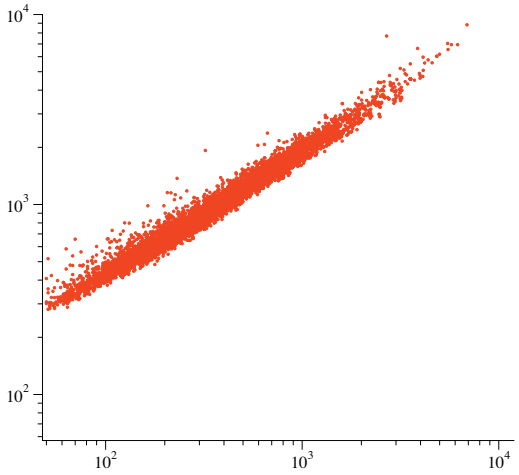
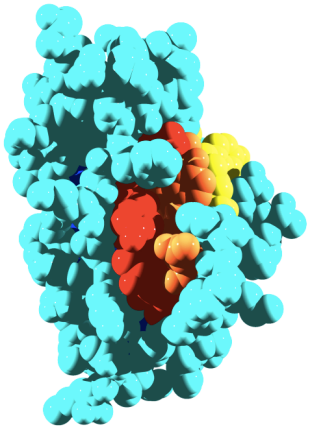
Scaling Laws, the Golden Ratio, & the Small-Angle Scattering of Biomolecules

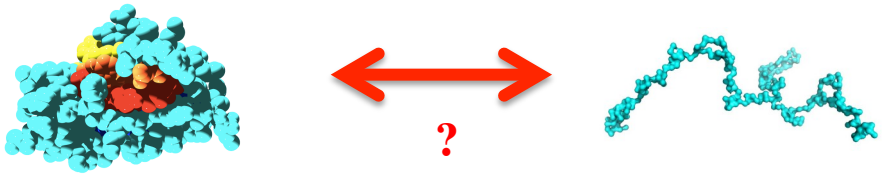




Max C. Watson

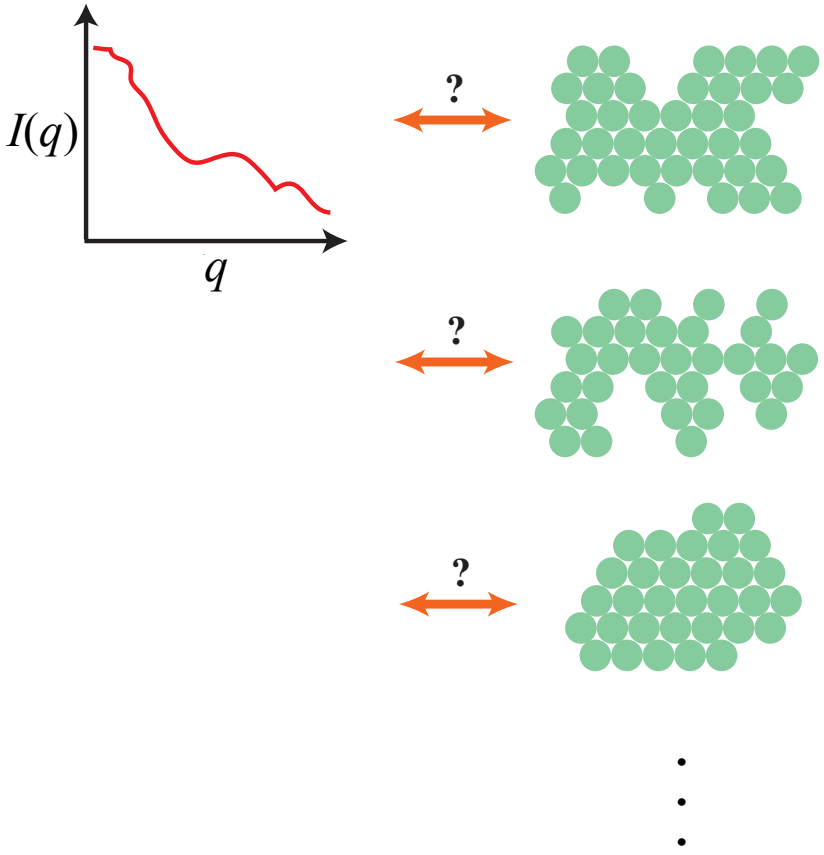
I) A fast method for calculating scattering intensities

