A Physics-Based Approach of Coarse-Graining the Cytoplasm of *Escherichia coli* (CGCYTO)

Qian Wang and Margaret S. Cheung

Department of Physics, University of Houston, Houston, Texas

Supporting Material

Method:

1. Clustering Method

Given a macromolecule, an atom i with coordinate of x_i is assigned into one of the clusters in iterative steps and there are C_1, C_2…C_N clusters. In the initial step, N is the number of atoms in a macromolecule and N may vary after each iteration. A cluster l has T atoms. The geometric center of C_l is $C_l = \frac{1}{l}\sum_{j=1}^{l} x_j$. In each iteration, a given atom i is assigned to cluster l if the Euclidean distance between x_i and the geometric center of the lth cluster C_l is less than D followed by an updated on C_l accordingly. All atoms are clustered iteratively into N clusters until there is no additional assignment of atoms among these clusters as C_l [l=1…N] converges.

2. Potential energy of coarse-grained macromolecular crowders

In order to maintain the shape of these macromolecular crowders in the simulation, bond strength, bond angle and dihedral energy were used.

The bond strength between bead i and bead j follows the formula below:

$$E_{ij}^{bond} = k_b(r_{ij} - r_{ij}^0)^2 \qquad \text{(S1)}$$

k_b = 250 kcal/mol. If bead i and j represent all-atomistic clusters A and B, respectively, then r_{ij}^0 is the distance between the geometry center of A and the geometry center of B.

The bond angle among three consecutive beads i, j and k has the following form:

$$E_{ijk}^{angle} = k_\theta(\theta_{ijk} - \theta_{ijk}^0)^2 \qquad \text{(S2)}$$

k_θ = 250 kcal/mol. If bead i, j and k represent all-atomistic clusters A, B and C, respectively, then θ_{ijk}^0 is the angle among the geometry center of A, geometry center of B and geometry center of C.

The dihedral energy for four consecutive beads *i, j, k,* and *l* has the following form:

$$E_{ijkl}^{dihedral} = \sum_{n=1,3} k_\varphi^n[1 - \cos(n \times (\varphi_{ijkl} - \varphi_{ijkl}^0))] \qquad \text{(S3)}$$

k_φ^1 = 1.2kcal/mol and k_φ^3 = 0.6kcal/mol. If i, j, k and l represent all-atomistic clusters A, B, C and D, respectively, then φ_{ijkl}^0 is the dihedral angle among the geometry center of A, geometry center of B, geometry center of C and geometry center of D.

3. Folding free energy calculation by the Scaled Particle Theory (SPT)

An analytical solution of the folding free energy difference ($\Delta\Delta G_{uf}$) between the bulk and crowded conditions is computed by the SPT with a mixture of spherical crowders. We have used two methods developed by Minton or Zhou: 1) $\Delta\Delta G_{uf}$ (SPT-Minton) satisfies the following formula [1],

$$\Delta\Delta G_{uf} = (1-\varphi)^{-1}\sum_i \varphi_i(3f_i + 3f_i^2 + f_i^3) + (1-\varphi)^{-2}\sum_i \varphi_i f_i \sum_i \varphi_i(9f_i/2 + 3f_i^2) + 3(1-\varphi)^{-3}(\sum_i \varphi_i f_i)^3$$
$$- (1-\varphi)^{-1}\sum_i \varphi_i(3U_i + 3U_i^2 + U_i^3) + (1-\varphi)^{-2}\sum_i \varphi_i U_i \sum_i \varphi_i(9U_i/2 + 3U_i^2) + 3(1-\varphi)^{-3}(\sum_i \varphi_i U_i)^3 \quad \text{(S4)}$$

$$f_i = \sqrt{5/3}R_g(fold)/r_i \quad \text{(S5)}$$

$$U_i = \sqrt{5/3}R_g(unfold)/r_i \quad \text{(S6)}$$

2) $\Delta\Delta G_{uf}$ (SPT-Zhou) satisfies the following formula [2],

$$\Delta\Delta G_{uf} = (1-\varphi)^{-1}\sum_i \varphi_i(3f_i + 3f_i^2 + f_i^3) + (1-\varphi)^{-2}\sum_i \varphi_i f_i \sum_i \varphi_i(9f_i/2 + 3f_i^2) + 3(1-\varphi)^{-3}(\sum_i \varphi_i f_i)^3$$
$$- 3\sum_i \varphi_i U_i^2(1 + 2/\sqrt{\pi}U_i) - 9\sum_{i,j} \varphi_i\varphi_j U_i U_j \ln(U_{i,j>}) \quad \text{(S7)}$$

