Fredholm Integral Equations in the Analysis of Optical Biosensor and Analytical Ultracentrifugation Data

Peter Schuck
Dynamics of Macromolecular Assembly
Laboratory of Bioengineering & Physical Science
National Institute for Biomedical Imaging and Bioengineering
National Institutes of Health
Bethesda, Maryland

Acknowledgments

Hacene Boukari
LIMB, NICHD, National Institutes of Health, USA

Zhaochun Chen, Robert Purcell
LID, NIAID, National Institutes of Health, USA

Jon Houtman, Lawrence Samelson
LCMB, NCI, National Institutes of Health, USA

Jim Arthos, Claudia Cicala, Anthony Fauci
LI, NIAID, National Institutes of Health, USA

Thomas Vorup-Jensen, Timothy Springer
Harvard Medical School

Shlomo Trachtenberg
The Hebrew University of Jerusalem

Richard Leapman
LBPS, NIBIB, National Institutes of Health, USA

Patrick Brown, Inna Gorshkova, Andrea Balbo, Juraj Svitel, Huaying Zhao
DMA, LBPS, NIBIB, National Institutes of Health, USA

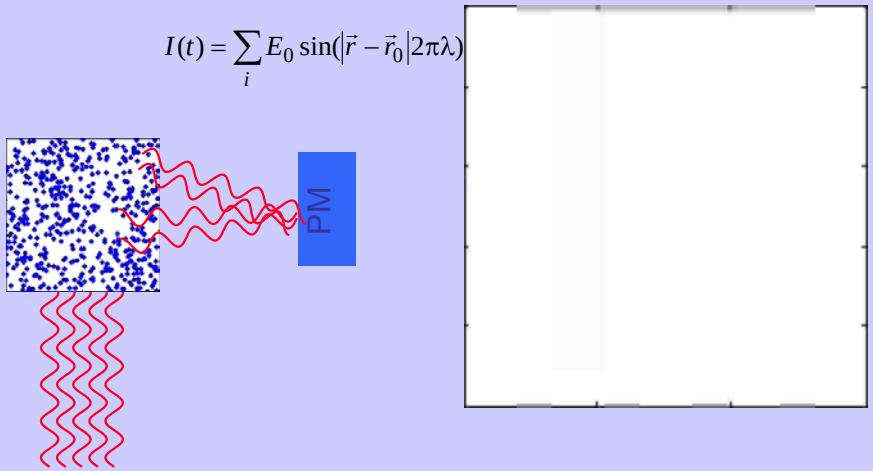
Background

- data $a(\tau)$ often report only indirectly on the quantity of interest $c(R)$

$$a(\tau) = \int_{R_{\min}}^{R_{\max}} c(R) g(R,\tau) dR$$

$$g^{(1)}(R,\tau) = a_1 e^{-a_2 q^2 R^{-1} \tau}$$

$$I(t) = \sum_i E_0 \sin(|\vec{r} - \vec{r}_0| 2\pi\lambda)$$



Background

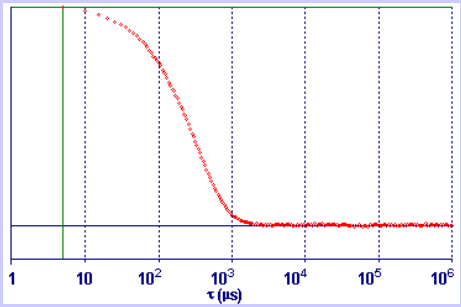
- data $a(\tau)$ often report only indirectly on the quantity of interest $c(R)$

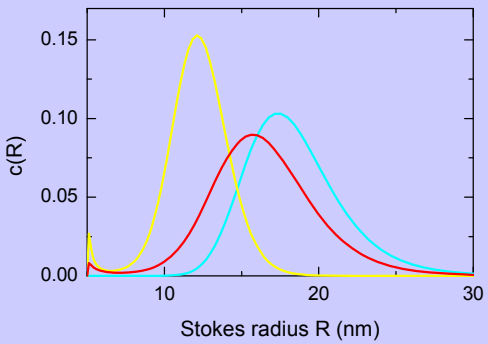
$$a(\tau) = \int_{R_{\min}}^{R_{\max}} c(R) g(R,\tau) dR \qquad g^{(1)}(R,\tau) = a_1 e^{-a_2 q^2 R^{-1} \tau}$$

- Stephen Provencher (1982): fit integral equation to experimental DLS data with CONTIN
- use Tikhonov regularization (or maximum entropy), scale α with F-statistics or $\Delta\chi^2 = 1$

$$\underset{c_i \geq 0}{Min} \left\{ \left\| \vec{a} - \mathbf{G}\vec{c} \right\|^2 + \alpha \left\| \mathbf{B}\vec{c} \right\|^2 \right\}$$

$a(\tau)$

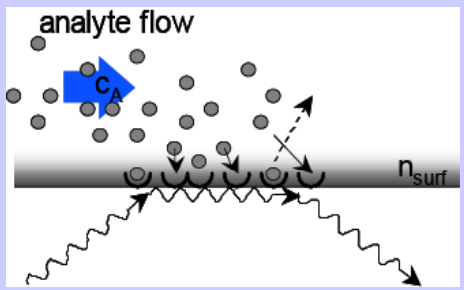


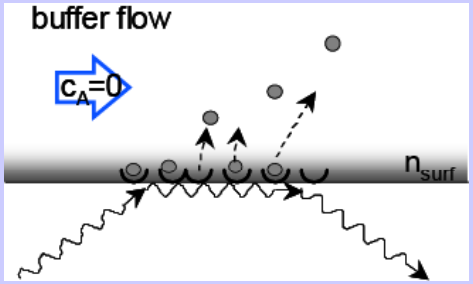


Physical Principle of the Experiment

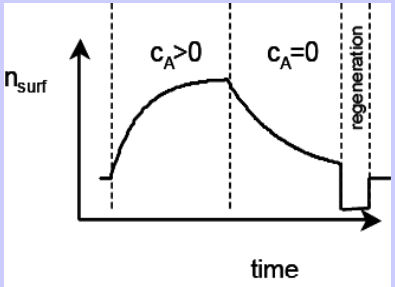
immobilization of protein to serve as binding sites (usually amine coupling to CM dextran)

detection with, e.g., surface plasmon resonance





- highly reproducible
- great signal/noise ratio
- flow-injection of analyte



For a single class of sites with equilibrium dissociation constant K_D and off-rate constant k_{off} (and on-rate constant $k_{on} = k_{off}/K_D$), we expect **piece-wise single-exponential traces**:

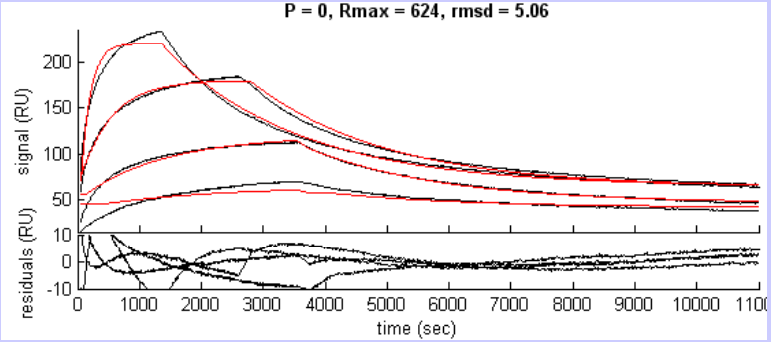
$$s(K_D,k_{off},c,t)=\begin{cases} b_0 & t<t_0 \\ s_a(k_{off},K_D,c,t)+b_c & t_0<t<t_0+t_c \\ s_d(k_{off},K_D,c,t)+b_0 & t>t_0+t_c \end{cases} \quad with$$