II) Scaling Laws & Molecular Disorder



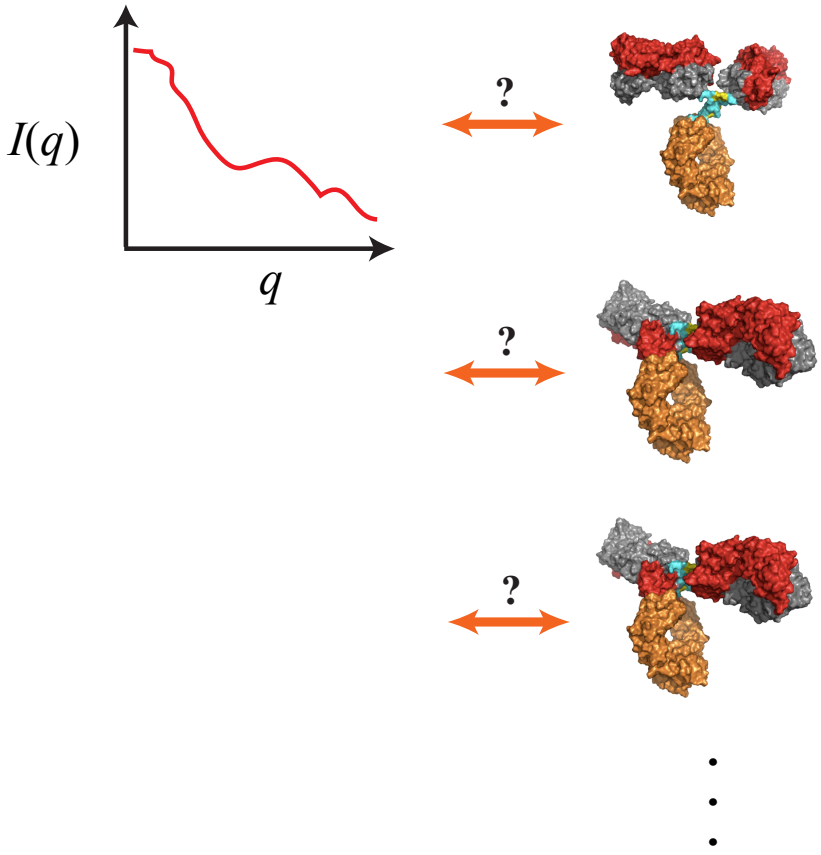
Shape Determination

Ab Initio Methods



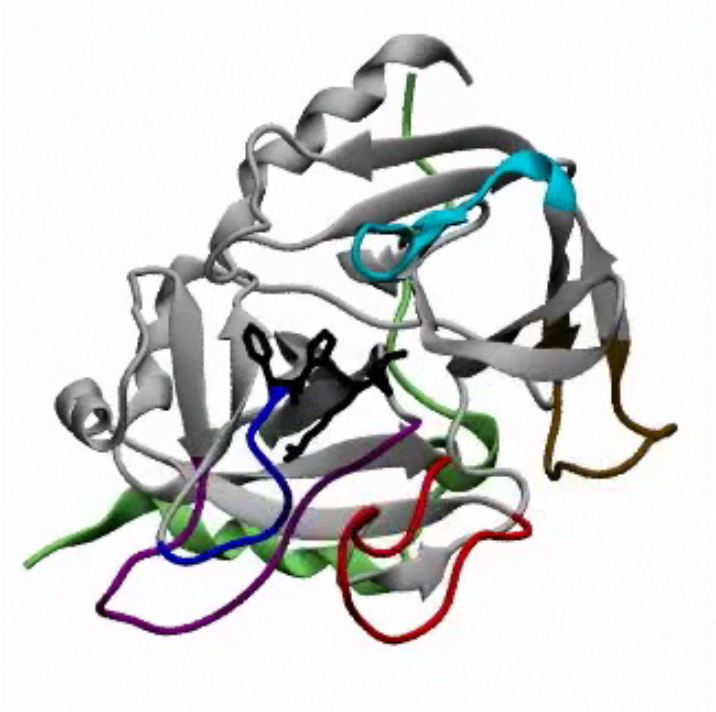
Chacon et al. 1998, Svergun 1999

Rigid Body Modeling

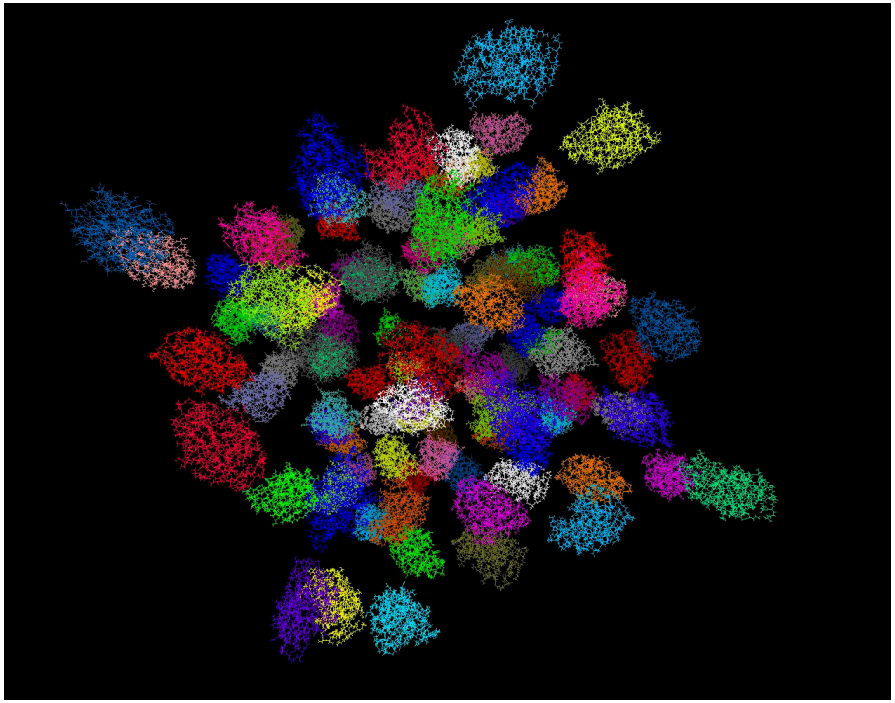


Wall et al. 2000, Curtis et al. 2012

Connection to Simulations

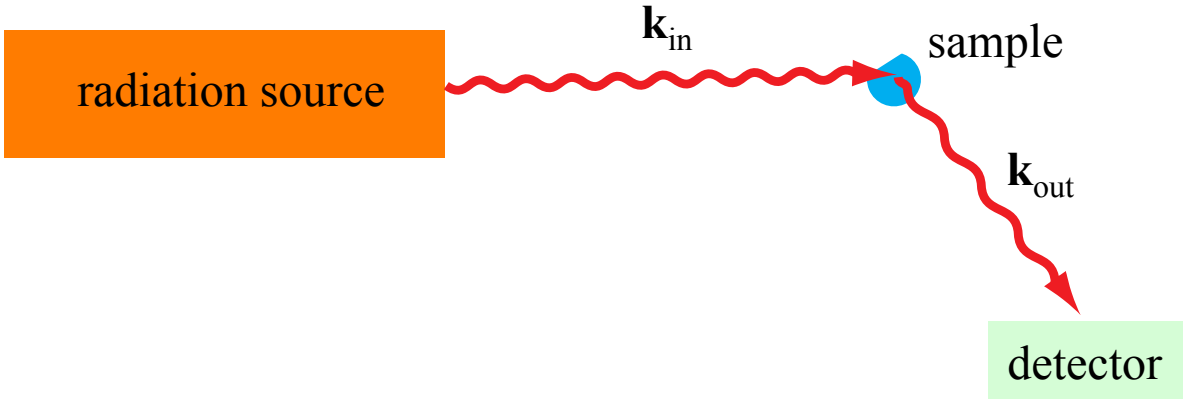


Single Protein
Fuglestad et al. 2012



Concentrated Protein Solutions

Scattering Fundamentals



amplitude $$A(\mathbf{q}) = \sum_{j}^{N} b_j e^{-i\mathbf{q}\cdot\mathbf{r}_j}$$

$\mathbf{q} \equiv \mathbf{k}_{out} - \mathbf{k}_{in}$ scattering vector

$\mathbf{r}_j$ position of atom j

b_j scattering length of atom j

N number of atoms

scattering intensity (flux) $$I(\mathbf{q}) = |A(\mathbf{q})|^2$$

experimental observable $$I(q) = \langle I(\mathbf{q}) \rangle_{\text{all directions}}$$

Easier Said Than Done

$$I(q) = \langle I(\mathbf{q}) \rangle_{\text{all directions}}$$