$$f_i = R_g(fold)/r_i \quad \text{(S8)}$$

$$U_i = R_g(unfold)/r_i \quad \text{(S9)}$$

φ_i is the volume of fraction of crowders for the *i*th type of crowder with radius r_i. *i* ranges from 1 to 50. R_g (fold) is the radius of gyration of apoazurin in the folded state and R_g (unfold) is the radius of gyration of apoazurin in the unfolded state. In this work, R_g (fold) = 15.1Å and R_g (unfold) = 35.1Å. These values were obtained from our coarse-grained molecular simulation at $1.32k_BT/\varepsilon$, the folding temperature T_f of apoazurin in the bulk condition.

4. Coarse-grained modeling of a protein

A side chain C_α model (SCM) [3] that includes two beads per amino acid (except glycine) was used to represent the structure of testing protein (apoazurin). The crystal structure of apoazurin was obtained from the Protein Data Bank (PDB:1E65). Apoazurin is a two-state fast folding protein in experiments[4]; therefore the use of structural-based model is adequate for modeling the folding of apoazurin. A detailed description of the structure-based (Go-type[5]) Hamiltonian can be found in a previous study[6]. Apoazurin was placed in a periodic cubic box and the length of the box is 1140Å. Crowder models were randomly assigned into the box. In addition, considering the heterogeneity of the crowders in the HS and PD model, 10 independent simulations were each performed where an apoazurin protein was randomly inserted into ten different void positions in a cytoplasmic model.

5. Computation of the local volume available for apoazurin in the presence of crowders

A void is formed by the fluctuation of particles in a crowded system. We computed a parameter, named "local free volume" $< LV_{free} >$ that measures the size of such void available to an apoazurin in a crowded medium. $< LV_{free} >$ is computed by the particle insertion method of Widom [7] shown in Figure S1: First, in a crowded solution, a defined sphere Z_A with radius of 114 Å is centered at the center of mass (COM) of an apoazurin. A hard sphere with a radius of

45.3Å (equals to $\sqrt{5/3}$ times 35.1Å, the radius of gyration of apoazurin in the unfolded state [1]) is randomly inserted into Z_A for N times. Note that the radius of 45.3Å is the same value as the one we have used for the unfolded state in the abovementioned SPT calculations. An attempted insertion of a testing hard sphere is accepted only when its volume does not overlap with other crowders' in Z_A. Set volume of Z_A to be $V=(4\pi/3)\times 114^3$ Å^3. If an attempt of insertion is accepted for N' times, then $< LV_{free} > = N'V/N$. <> represents an ensemble average.

REFERENCE:

1. Minton, A.P., *Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: Macromolecular crowding and protein stability revisited* Biophys. J., 2005. **88**: p. 971-985.
2. Zhou, H.-X., *Effect of mixed macromolecular crowding agents on protein folding*. Proteins, 2008. **72**: p. 1109-1113.
3. Cheung, M.S., et al., *Exploring the interplay of topology and secondary structural formation in the protein folding problem*. J. Phys. Chem. B, 2003. **107**: p. 11193-11200.
4. Chen, M., et al., *Correlation between protein stability cores and protein folding kinetics: a case study on Pseudomonas aeruginosa apoazurin*. Structure, 2006. **14**: p. 1401-1410.
5. Taketomi, H., Y. Ueda, and N. Go, *Studies on Protein Folding, Unfolding and Fluctuations by Computer-Simulation .1. Effect of Specific Amino-Acid Sequence Represented by Specific Inter-Unit Interactions*. Int J Pept Protein Res, 1975. **7**: p. 445-459.
6. Christiansen, A., et al., *Factors defining effects of macromolecular crowding on protein stability: an in vitro/in silico case study using cytochrome c*. Biochemistry, 2010. **49**: p. 6519-6530.
7. Widom, B., *Some topics in the theory of fluids*. J. Chem. Phys., 1963. **39**: p. 2808-2812.
8. Chodera, J.D., et al., *Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations*. J. Chem. Theory Comput., 2007. **3**: p. 26-41.

