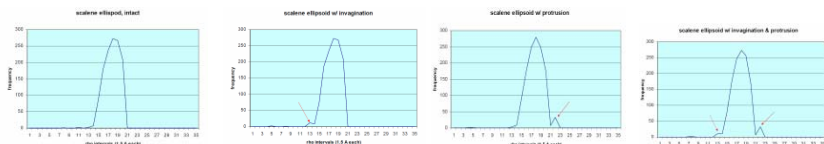
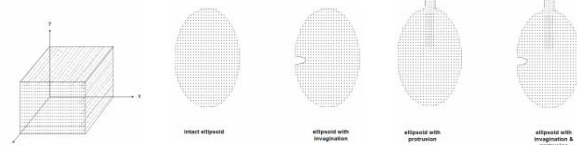


Phi and Theta Binning of a Protein in Spherical Coordinates



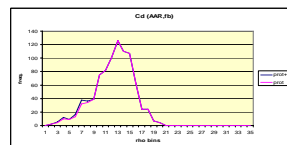
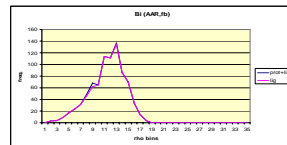
The Theoretical Case:
 Artificial Protein in the form of a Scalene Ellipsoid Grid of Points

Construction of Artificial Protein

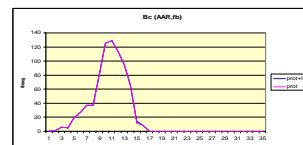
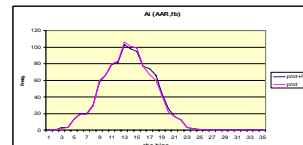


The Results (1.):

Group 1:
 bumps on lagging side of main peak

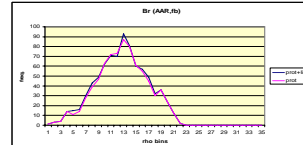
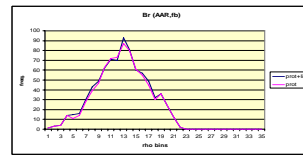


Group 2:
 bumps on leading side of main peak

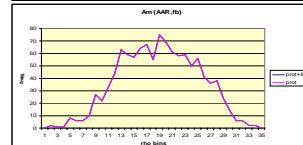
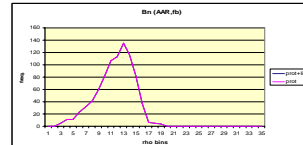


The Results (1.) [cont'd.]:

Group 3:
 bumps on both lagging and leading sides of main peak



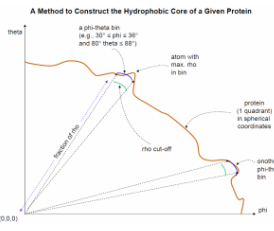
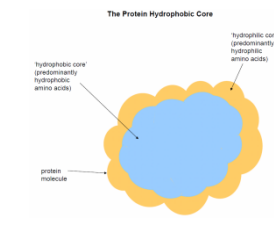
Group 4:
 ligand has no / negligible effect on protein overall shape