Exact Result: $$I(q) = \sum_j^N \sum_k^N b_j b_k \frac{\sin(q|\mathbf{r}_j - \mathbf{r}_k|)}{q|\mathbf{r}_j - \mathbf{r}_k|}$$ (Debye 1915)

Summing over all pairs is an $O(N^2)$ calculation

For one protein, $N \sim 10^3 - 10^6$

Even worse for: multiple proteins
flexible domains
shape determination

A Simple Alternative

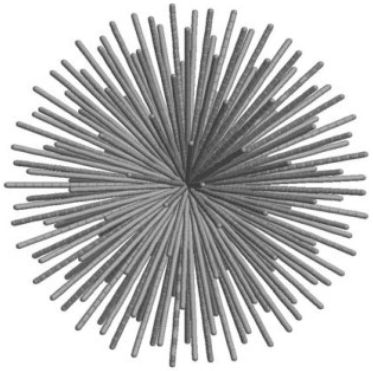
$$I(q) = \langle I(\mathbf{q}) \rangle_{\text{all directions}}$$

q:	scattering vector
$I(\mathbf{q})$:	**q**-dependent scattering intensity
$I(q)$:	measured scattering intensity

Numerically calculating $I[\mathbf{q}]$ at a given **q** scales as $O[N]$ (N = number of atoms)

To get $I(q)$, just average over $I[\mathbf{q}]$ for many **q**'s :

$\mathbf{q}^{(j)}$

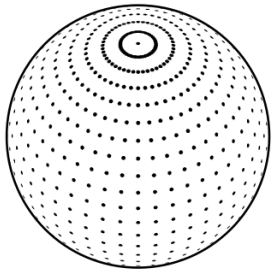


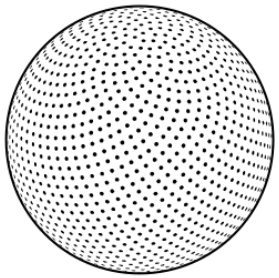
$$I(q) \approx \frac{1}{n}\left[\sum_j I\left(\mathbf{q}^{(j)}\right)\right]$$

Calculating I(q) then scales as $O[nN]$

Generating a Quasi-Spherical Lattice

Bad!