Table S1. Structural characteristics for 50 kinds of macromolecules. See the definitions of V_p, R_{eff}, Δ, S, $V_{co\text{-}macro}$, $V_{co\text{-}sphere}$, and λ in the Method Section.

Name	V_p (Å^3)	R_{eff} (Å)	Δ	S	$V_{co\text{-}macro}$ (Å^3)	$V_{co\text{-}sphere}$ (Å^3)	λ
Ppa	223386	37.6	0	0	624901	576847.57	0.92
GapA	268037	40	0	0	739953	659501	0.89
Adk	44674	22	0.02	0	214224	195624.49	0.91
Pyr	641323	53.5	0.02	-0.01	1488390	1288078	0.87
CspC	13928	14.9	0.02	-0	110300	101376.82	0.92
Pnp	414100	46.2	0.02	-0.01	1023966	915620.37	0.89
GlnA	1194339	65.8	0.02	-0.01	2455160	2130077.5	0.87
PpiB	32695	19.8	0.03	0.01	176464	162272.12	0.92
50S	2658992	85.9	0.04	-0.01	5059822	4181607.1	0.83
Bcp	22962	17.6	0.05	0	145825	132580.04	0.91
PurA	177861	34.9	0.05	0.03	541574	489406.74	0.9
TufA	170376	34.4	0.06	-0.02	547255	474648.72	0.87
DapA	241746	38.6	0.06	-0	670665	611157.97	0.91
Mop	1969190	77.8	0.06	0.03	3870926	3235861.1	0.84
Eno	164550	34	0.08	0.03	508275	463075.17	0.91
DnaK	79355	26.7	0.08	0	320937	281532.77	0.88
Rpo	618409	52.9	0.09	0.05	1559367	1251510	0.8
CysK	120284	30.6	0.09	0.02	427388	372176.36	0.87
Hns	10929	13.8	0.1	0.02	100860	89673.279	0.89
Hup	29059	19.1	0.1	0.02	173491	151520.83	0.87
MetE	166072	34.1	0.12	0.07	533884	466106.23	0.87
SodA	85575	27.3	0.13	0.08	333185	295875.91	0.89
GlyA	155914	33.4	0.13	0.09	494315	445771.06	0.9
IcdA	170979	34.4	0.14	0.11	543143	475842.19	0.88
PanB	265733	39.9	0.15	-0.11	786045	655298.71	0.83
IlvC	96004	28.4	0.15	0.11	360161	319411.9	0.89
SerC	146045	32.7	0.16	0.11	480139	425762.88	0.89
SodB.pqr	78262	26.5	0.16	0.11	323602	278986.25	0.86
AhpC	361605	44.2	0.17	-0.14	1098769	825592.85	0.75
30S	1458112	70.3	0.21	0.13	3107966	2513442.5	0.81
Mdh	119180	30.5	0.21	0.17	422320	369829.89	0.88
GpmA	105080	29.3	0.23	0.19	394194	339437.7	0.86
Suc	268156	40	0.23	0.22	774916	659717.88	0.85
tRNA-GLN	41655	21.5	0.25	0.08	256529	187470.72	0.73

UspA	60387	24.3	0.25	0.24	268088	236035.57	0.88
Pgk	78587	26.6	0.26	0.26	322216	279744.3	0.87
TpiA	97232	28.5	0.27	0.27	370077	322144.85	0.87
Upp	93186	28.1	0.28	0.22	366016	313110.99	0.86
Fba	145047	32.6	0.28	0.28	493950	423724.82	0.86
Asd	146291	32.7	0.29	0.3	486382	426264.81	0.88
GltD	194358	35.9	0.3	0.26	621910	521525.91	0.84
PurC	88909	27.7	0.31	0.33	366271	303466.26	0.83
RpiA	82890	27	0.33	0.37	337562	289714.53	0.86
tRNA-CYS	42428	21.6	0.36	0.34	253484	189572.75	0.75
FusA	144363	32.5	0.37	0.44	525571	422326.38	0.8
tRNA-PHE	44208	21.9	0.37	0.36	259957	194375.51	0.75
Tig	119772	30.6	0.38	0.47	503876	371088.71	0.74
Efp	41367	21.5	0.39	0.42	246304	186684.94	0.76
Tsf	131649	31.6	0.52	0.7	539382	396080.3	0.73
Frr	39978	21.2	0.71	1.18	245654	182874.7	0.74

Table S2. Structural characteristics of partitioned spheres for a macromolecule that cannot be modeled as a single hard sphere. See the definition of R_{eff}, Δ, S, λ, and λ' in the Method section.