$$s_a(K_D,k_{off},c,t)=\left(1+K_D/c\right)^{-1}e^{-k_{off}(1+c/K_D)(t-t_0)}$$

$$s_d(K_D,k_{off},c,t)=s_a(K_D,k_{off},c,t_c)e^{-k_{off}(t-t_c)}$$

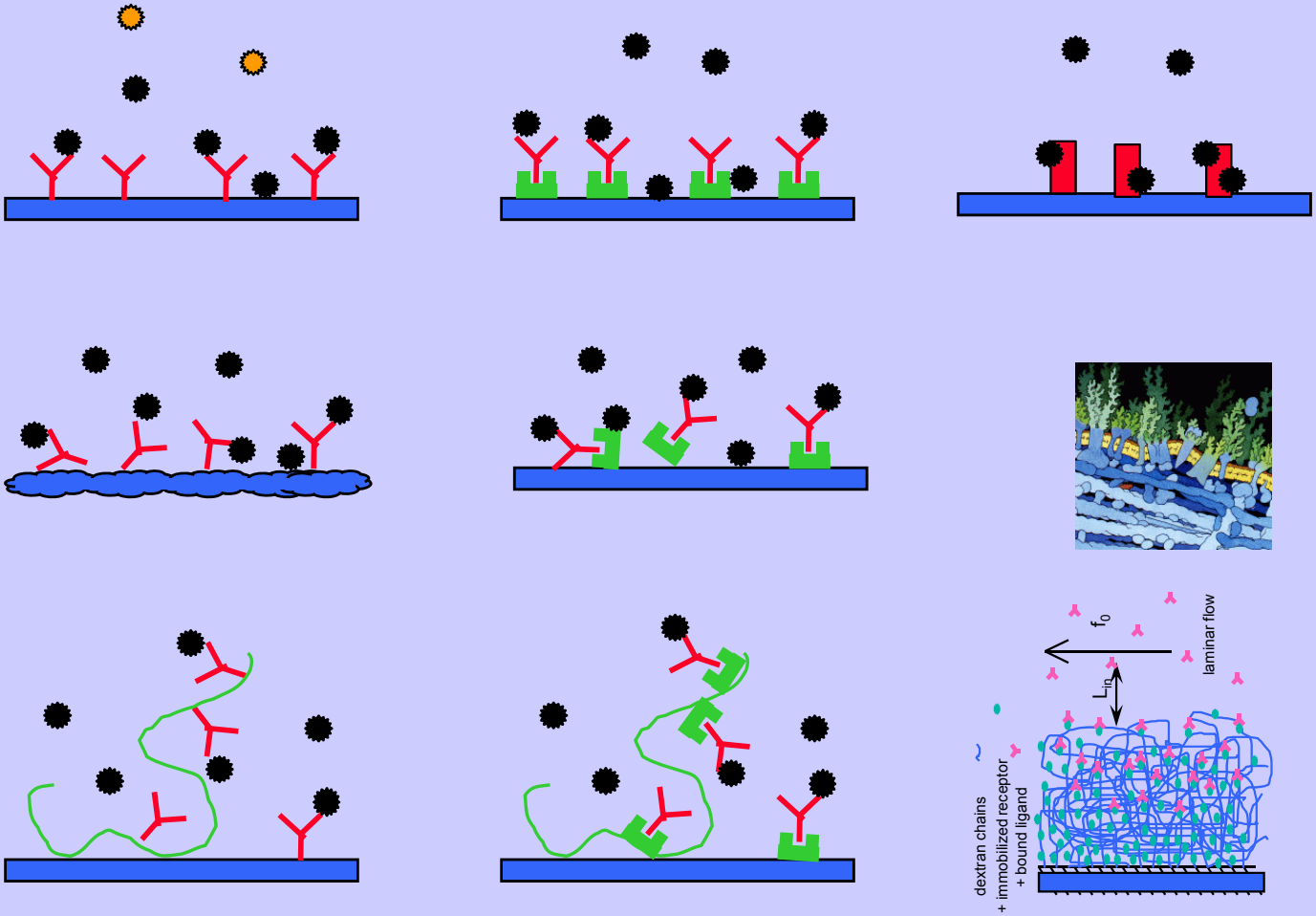
Ann. Rev. Biophys. Biomol. Struct. 26(1997):541

Data: family of traces at different concentrations (obtained from the same surface)
Fit with single discrete site model



→ experimental data are highly reproducible, but usually not single exponentials
(unless truncated, or collected at extremely low signal/noise)

“God made the solid state, he left the surface to the devil” (Enrico Fermi)



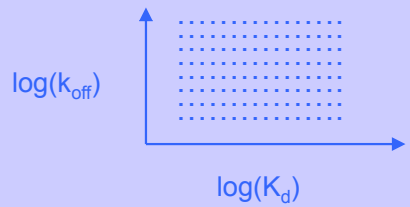
→ **ensemble of surface sites may exhibit a continuum of free energies of binding and rate constants**

The Model

Let $P(K_D,k_{off})dK_D dk_{off}$ be the population of molecules with affinity between K_D and K_D+dK_D, and with chemical off-rate constants between k_{off} and $k_{off}+d\,k_{off}$

$$P_{tot} = \int_{K_{D,\min}}^{K_{D,\max}} \int_{k_{off,\min}}^{k_{off,\max}} P(K_D, k_{off})\, dk_{off}\, dK_D$$

$$s_{obs}(c,t) = \int_{K_{D,\min}}^{K_{D,\max}} \int_{k_{off,\min}}^{k_{off,\max}} s(k_{off}, K_D, c, t) P(k_{off}, K_D)\, dk_{off}\, dK_D$$



without transport limitation

linear, independent sites

$$s(K_D, k_{off}, c, t) = \begin{cases} b_0 & t < t_0 \\ s_a(k_{off}, K_D, c, t) + b_c & t_0 < t < t_0 + t_c \\ s_d(k_{off}, K_D, c, t) + b_0 & t > t_0 + t_c \end{cases}$$

$$s_a(K_D, k_{off}, c, t) = \left(1 + K_D/c\right)^{-1} e^{-k_{off}(1+c/K_D)(t-t_0)}$$

$$s_d(K_D, k_{off}, c, t) = s_a(K_D, k_{off}, c, t_c) e^{-k_{off}(t-t_c)}$$

with a single 'effective' transport rate constant

compartment model for transport,
large system of rate equations coupled by competition for surface concentration of analyte

$$s_{tot}(c,t) = \sum_i s_i(c,t)$$

$$dc_s/dt = k_t(c_0 - c_s) - j_r$$

$$j_r = \sum_i \frac{ds_i}{dt} = k_{on,i}\left(P(K_{D,i}, k_{off,i}) - s_i\right) c_s - k_{off,i} s_i$$

Svitel et al. (2003) Biophys. J. 84:4062-4077
Svitel et al. (2007) Biophys. J. 92:1742

naïve inversion would be ill-conditioned, amplify noise

➔need regularization (e.g., CONTIN)

➔selects from all statistically indistinguishable fits the most parsimonious one

➔F-statistics sets an upper limit for acceptable rmsd values

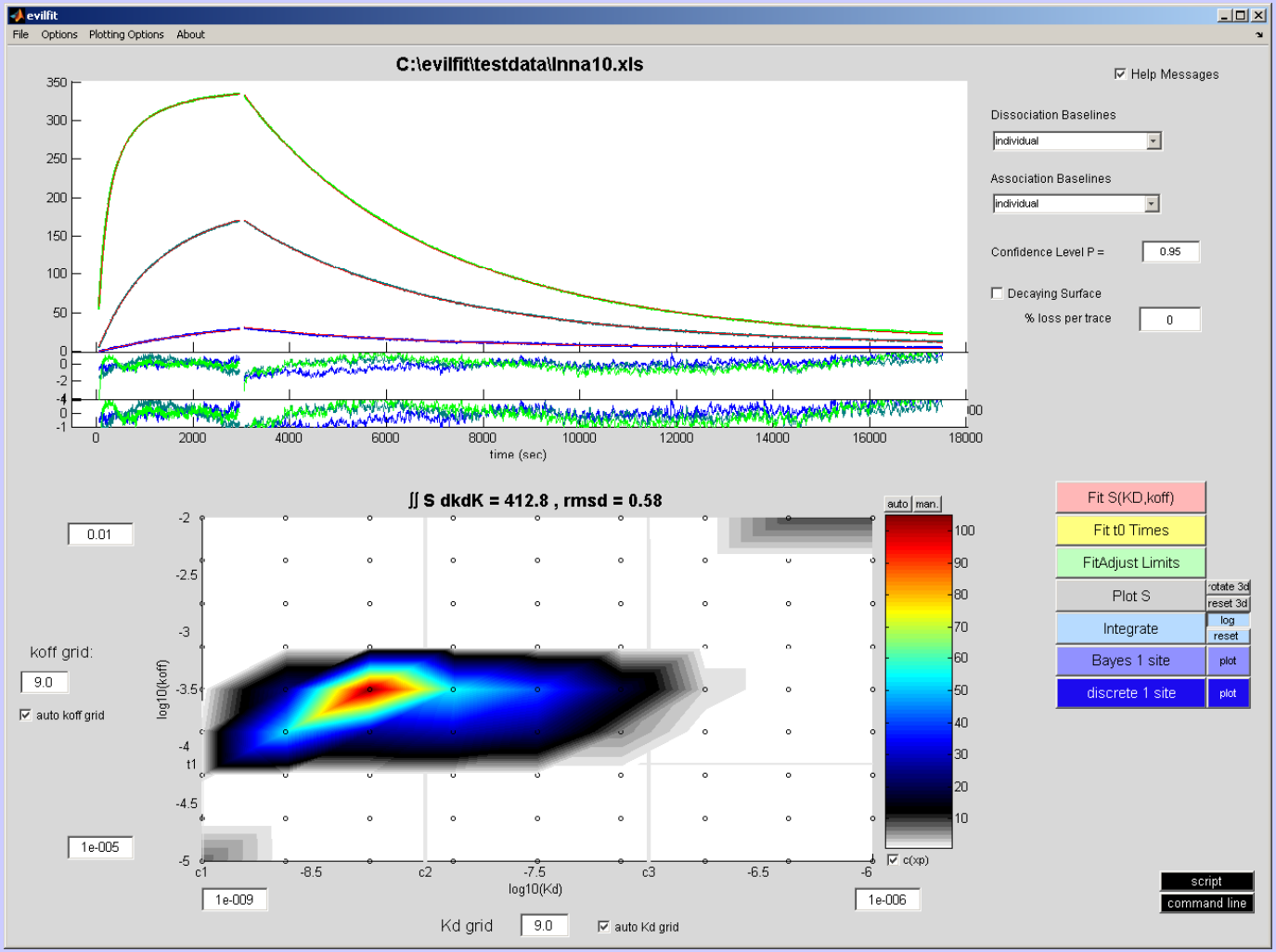
➔maximum entropy or Tikhonov-Phillips regularization