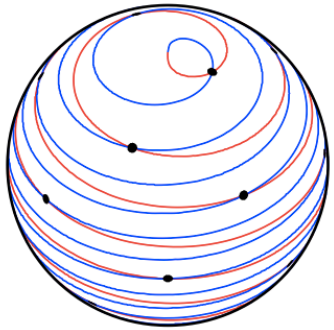




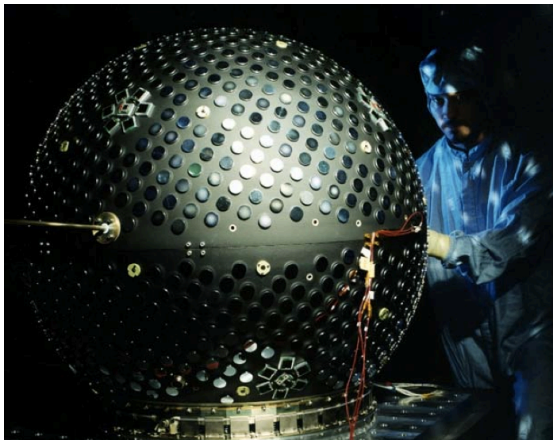
Good

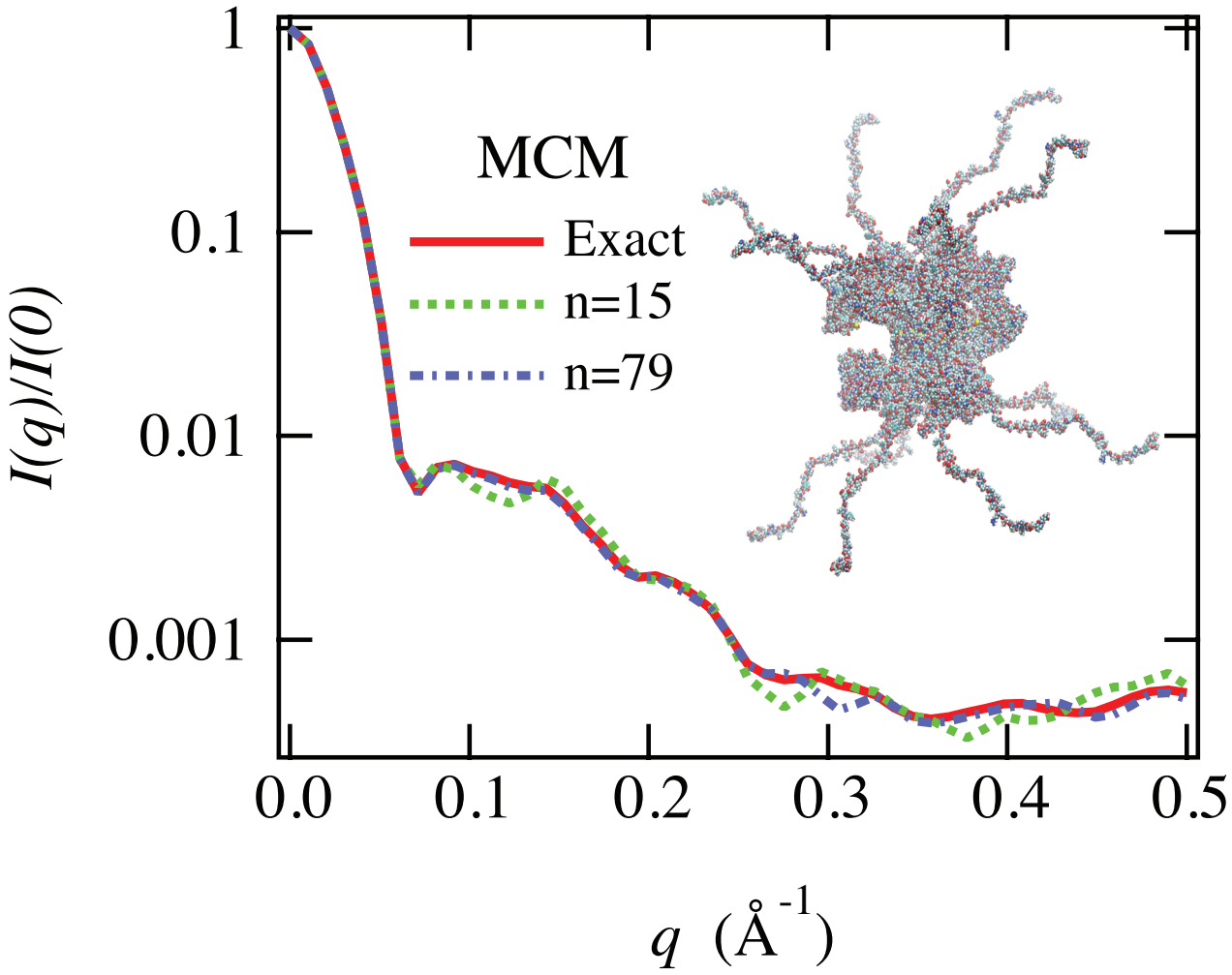
Fibonacci Lattice Built Using Golden Ratio:

$\Phi = 1.618\ldots$



(González 2010)

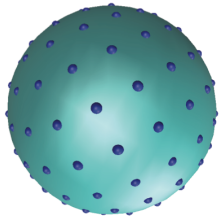




(Krueger et al, 2011)

Golden Vector Method

Step 1: generate n scattering vectors $\mathbf{q}^{(j)}$ on quasi-spherical lattice using golden ratio



Step 2: calculate $I\,[\mathbf{q}^{(j)}]$ for each $\mathbf{q}^{(j)}$

Step 3: average over all $I\,[\mathbf{q}^{(j)}]$: $I\,(q) \approx \frac{1}{n}\left(\sum_{\mathbf{q}^{(j)}\in \text{lattice}} I[\mathbf{q}^{(j)}]\right)$