Name		Δ	S	R_{eff} (Å)	λ	λ'
tRNA-GLN		0.25	0.08	21.5	0.73	0.95
	Bead 1	0.06	0.03	10.5	0.9	
	Bead 2	0.23	0.02	8.6	0.84	
	Bead 3	0.1	-0.1	16.3	0.91	
	Bead 4	0.24	0.22	15.1	0.86	
UspA		0.25	0.24	24.3	0.88	0.92
	Bead 1	0.02	0	18.9	0.9	
	Bead 2	0.03	0	19.7	0.92	
Pgk		0.26	0.26	26.6	0.87	0.93
	Bead 1	0.02	0	20.8	0.92	
	Bead 2	0.09	0.01	20.8	0.91	
TpiA		0.27	0.27	28.5	0.87	0.94
	Bead 1, 2	0.04	0.01	22.6	0.92	
Upp		0.28	0.22	28.1	0.86	0.90
	Bead 1, 2	0.07	0.01	22.1	0.9	
Fba		0.28	0.28	32.6	0.86	0.91
	Bead 1, 2	0.03	0	25.9	0.89	
Asd		0.29	0.3	32.7	0.88	0.93
	Bead 1, 2	0.06	0	26.1	0.89	
GltD		0.3	0.26	35.9	0.84	0.92
	Bead 1	0.09	0.03	25.5	0.88	
	Bead 2	0.17	0.13	23	0.9	
	Bead 3	0.09	0.04	25.4	0.89	
PurC		0.31	0.33	27.7	0.83	0.91
	Bead 1, 2	0.08	-0	22.1	0.88	
RpiA		0.33	0.37	27	0.86	0.94
	Bead 1	0.06	0	21.5	0.9	
	Bead 2	0.05	0	21.6	0.9	
tRNA-CYS		0.36	0.34	21.6	0.75	0.94
	Bead 1	0.22	0.18	13.7	0.82	
	Bead 2	0.12	0	16.2	0.89	
	Bead 3	0.21	0.18	14.5	0.9	
FusA		0.37	0.44	32.5	0.8	0.90
	Bead 1	0.11	0.06	27.2	0.86	
	Bead 2	0.15	0.08	24.6	0.8	
tRNA-PHE		0.37	0.36	21.9	0.75	0.95

	Bead 1	0.24	0.21	13.6	0.81	
	Bead 2	0.1	0.03	16.2	0.91	
	Bead 3	0.25	0.22	15.2	0.88	
Tig		0.38	0.47	30.6	0.74	0.92
	Bead 1	0.11	0.04	14.4	0.89	
	Bead 2	0.12	-0	14	0.89	
	Bead 3	0.03	-0	22.9	0.85	
	Bead 4	0.12	-0.1	14.2	0.85	
	Bead 5	0.2	0.17	16.3	0.9	
Efp		0.39	0.42	21.5	0.76	0.93
	Bead 1	0.18	0.11	14.8	0.87	
	Bead 2	0.01	0	15	0.91	
	Bead 3	0.14	0.1	14.2	0.89	
Tsf		0.52	0.7	31.6	0.73	0.86
	Bead 1	0.21	0.1	21.7	0.8	
	Bead 2	0.22	0.19	23.1	0.87	
	Bead 3	0.17	0	20.4	0.81	
Frr		0.71	1.18	21.2	0.74	0.98
	Bead 1	0.14	0.1	14.5	0.9	
	Bead 2	0.17	0.14	14.5	0.89	
	Bead 3	0.08	0.04	15.2	0.92	
AhpC		0.17	-0.1	44.2	0.75	0.97
	Bead 1,2,3,4,5	0.16	0.1	25.9	0.89	

Table S3. Composition of macromolecules for an *E. coli* cytoplasmic PD model. The second and the fifth columns are the number of partitioned beads for each macromolecular species. The third and the sixth columns are the number of macromolecules in the system.