➔with a Bayesian approach to incorporate prior expectation

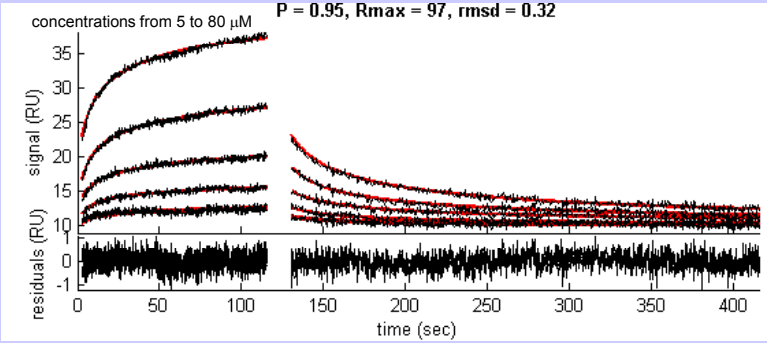
$$Min = \sum_{data\ points} \left(signal_{\exp} - signal_{\text{model}} \right)^2 + \alpha \int \left| \frac{d^2}{dP^2} \left(\frac{P}{prior} \right) \right|^2 dP$$

- prior = constant: all binding parameters are equally likely
- prior = δ-function: there should be a single class of sites
- specific shape of prior could be from previous experiment, mutant, etc.

EVILFIT: public domain GUI for distribution analysis of optical biosensor data in MATLAB

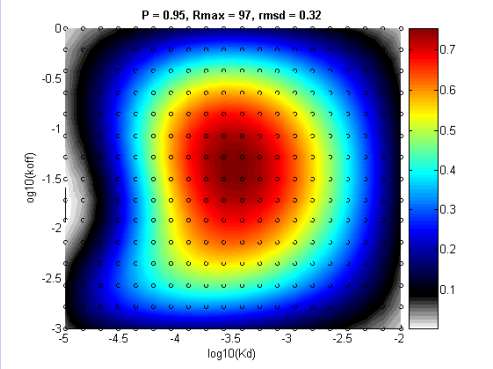


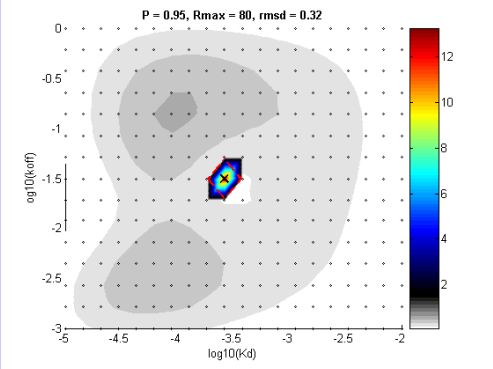
example for the effect of different prior: data with low information content

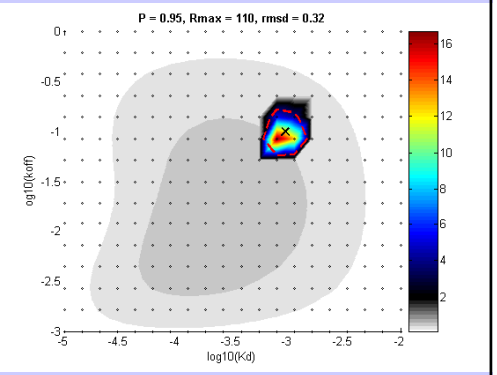


prior = constant:
calculates the broadest possible distribution

prior = δ-function:
gives the distribution closest to a single δ-peak
but not sacrificing quality of fit

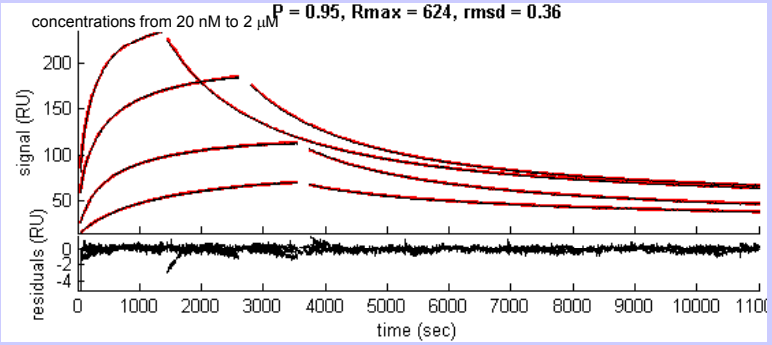






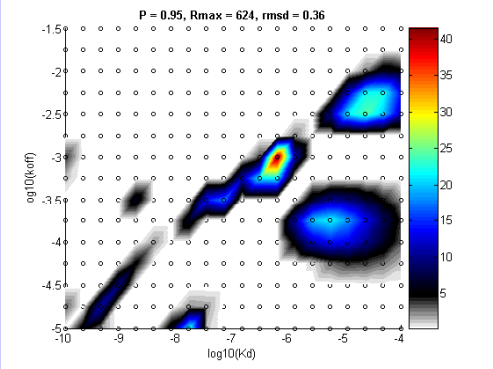
→ probing the space of possible distributions that fit the data
→ this flexibility of interpretation will be small for data with high signal/noise ratio

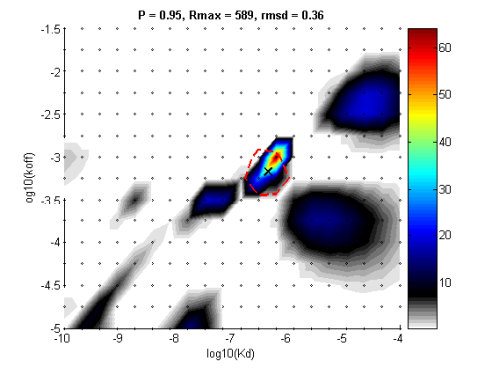
example for the effect of different prior: data with high information content

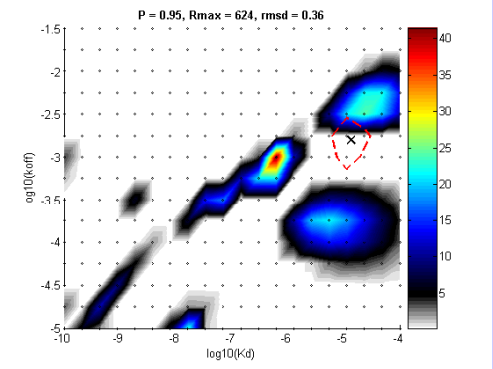


prior = constant:
calculates the broadest possible distribution

prior = δ-function:
gives the distribution closest to a single δ-peak
but not sacrificing quality of fit