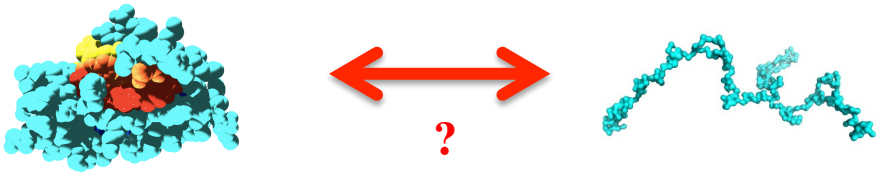
- Speed scales as $O[nN]$
- For given level of accuracy, 2-8 times faster than Spherical Harmonic Method (CRYSON)
- Watson and Curtis, *Journal of Applied Crystallography,* 2013

Summary

- Golden Vector Method: simple yet powerful
  - good for any collection of atoms
  - easy to customize
- <u>Ongoing Work at NCNR:</u>
  - analysis of polymer simulations (Mike Hore)
  - incorporate solvent effects and web server (Hailang Zhang and Joseph Curtis)

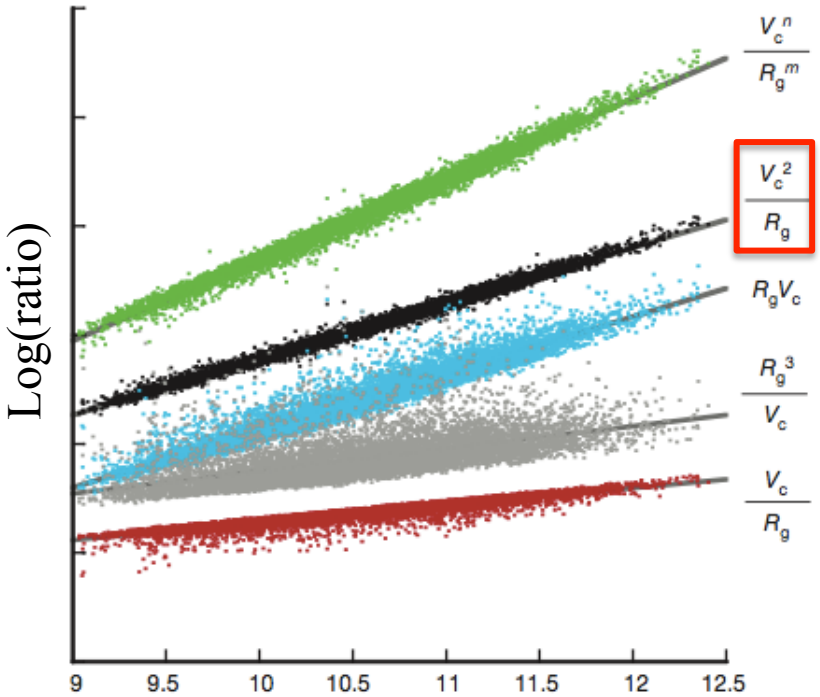
I) A fast method for calculating scattering intensities

II) Scaling Laws & Molecular Disorder



A Curious Result for Compact Proteins:

$$V_c \equiv \left(\int_{\text{all } q} \frac{I(q)}{I(0)} \, q \, \mathrm{d}q \right)^{-1} \qquad R_g = \text{radius of gyration}$$



Rambo & Tainer 2013

$$\text{molecular weight} \propto \frac{V_c^2}{R_g}$$

What is V_c really?

$$V_c\left(q_{\mathrm{m}}\right) \equiv \left(\int_0^{q_{\mathrm{m}}} \frac{I\left(q\right)}{I\left(0\right)} \, q \, \mathrm{d}q \right)^{-1} \qquad q_{\mathrm{m}} = \text{adjustable parameter}$$

Properties well known for $q_{\mathrm{m}} \to \infty$, but what about finite q_{m} ?

It can be shown that:

$$V_c\left(q_{\mathrm{m}}\right) = \frac{2I\left(0\right)}{q_{\mathrm{m}}^2 J\left(q_{\mathrm{m}}\right)}$$

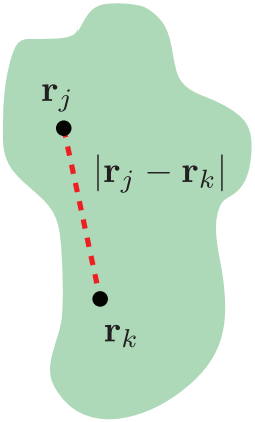
$$J\left(q_{\mathrm{m}}\right) = \sum_j^N \sum_k^N b_j b_k \, \mathrm{sinc}^2\left(q_{\mathrm{m}} |\mathbf{r}_j - \mathbf{r}_k| / 2\right) \qquad \text{and} \qquad \mathrm{sinc}\left(x\right) \equiv \frac{\sin\left(x\right)}{x}$$