Name	# of beads	# of macromolecules	Name	# of beads	# of macromolecules
Ppa	1	25	IlvC	1	50
GapA	1	28	SerC	1	31
Adk	1	39	SodB.pqr	1	25
Pyr	1	8	AhpC	5	20
CspC	1	202	30S	1	28
Pnp	1	8	Mdh	1	36
GlnA	1	3	GpmA	1	11
PpiB	1	20	Suc	1	11
50S	1	28	tRNA-GLN	4	104
Bcp	1	22	UspA	2	20
PurA	1	11	Pgk	2	73
TufA	1	507	TpiA	2	14
DapA	1	6	Upp	2	31
Mop	1	6	Fba	2	17
Eno	1	50	Asd	2	11
DnaK	1	31	GltD	3	8
Rpo	1	11	PurC	2	20
CysK	1	36	RpiA	2	8
Hns	1	20	tRNA-CYS	3	104
PanB	1	6	FusA	2	62
Hup	1	34	tRNA-PHE	3	104
MetE	1	596	Tig	5	25
SodA	1	36	Efp	3	39
GlyA	1	42	Tsf	3	34
IcdA	1	120	Frr	3	20

Table S4. Difference in the folding free energy ($\Delta\Delta G_{uf}$) between the bulk and the crowded environment at several conditions at 1.32 k_BT/ε calculated by the coarse-grained molecular simulations and the scaled particle theory (SPT). See the definition of $\Delta\Delta G_{uf}$(SPT-Minton) and $\Delta\Delta G_{uf}$(SPT-Zhou) in the method section. The unit is in kcal/mol. The deviation of ($\Delta\Delta G_{uf}$) in HS and PD models from 10 independent simulations is shown in the parentheses. The error bar for the simulation data of the HS and the PD model is calculated by the definition of the standard deviation divided by N samples and N=10. The error bar for the simulation data of the F70 model is calculated by the WHAM program[8].

	F70, ϕ_c=25%	F70, ϕ_c=32%	F70, ϕ_c=40%	HS, ϕ_c=32%	PD, ϕ_c=32%
$\Delta\Delta G_{uf}$(Simulation)	-1.34±0.08	-1.87±0.09	-3.25±0.07	-5.22±0.13 (0.41)	-4.75±0.24 (0.76)
$\Delta\Delta G_{uf}$(SPT-Zhou)*	-0.28	-0.31	-0.24	-1.48	N/A
$\Delta\Delta G_{uf}$(SPT-Minton)	-2.15	-3.55	-5.78	-8.27	N/A

Figure S1. Illustration of the algorithm for the calculation of the local free volume (see the definition in the Method section) for a test protein apoazurin. The largest black circle represents a sphere of radius 114Å, the red circle represents the coarse-grained macromolecules and the blue circle with broken lines represents the two possible positions for a test bead upon an insertion process. The radius of the blue sphere is 45.3Å.

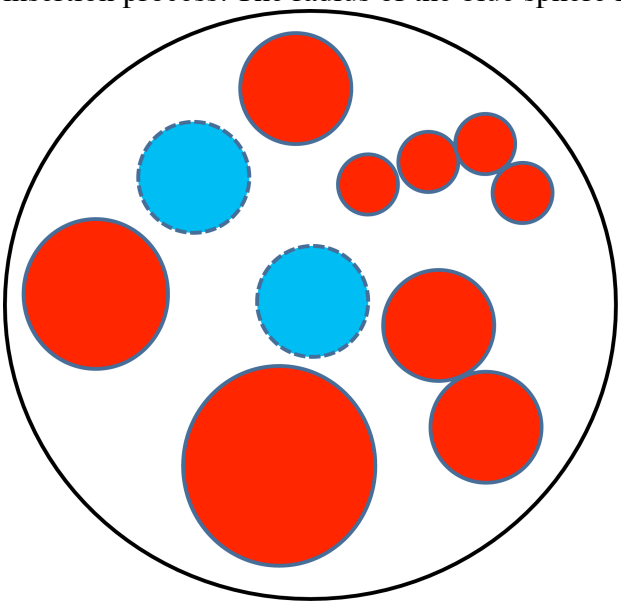

Figure S2. Radius of gyration of apoazurin as a function of temperature in the presence of cytoplasm made of (A) HS model and (B) PD model. For each model, apoazurin was randomly inserted into 10 different voids in the cytoplasm. 10 independent simulations were performed and each trajectory was represented by a curve in the figure.