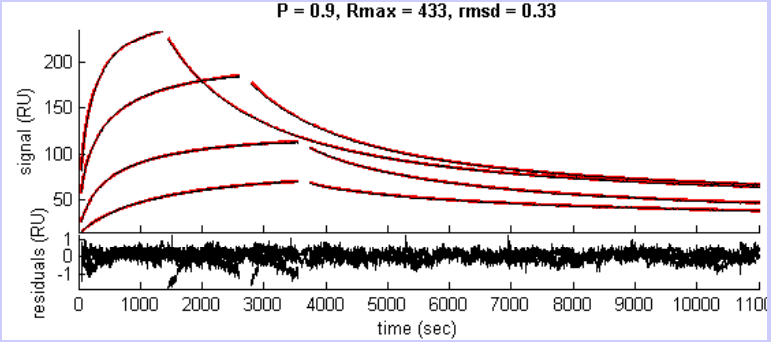


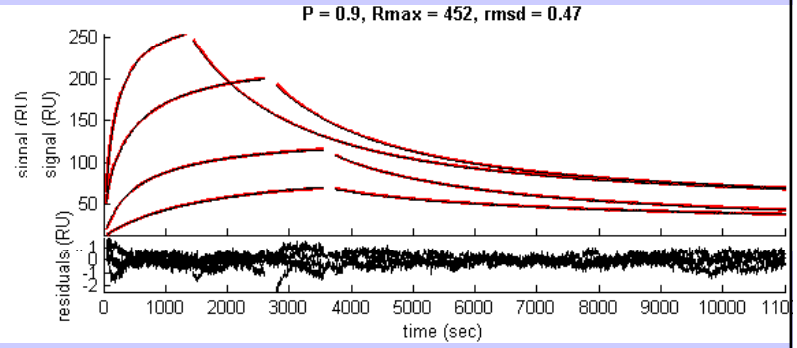


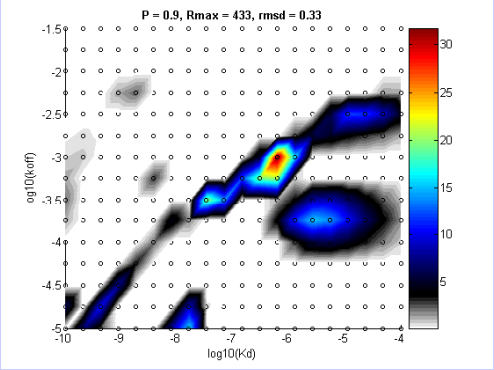


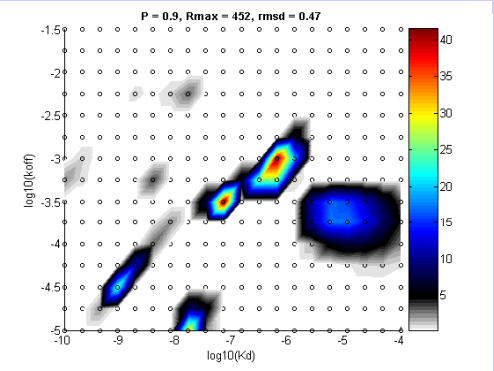
probing the space of possible distributions that fit the data

replicate experiments, same surface density of sites



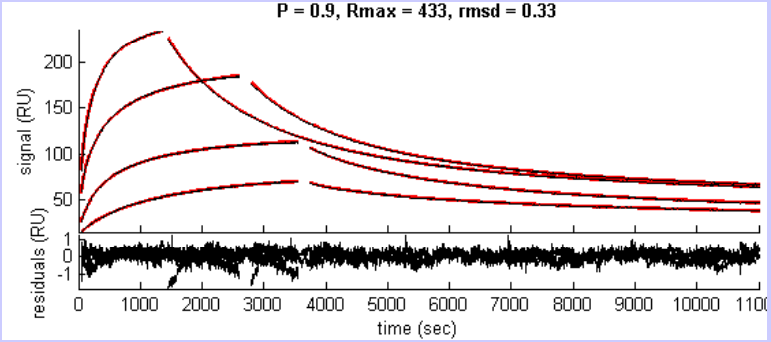


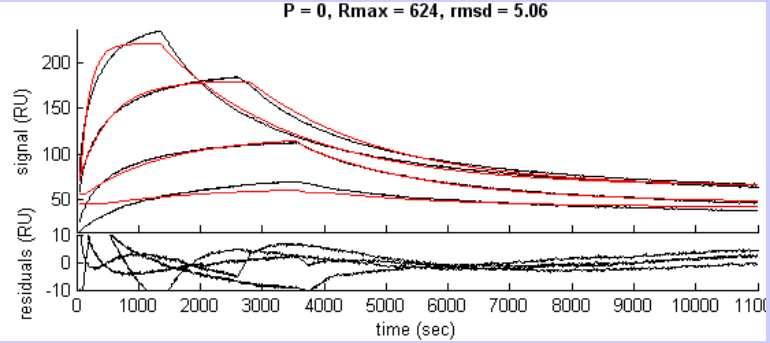


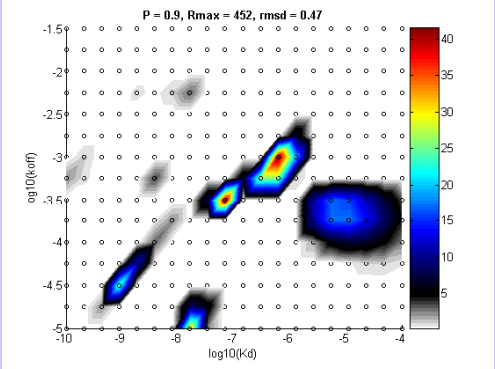


Gorshkova et al., Langmuir 24 (2008) 11577

replicate experiments, same surface density of sites





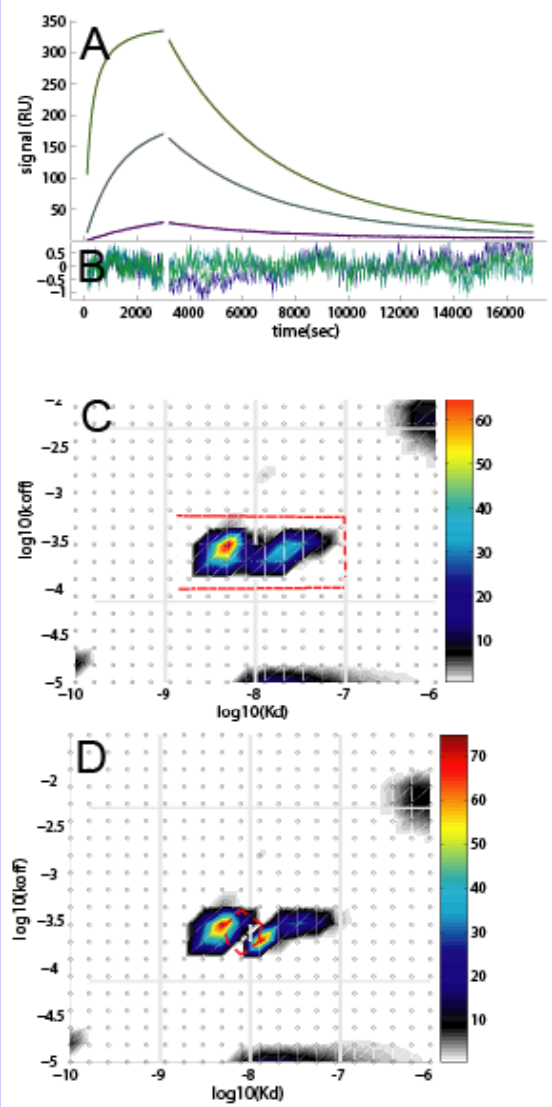


➔from computer simulations: with experimental signal/noise ratio of SPR data, we need > 3-5fold affinity difference in order to resolve two different sites

Gorshkova et al., Langmuir 24 (2008) 11577

Application: Microheterogeneity

Binding of an antigen to its monoclonal antibody immobilized on a short-chain carboxy-methyl dextran surface. Analyte concentrations 1, 10, and 100 nM



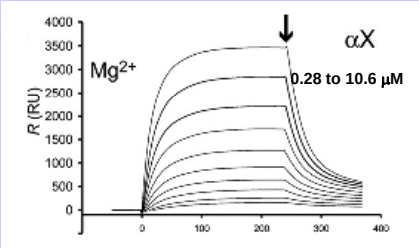
→main peak cannot be regarded as a single site