The Test Set: Laskowski et al., *Proteins*, 1996

Structure PDB ID	Shortname designation	LIG # or LIG #	Ligand Name
1A20	1A20	1	ATP
1A21	1A21	1	ATP
1A22	1A22	1	ATP
1A23	1A23	1	ATP
1A24	1A24	1	ATP
1A25	1A25	1	ATP
1A26	1A26	1	ATP
1A27	1A27	1	ATP
1A28	1A28	1	ATP
1A29	1A29	1	ATP
1A30	1A30	1	ATP
1A31	1A31	1	ATP
1A32	1A32	1	ATP
1A33	1A33	1	ATP
1A34	1A34	1	ATP
1A35	1A35	1	ATP
1A36	1A36	1	ATP
1A37	1A37	1	ATP
1A38	1A38	1	ATP
1A39	1A39	1	ATP
1A40	1A40	1	ATP
1A41	1A41	1	ATP
1A42	1A42	1	ATP
1A43	1A43	1	ATP
1A44	1A44	1	ATP
1A45	1A45	1	ATP
1A46	1A46	1	ATP
1A47	1A47	1	ATP
1A48	1A48	1	ATP
1A49	1A49	1	ATP
1A50	1A50	1	ATP
1A51	1A51	1	ATP
1A52	1A52	1	ATP
1A53	1A53	1	ATP
1A54	1A54	1	ATP
1A55	1A55	1	ATP
1A56	1A56	1	ATP
1A57	1A57	1	ATP
1A58	1A58	1	ATP
1A59	1A59	1	ATP
1A60	1A60	1	ATP
1A61	1A61	1	ATP
1A62	1A62	1	ATP
1A63	1A63	1	ATP
1A64	1A64	1	ATP
1A65	1A65	1	ATP
1A66	1A66	1	ATP
1A67	1A67	1	ATP
1A68	1A68	1	ATP
1A69	1A69	1	ATP
1A70	1A70	1	ATP
1A71	1A71	1	ATP
1A72	1A72	1	ATP
1A73	1A73	1	ATP
1A74	1A74	1	ATP
1A75	1A75	1	ATP
1A76	1A76	1	ATP
1A77	1A77	1	ATP
1A78	1A78	1	ATP
1A79	1A79	1	ATP
1A80	1A80	1	ATP
1A81	1A81	1	ATP
1A82	1A82	1	ATP
1A83	1A83	1	ATP
1A84	1A84	1	ATP
1A85	1A85	1	ATP
1A86	1A86	1	ATP
1A87	1A87	1	ATP
1A88	1A88	1	ATP
1A89	1A89	1	ATP
1A90	1A90	1	ATP
1A91	1A91	1	ATP
1A92	1A92	1	ATP
1A93	1A93	1	ATP
1A94	1A94	1	ATP
1A95	1A95	1	ATP
1A96	1A96	1	ATP
1A97	1A97	1	ATP
1A98	1A98	1	ATP
1A99	1A99	1	ATP
1A100	1A100	1	ATP

(2.) Hydrophobic Inner Core and Hydrophilic Outer Layer Of a Protein and How to Separate Them



The Results (2.):



Representing Protein 3D Structures in Spherical Coordinates -- Two Applications: 1. Detection of Invaginations, Protrusions and Potential Ligand Binding Sites; and 2. Separation of Protein Hydrophilic Outer Layer from the Hydrophobic Core

Vicente M. Reyes, Ph.D., Assistant Professor

Dept. of Biological Sciences, Rochester Institute of Technology, Rochester, NY 14623

A Fortran 90 program was written to convert a protein 3D structure PDB file in Cartesian coordinates to spherical coordinates (ρ , ϕ , θ), with the centroid (center of mass) of the protein molecule as origin. We investigated the utility of this representation in the (1.) detection of invaginations, protrusions and potential ligand binding sites (LBSs) on the protein surface, and (2.) separation of the hydrophilic outer layer (HOL) from the hydrophobic inner core (HIC) of protein molecules. The dataset of Laskowski et al., (Prot. Sci., 1996), composed of 67 single-chain protein structures, was used as test set in evaluating both applications. Both ϕ and θ are partitioned into suitable intervals (e.g., 6- and 8-degree intervals, respectively), giving rise to 1,350 ϕ - θ bins partitioning all of 3D space. The atom with maximum ρ in each ϕ - θ bin is sequestered. In the first application, this step is done in both the liganded and unliganded form of the query protein and the frequency distribution of the maximum ρ values from both forms are plotted superimposed on each other. Invaginations on the protein surface give rise to subpeaks or shoulders on the lagging side of the main peak, while protrusions give rise to similar shoulders, but on the leading side of the main peak. We find that most LBSs are associated with such subpeaks and therefore propose that such subpeaks are potential LBSs. In the second application, a suitable cutoff value for ρ , e.g., 0.95ρ , is adopted for each ϕ - θ bin: all atoms with ρ values less than this cutoff value are considered part of the HIC, and those with equal or greater ρ values part of the HOL. Except for a very few special cases, we show that all of the proteins in the Laskowski dataset, after undergoing our HOL-HIC separation procedure, give rise to an outer layer that is significantly more enriched in hydrophilic amino acid residues, an inner core that is significantly more enriched in hydrophobic amino acid residues. In addition, a quick but effective way of determining active sites in the HIC and protein-protein interaction (PPI) interfaces in the HOL was derived. Once the HIC and the HOL are separated, the HIC may be searched for His, Glu, Asp, or Cys residues as potential active sites, and the HOL searched for clusters of hydrophobic amino acid residues as potential PPI interfaces (data not shown). We conclude that spherical coordinate representation of protein structures is a useful alternative to Cartesian coordinate representation, and may well find other useful applications beyond the ones described here.