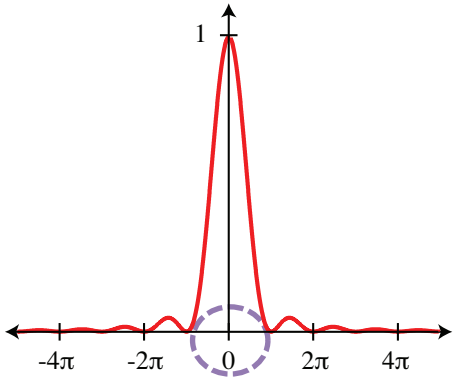
$\mathbf{r}_j$	position of atom j
b_j	scattering length of atom j
N	number of atoms

$$J\left(q_{\mathrm{m}}\right)=\sum_{j}^{N}\sum_{k}^{N}b_j b_k\,\mathrm{sinc}^2\left(q_{\mathrm{m}}|\mathbf{r}_j-\mathbf{r}_k|/2\right)$$

j, k each run over all atoms

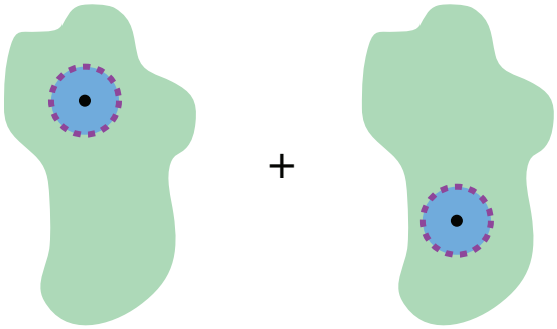
$$\mathrm{sinc}^2\left(x\right)$$





q_{m} describes an effective probe size

$$J\left(q_m\right)=$$



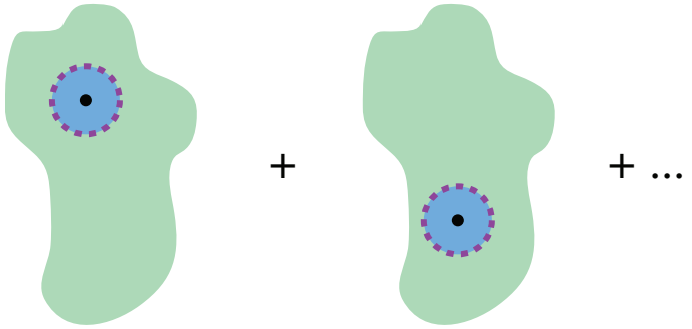
+

+ ...

k =1 k =2

Compact Molecules:

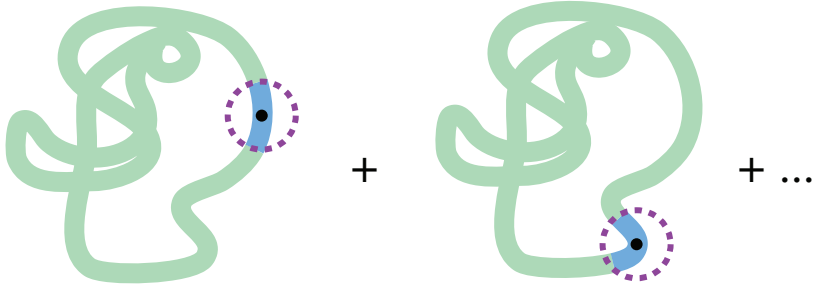
$$J\left(q_m\right) =$$



k =1 k =2

Disordered molecules:

$$J\left(q_m\right) =$$



k =1 k =2

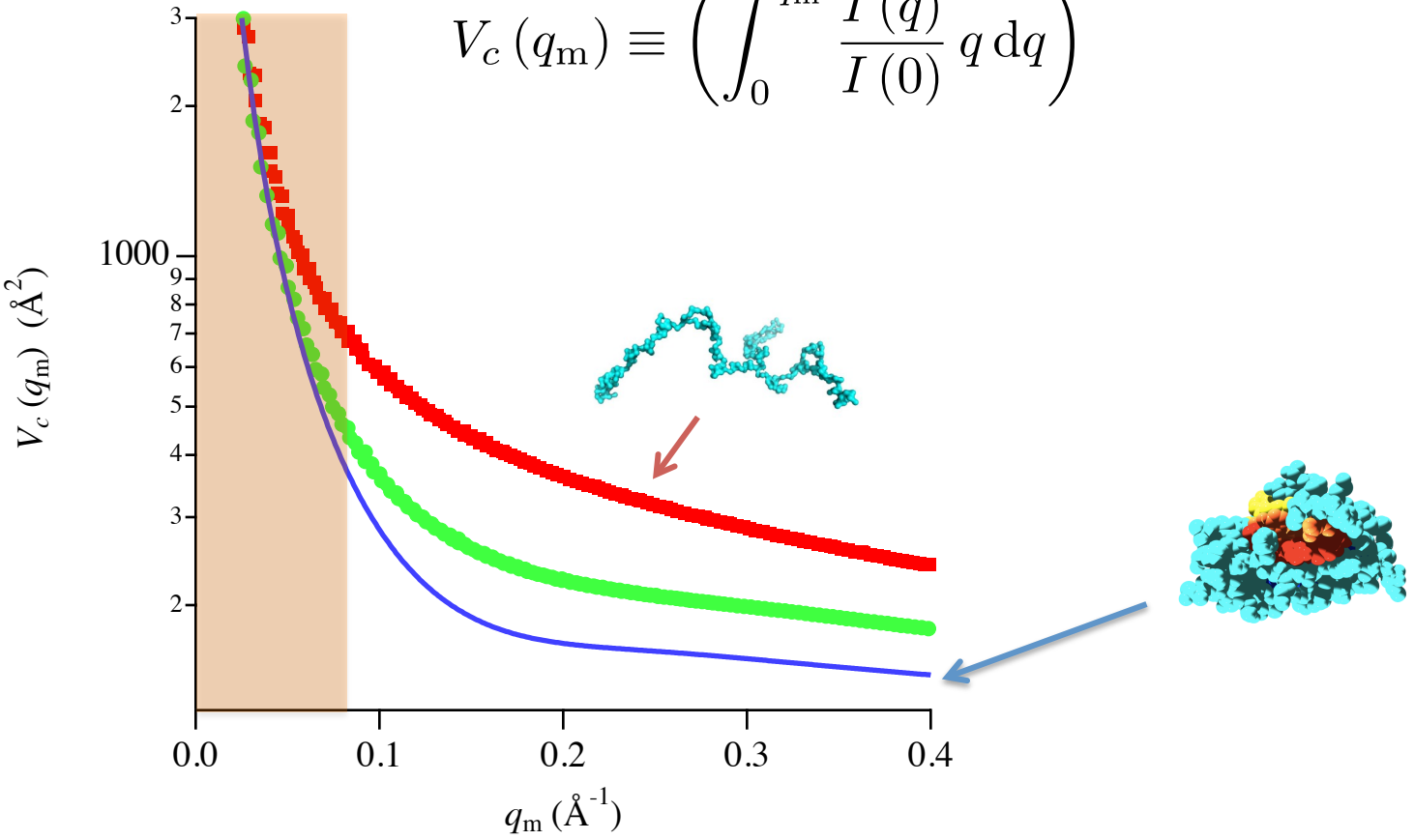
smaller than $J\left(q_{\mathrm{m}}\right)$ for compact molecules