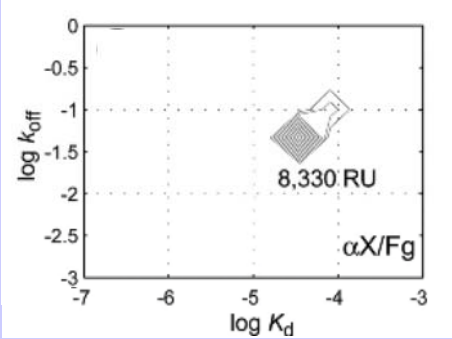
Application: Continuum of Sites

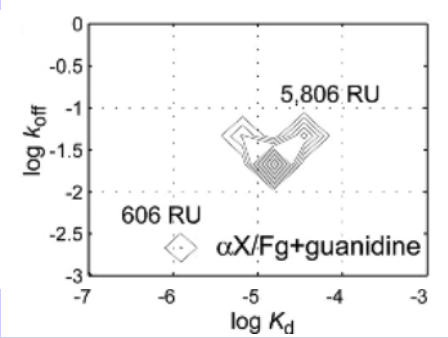
soluble integrin αX domain binding to immobilized fibrinogen:

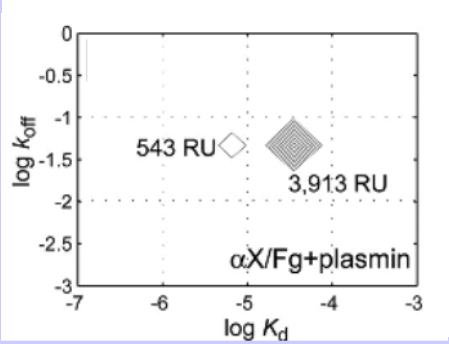
native, proteolyzed, and guanidine-treated fibrinogen

as a model for leukocytes recognition of structurally decayed protein









Vorup-Jensen et al., PNAS (2005) 102:1614

Summary: Functional Heterogeneity of Surface Sites in Biosensing

→accounting for surface heterogeneity gives (for the first time) an excellent description of experimental data in most cases

→ inversion may not be unique, but choice of Bayesian prior allows to probe the space of possible distributions consistent with the data

→ can discriminate conformational change and mass transport processes

→allows naturally to distinguish between 'specific' and 'non-specific' sites

→reveals 'number of sites'

→reveals micro-heterogeneity

→ appropriate for 'complex' functionalized surfaces, optimization of surfaces

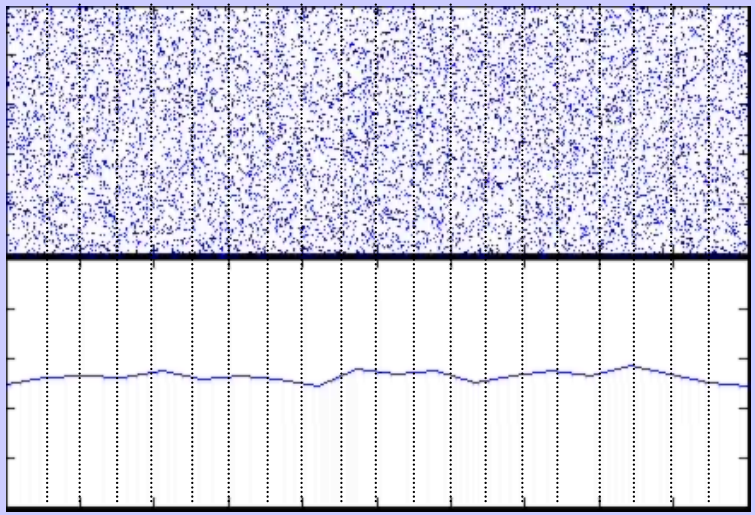
→ implemented in shareware software EVILFIT

→ LIMITATION: Information content of experiment, essentially exponential kernel

Sedimentation Velocity

analytical ultracentrifugation:
diffusion with bias due to centrifugal force

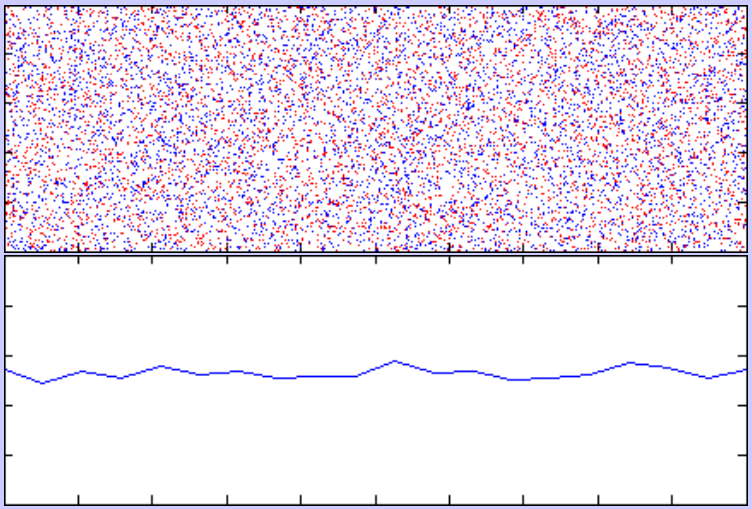
#particles
per volume



(σ = 0.025; δ = 0.0005)

Sedimentation Velocity

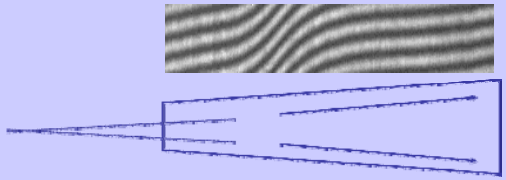
random walk with bias, two non-interacting components



(σ_1 = 0.01; δ_1 = 0.00005; σ_2 = 0.025; δ_2 = 0.0005)

Sedimentation Velocity

laser interferometry (refractive index ~ mass)

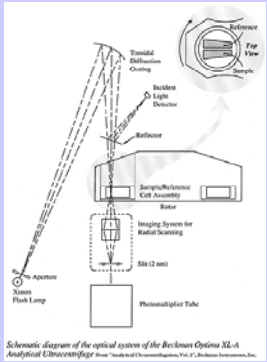


Physical Principle of the Experiment

- Nobel Prize 1926

- sample in optically transparent compartment
- inserted in rotor ~ 300,000 g
- measure evolution of radial concentration distribution
- multiple optical systems (absorbance, Rayleigh interferometry)
- since ~ 10 years, fast finite element solutions of Lamm equation (msec)

$$\frac{\partial \chi}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial \chi}{\partial r} - \chi s \omega^2 r^2 \right]$$

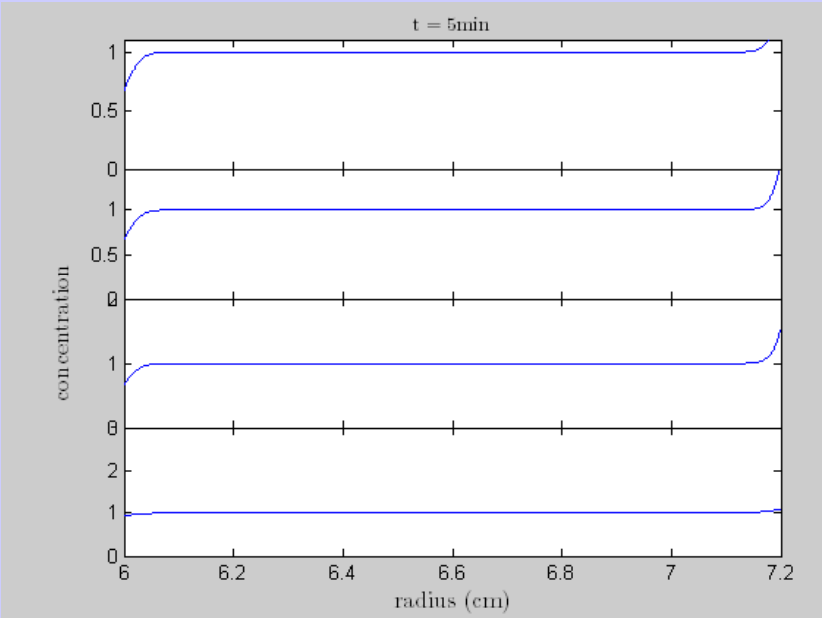


1000 kDa, 30 S

100 kDa, 5 S

10 kDa, 1 S

1 kDa, 0.3 S



$$s \sim \left(M(1 - \overline{v}\rho) \right)^{2/3}$$

$$D(s) \sim as^b \sim s^{-1/2} f_r^{3/2}$$

$$s = \frac{DM\left(1 - \overline{v}\rho\right)}{RT}$$

Biophys. J. 74 (1998) 466
Comput. Phys. Commun. 178 (2008) 105

The Model

Size-and-shape distribution $c\left(s, f_r\right)$

$$a(r,t) = \int_{s_{\min}}^{s_{\max}} \int_{f_{\min}}^{f_{\max}} c\left(s, f_r\right) \chi_1\left(s, f_r, r, t\right) ds df_r$$