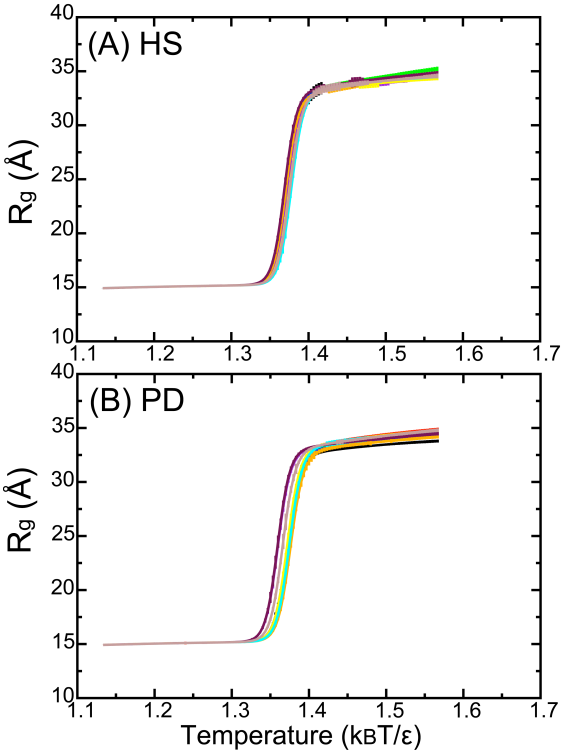

Figure S3. Distribution of the potential energy in replica exchange molecular dynamics simulations. Temperature ranges from $1.13k_BT/\varepsilon$ to $1.57k_BT/\varepsilon$.

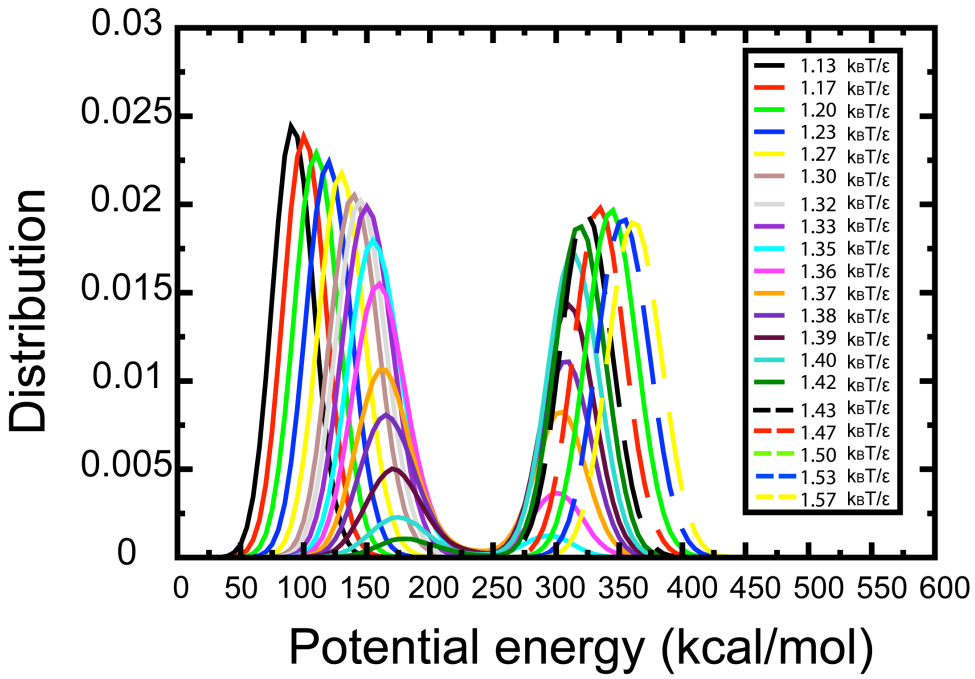

Figure S4. Illustration of all-atomistic structure (AA), PD model (PD) and HS model (HS) for three proteins: ribosome recycling factor (Frr); alkyl hydroperoxide reductase subunit C (AhpC) and phosphoglycerate kinase (PGK). Our test protein is apoazurin, which is showed both in new cartoon representation and a cyan bead. The radius of the cyan bead equals to the radius of gyration of apoazurin.