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

$I(q)$

$\hat{q}$

$q$

Rigid Body Modeling

$I(q)$

$\hat{q}$

$q$

Chacon et al. 1998, Svergun 1999

Wall et al. 2000, Curtis et al. 2012
Connection to Simulations

Single Protein
Fuglestad et al. 2012

Concentrated Protein Solutions
Scattering Fundamentals

The diagram illustrates the process of scattering, starting with a radiation source, passing through a sample, and ending with a detector. The key variables are:

- $k_{in}$ and $k_{out}$: the in- and outgoing wavevectors, respectively.
- $b_j$: the scattering length of atom $j$.
- $r_j$: the position of atom $j$.
- $N$: the total number of atoms.
- $q = k_{out} - k_{in}$: the scattering vector.

The amplitude of the scattering is given by:

$$A(q) = \sum_{j}^{N} b_j e^{-i q \cdot r_j}$$

The scattering intensity (flux) is:

$$I(q) = |A(q)|^2$$

The experimental observable is:

$$I(q) = \langle I(q) \rangle_{all \ directions}$$
Easier Said Than Done

\[ I(q) = \langle I(q) \rangle_{\text{all directions}} \]

Exact Result:  
\[ I(q) = \sum_{j}^{N} \sum_{k}^{N} b_{j}b_{k} \frac{\sin(q|r_{j} - r_{k}|)}{q|r_{j} - r_{k}|} \]  
(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

\[ I(q) = \langle I(q) \rangle_{\text{all directions}} \]

- **q**: scattering vector
- **I(q)**: \( q \)-dependent scattering intensity
- **I(q)**: measured scattering intensity

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

To get \( I(q) \), just average over \( I[q] \) for many \( q \)'s:

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

Calculating \( I(q) \) then scales as \( O[nN] \)
Generating a Quasi-Spherical Lattice

Bad!  Good

Fibonacci Lattice Built Using Golden Ratio:
$\Phi = 1.618…$

(González 2010)
\[ \frac{I(q)}{I(0)} \]

\( q \ (\text{Å}^{-1}) \)

(MCM)

- **Exact**
- \( n=15 \)
- \( n=79 \)

(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)
Summary

• Golden Vector Method: simple yet powerful
  – good for any collection of atoms
  – easy to customize

• **Ongoing Work at NCNR:**
  
  – 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:

\[ V_c \equiv \left( \int \frac{I(q)}{I(0)} q \, dq \right)^{-1} \]

\[ R_g = \text{radius of gyration} \]

\[ \text{molecular weight} \propto \frac{V_c^2}{R_g} \]
What is $V_c$ really?

$$V_c (q_m) \equiv \left( \int_0^{q_m} \frac{I(q)}{I(0)} q \, dq \right)^{-1} \quad q_m = \text{adjustable parameter}$$

Properties well known for $q_m \to \infty$, but what about finite $q_m$?

It can be shown that:

$$V_c (q_m) = \frac{2I(0)}{q_m^2 J(q_m)}$$

$$J(q_m) = \sum_j \sum_k b_j b_k \text{sinc}^2 \left( q_m |r_j - r_k| / 2 \right) \quad \text{and} \quad \text{sinc}(x) \equiv \frac{\sin(x)}{x}$$

$r_j$ position of atom $j$

$b_j$ scattering length of atom $j$

$N$ number of atoms
\[ J(q_m) = \sum_{j}^{N} \sum_{k}^{N} b_j b_k \text{sinc}^2 (q_m |r_j - r_k|/2) \quad j, k \text{ each run over all atoms} \]

\[ \text{sinc}^2 (x) \]

\[ q_m \text{ describes an effective probe size} \]
Compact Molecules:

\[ J(q_m) = \]

\[ k = 1 \]

\[ k = 2 \]

Disordered molecules:

\[ J(q_m) = \]

\[ k = 1 \]

\[ k = 2 \]

smaller than \( J(q_m) \) for compact molecules

\[ V_c(q_m) = \frac{2I(0)}{q_m^2 J(q_m)} \]

\( V_c(q_m) \) is *larger* for disordered molecules, for a given number of atoms
$V_c$ probes molecular disorder

$$V_c(q_m) \equiv \left( \int_0^{q_m} \frac{I(q)}{I(0)} q \, dq \right)^{-1}$$

Information about disorder is even contained at very low $q_m$.
Scaling Laws for $V_c$

$q_m^{SAS} = (0.2 - 0.5) \text{Å}^{-1}$  \hspace{1cm} \text{Upper limit for Small-Angle Scattering (SAS)}

Calculate $V_c \left( q_m^{SAS} \right)$:

Guinier Approximation:

\[
\frac{I(q)}{I(0)} \approx \text{Exp} \left( -\frac{R_g^2 q^2}{3} \right) \quad (qR_g \leq 1)
\]

\[
V_c \left( q_m^{SAS} \right) \approx \frac{2}{3} R_g^2
\]

For compact proteins:

\[
R_g \propto N^{1/3} \quad N = \text{number of atoms}
\]

\[
V_c \left( q_m^{SAS} \right) \propto N^{2/3}
\]
Theory Agrees with Compact Protein Data

$V_c (q_m = 0.2 \text{ Å}^{-1})$

$V_c \propto N^{2/3}$ explains Rambo and Tainer finding: molecular weight $\propto \frac{V_c^2}{R_g}$
Disordered Proteins Have a Larger $V_c$

$V_c (q_m = 0.2 \text{ Å}^{-1})$

Position in plot can be used to quantify level of disorder
Summary of $V_c$

$$V_c (q_m) = \left( \int_0^{q_m} \frac{I(q)}{I(0)} q \, dq \right)^{-1}$$

- 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…)

![Graph](image_url)
Acknowledgements

• Joseph Curtis, Susan Krueger, Michi Nagao, Nick Clark

• Dan Neumann and Rob Dimeo

• National Research Council