$$V_{\mathrm{c}}\left(q_{\mathrm{m}}\right) = \frac{2I\left(0\right)}{q_{\mathrm{m}}^2 J\left(q_{\mathrm{m}}\right)}$$

$V_{\mathrm{c}}\left(q_{\mathrm{m}}\right)$ is *larger* for disordered molecules, for a given number of atoms

V_c probes molecular disorder

$$V_c\left(q_{\mathrm{m}}\right) \equiv \left(\int_0^{q_{\mathrm{m}}} \frac{I\left(q\right)}{I\left(0\right)}\, q\, \mathrm{d}q\right)^{-1}$$



information about disorder is even contained at very low q_{m}

Scaling Laws for V_c

$q_{\mathrm{m}}^{SAS} = (0.2 - 0.5)\ \text{Å}^{-1}$ Upper limit for Small-Angle Scattering (SAS)

Calculate $V_c\left(q_{\mathrm{m}}^{SAS}\right)$:

Guinier Approximation: $\dfrac{I(q)}{I(0)} \approx \mathrm{Exp}\left(-\dfrac{R_g^2 q^2}{3}\right) \qquad (qR_g \lesssim 1)$

For compact proteins:

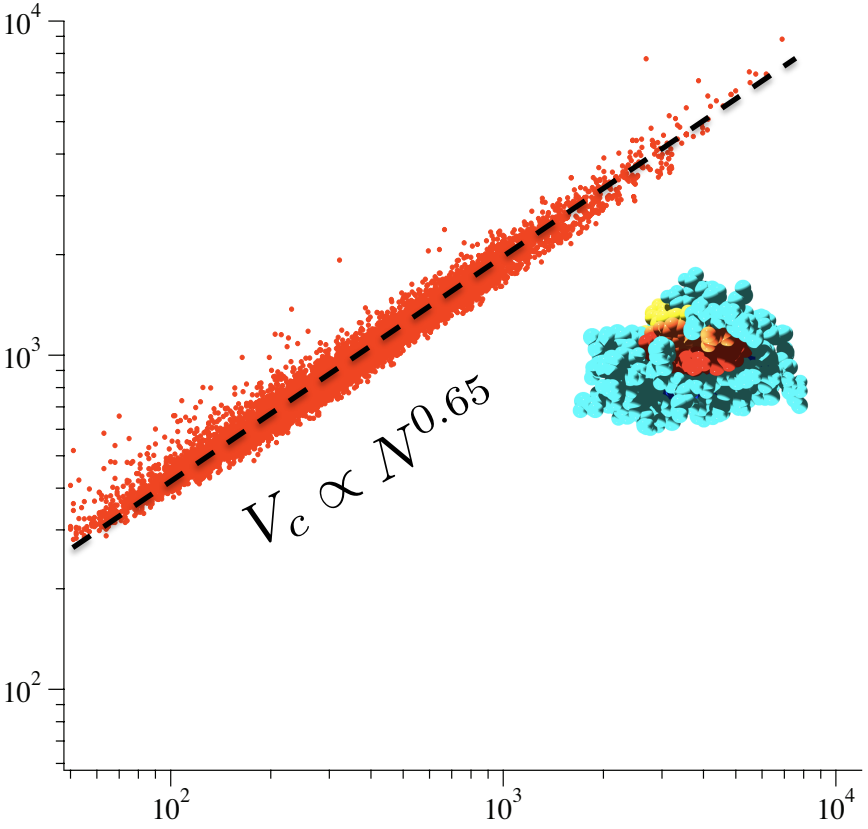
$$V_c\left(q_{\mathrm{m}}^{SAS}\right) \approx \frac{2}{3} R_g^2$$

$R_g \propto N^{1/3}$ N = number of atoms

$$\boxed{V_c\left(q_{\mathrm{m}}^{SAS}\right) \propto N^{2/3}}$$

Theory Agrees with Compact Protein Data

V_c (q_m=0.2 Å^{-1})