Sedimentation coefficient distribution $c\left(s\right)$

$$a(r,t) = \int_{s_{\min}}^{s_{\max}} c\left(s\right) \chi_1\left(s, f_{r,w}, r, t\right) ds$$

→ discretized problem of the form

$$\underset{c_k \geq 0}{Min}\left[\sum_l \left(a_l - \sum_{k=1}^{K} c_k \chi_{kl}\right)^2 + \alpha \sum_{k,\kappa} c_k H_{k\kappa} c_\kappa\right]$$

→ typically with *l* = *10,000 – 100,000* data points, distribution discretized to *k* = *100 – 1000* species, use normal equations and NNLS

→ with Tikhonov regularization $\alpha \sum_{k,\kappa} c_k H_{k\kappa} c_\kappa$ for broad continuous particle distributions

→ maximum entropy $\alpha \sum_k c_k \log(c_k)$ for biological macromolecules

→ scale regularization iteratively with goal to achieve confidence limits from F-statistics

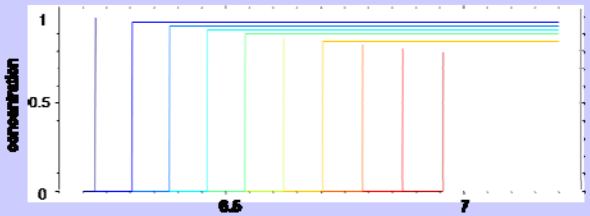
→ Bayesian prior distribution to probe space of possible (statistically indistinguishable) solutions

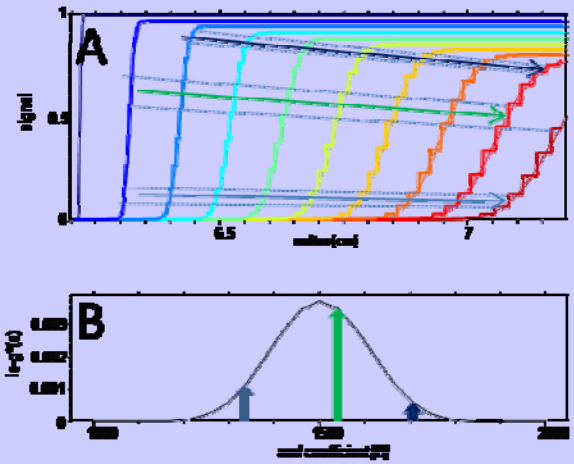
→ **implemented in public domain software SEDFIT and SEDPHAT**

Biophys. J. 78 (2000) 1606
Biophys. J. 89 (2005) 651
Biomacromolecules 8 (2007) 2011

Applications: Very Large Particles

for $D = 0$ it is $\chi_1(r,t) = e^{-2s\omega^2 t} H(r - me^{s\omega^2 t})$





→ calculating distribution is similar to differentiation of noisy data

$$a(r,t) = \int c(s)\, e^{-2s\omega^2 t} H(r - me^{s\omega^2 t}) ds$$

$$\frac{da}{dr} = \int c(s) \frac{r^{*2}}{m^2} \delta(r - r^*) ds$$

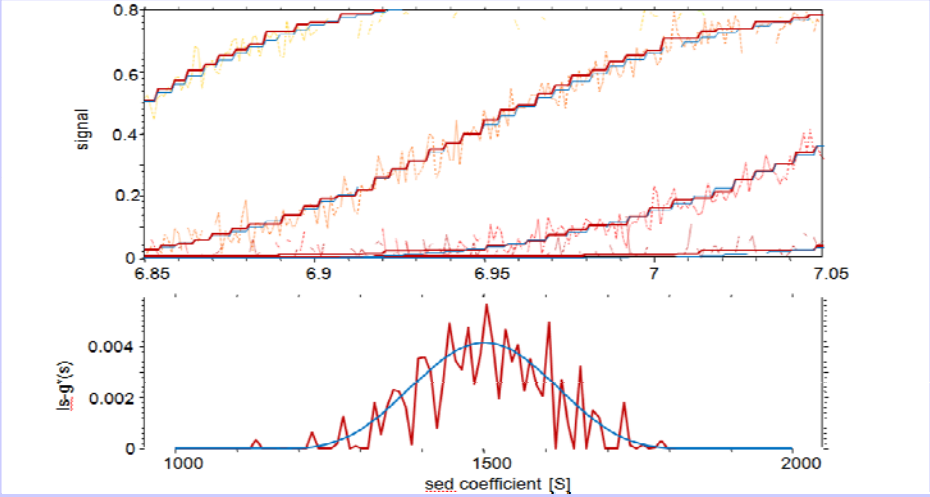
$$= \int c(s) \frac{r^{*2}}{m^2} \left(\frac{dr^*}{ds^*} \right) \delta(s - s^*) ds$$

$$c(s^*) = \frac{\omega^2 t r^3}{m^2} \frac{da}{dr}$$

→clearly highly ill-conditioned if

$$\Delta s < noise \times \frac{m^2}{\omega^2 t r^3} \left(da/dr \right)^{-1}$$

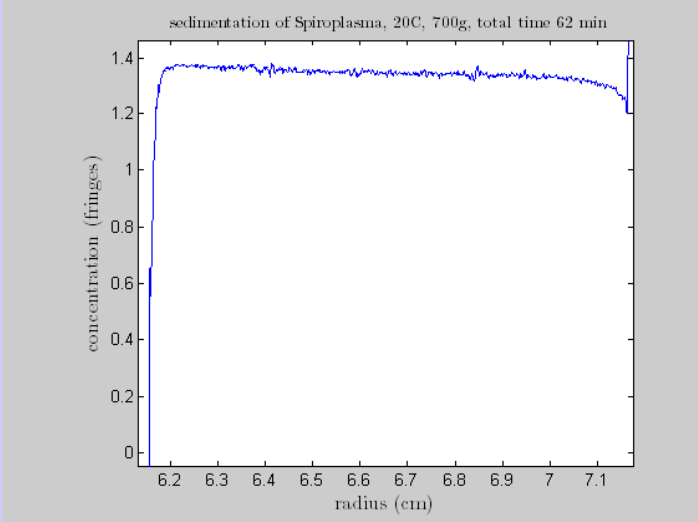
→ resolution of species with different s limited by signal/noise and rotor speed

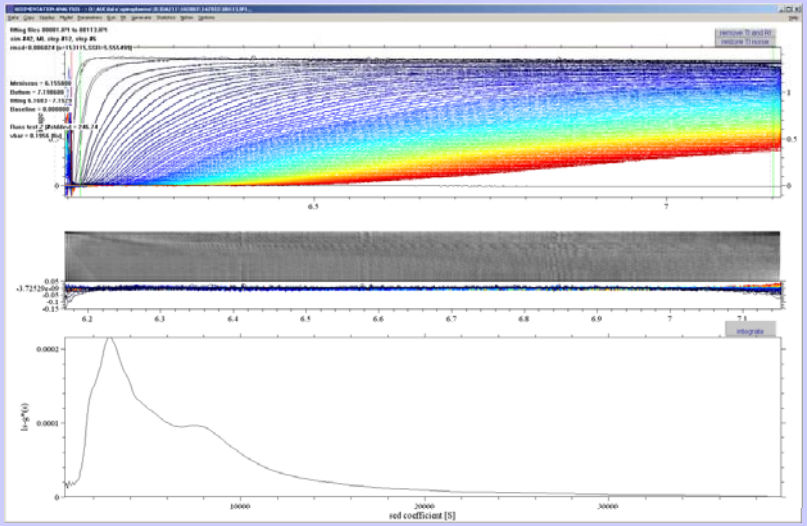


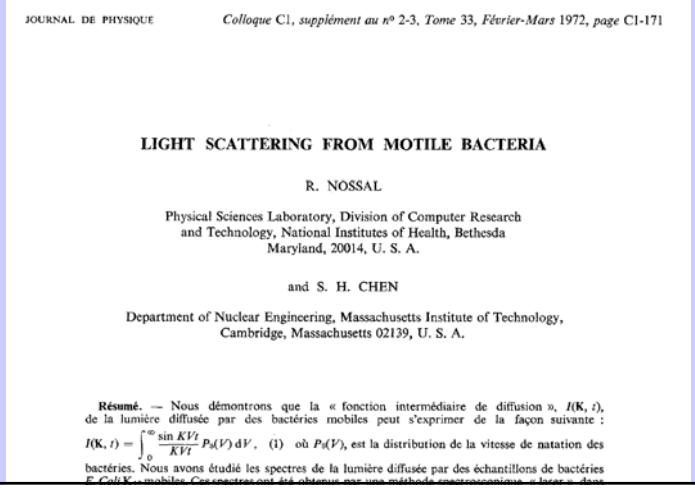
→ boundary spread is sufficiently different from diffusion effect

Applications: Very Large Particles

Sedimentation Velocity







JOURNAL DE PHYSIQUE *Colloque C1, supplément au n° 2-3, Tome 33, Février-Mars 1972, page C1-171*

LIGHT SCATTERING FROM MOTILE BACTERIA

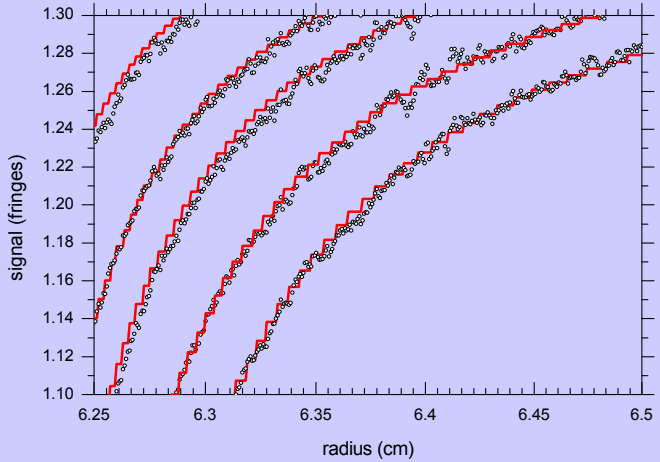
R. NOSSAL

Physical Sciences Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda Maryland, 20014, U. S. A.

and S. H. CHEN

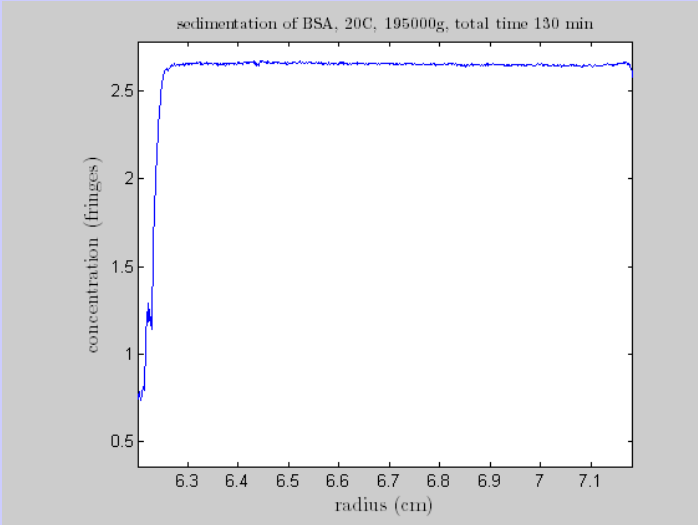
Department of Nuclear Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, U. S. A.

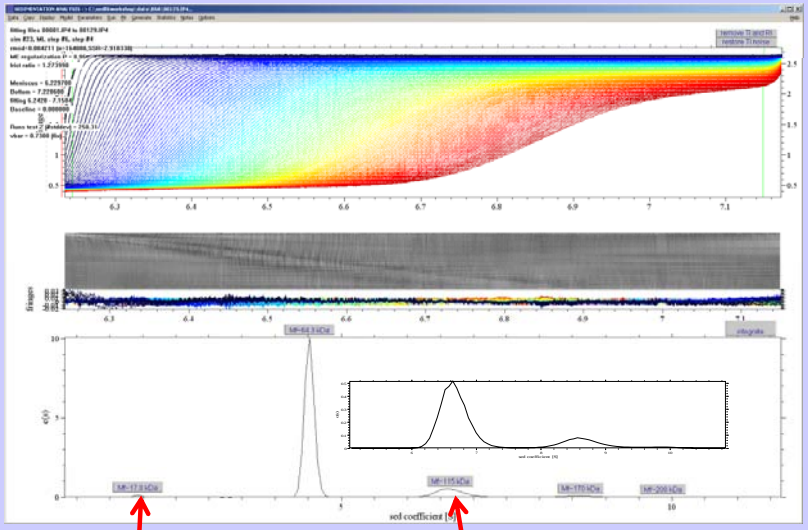
Résumé. — Nous démontrons que la « fonction intermédiaire de diffusion », $I(\mathbf{K}, t)$, de la lumière diffusée par des bactéries mobiles peut s'exprimer de la façon suivante : $I(\mathbf{K}, t) = \int_0^\infty \frac{\sin KVt}{KVt} P_B(V)\, dV$, (1) où $P_B(V)$, est la distribution de la vitesse de natation des bactéries. Nous avons étudié les spectres de la lumière diffusée par des échantillons de bactéries

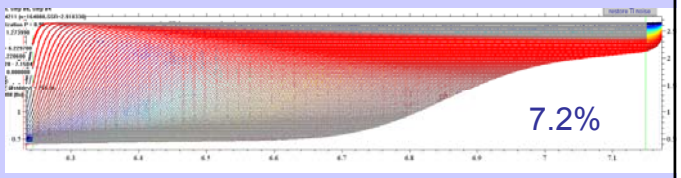


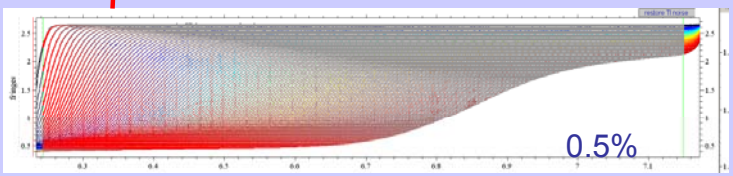
Applications: Protein Complexes

Sedimentation Velocity









→ approximately like step-functions, but diffusion broadened

→ optimize average scaling parameters in non-linear regression, as well as meniscus position

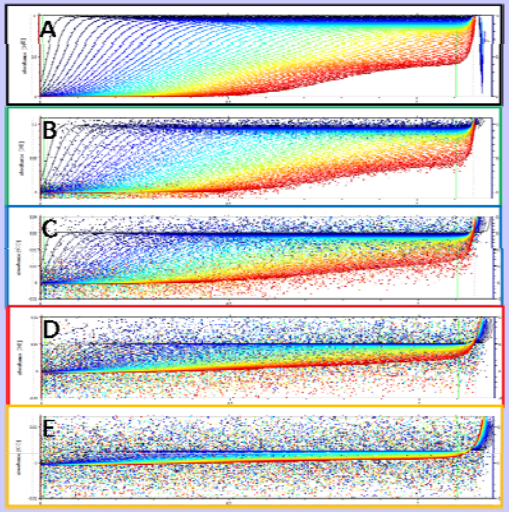
→ diffusion creates additional correlation between species of similar s

→ **highly quantitative for trace species ~ 1 %**

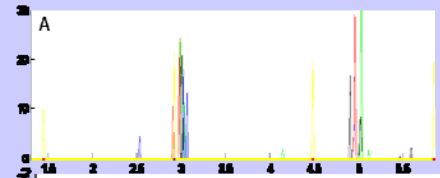
→ **very precise s-values**

→ **time-average s-values of systems of rapidly interacting molecules, interpreted as binding isotherms**