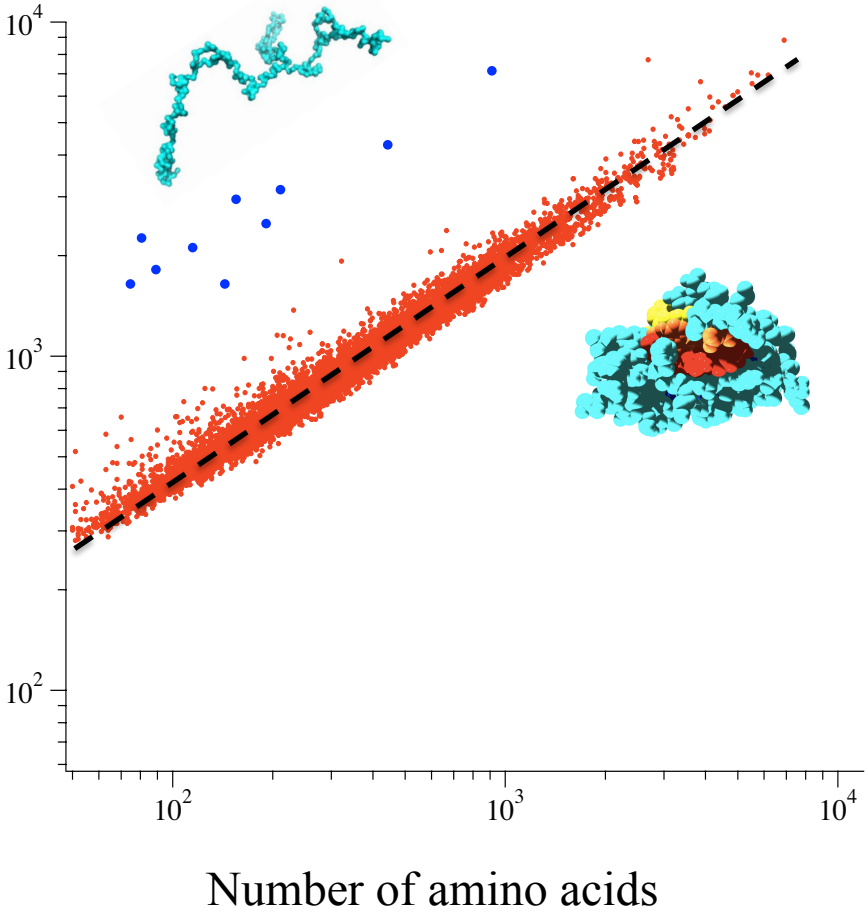


Number of amino acids

$V_c \propto N^{2/3}$ explains Rambo and Tainer finding:

molecular weight $\propto \frac{V_c^2}{R_g}$

Disordered Proteins Have a Larger V_c

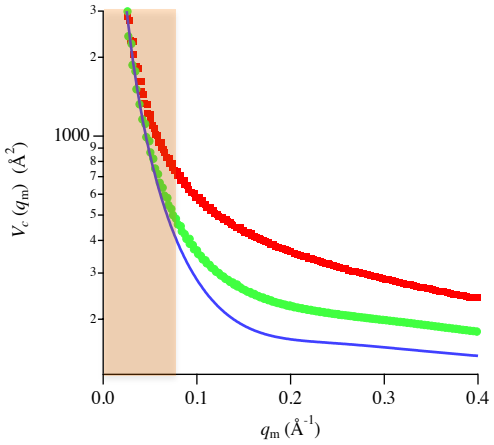


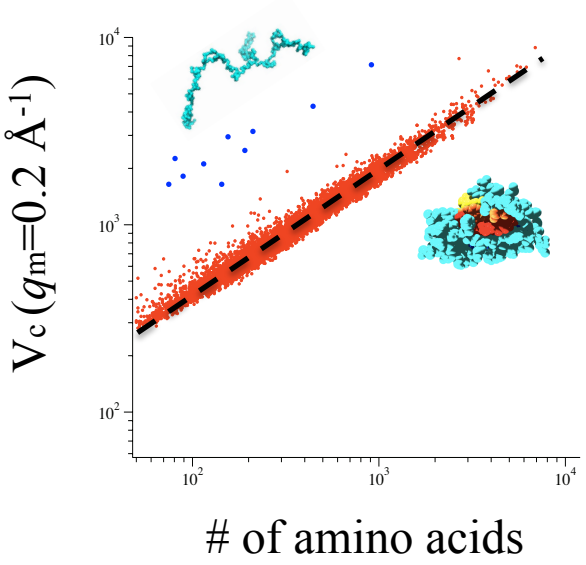
Position in plot can be used to quantify level of disorder

Summary of V_c

$$V_c\left(q_{\mathrm{m}}\right) \equiv \left(\int_0^{q_{\mathrm{m}}} \frac{I\left(q\right)}{I\left(0\right)}\, q\, \mathrm{d}q\right)^{-1}$$

- Concentration independent, relative scale, easy to calculate
- Useful for comparing two molecules with roughly same # of atoms in a model-free way
- Contains good information even at low q_{m}





- Can estimate molecular weight (compact)
- Can be applied to other molecules too (e.g. RNA, polymers…)

Acknowledgements

- Joseph Curtis, Susan Krueger, Michi Nagao, Nick Clark
- Dan Neumann and Rob Dimeo
- National Research Council