Applications: Regularization

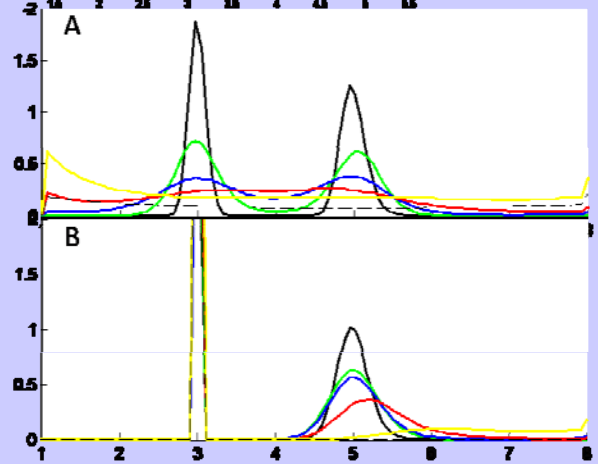


$$\alpha \sum_k c_k \log(c_k)$$

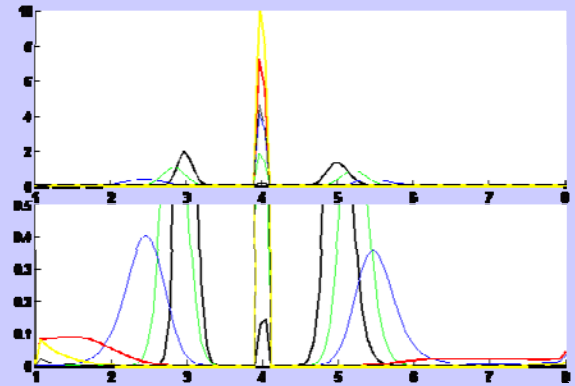


$$\alpha \sum_k c_k \log(\frac{c_k}{p_k})$$

$$p_k = \begin{cases} 1 & for\ s_k = 3 \\ 0 & else \end{cases}$$

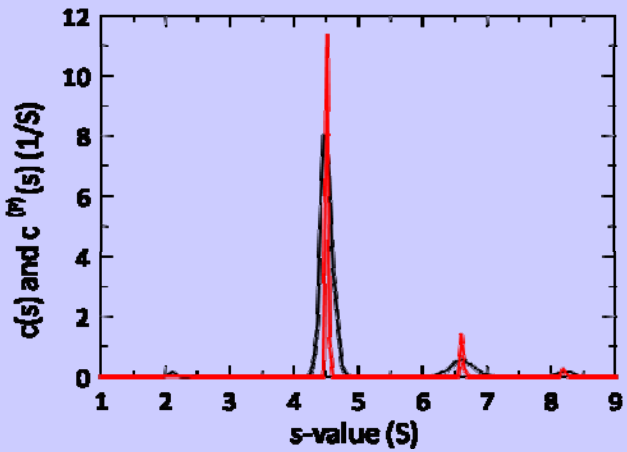


$$p_k = \begin{cases} 1 & for\ s_k = 4 \\ 0 & else \end{cases}$$



Applications: Regularization

BSA

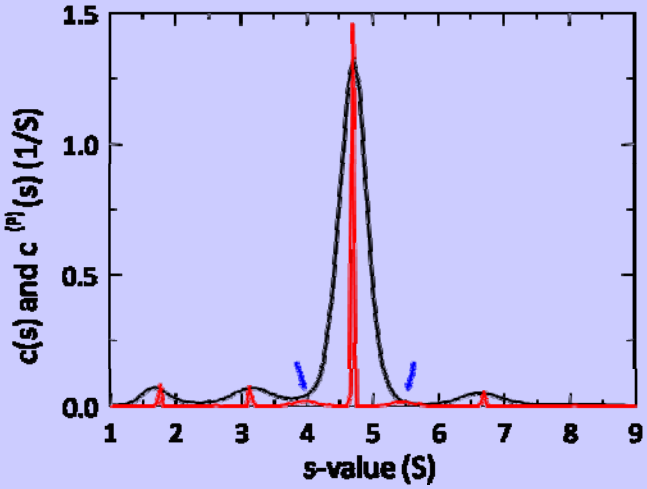


$$\alpha \sum_k c_k \log(c_k)$$

$$\alpha \sum_k c_k \log(\frac{c_k}{p_k})$$

with p_k from integration of peaks of initial distribution c_k

NKR



Applications: Multi-Signal SV Analysis

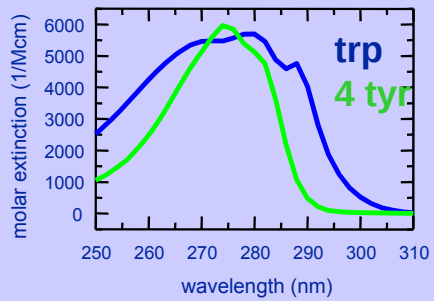
with multiple signals $a_\lambda(r,t)$: sedimentation coefficient distributions of component k

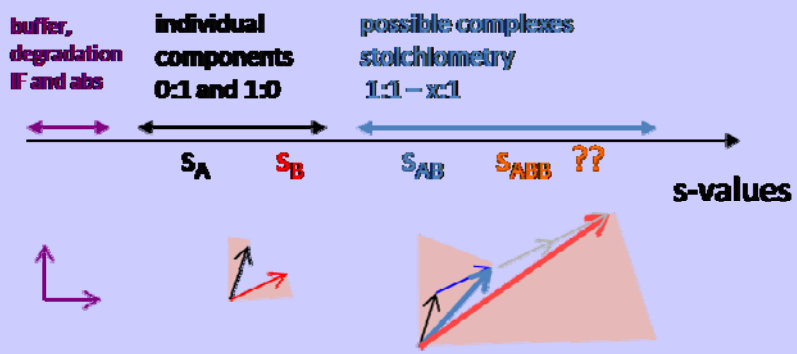
$$a_\lambda(r,t) \cong \sum_{k=1}^{K} \varepsilon_{k\lambda} \int_{s_{\min}}^{s_{\max}} \hat{c}_k(s)\chi_1(s, f_{r,w}, r, t)ds \qquad \lambda = 1...\Lambda,\ K \le \Lambda$$

$a_\lambda(r,t)$ sedimentation signal at wavelength λ (or RI)

$\varepsilon_{k\lambda}$ extinction coefficient of component k at signal λ

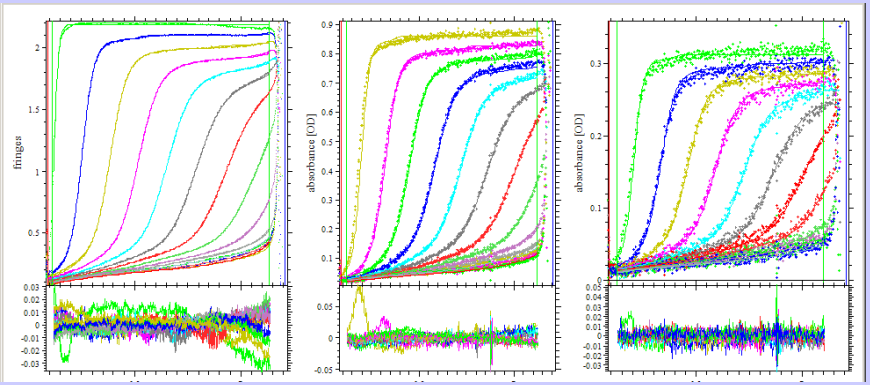
- different fraction of tryptophan residues will give characteristic RI / ε_{280}
- ratio of tyrosine and tryptophan → ε_{280} / ε_{250}
- glycosylation will contribute to signal RI / ε_{280}
- chromophoric labels
- different spectral decomposition in different s-ranges





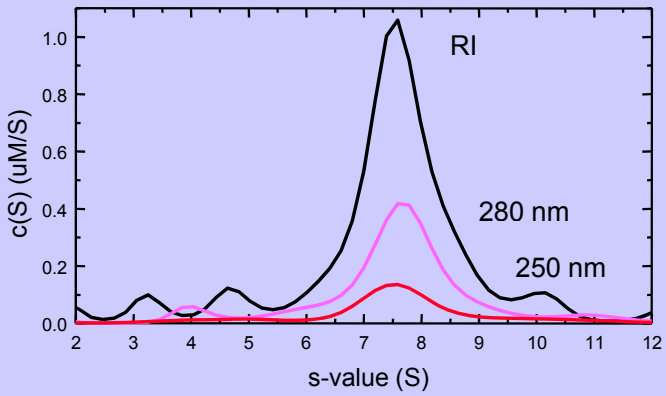
PNAS 102 (2005) 81

mix IgG and aldolase

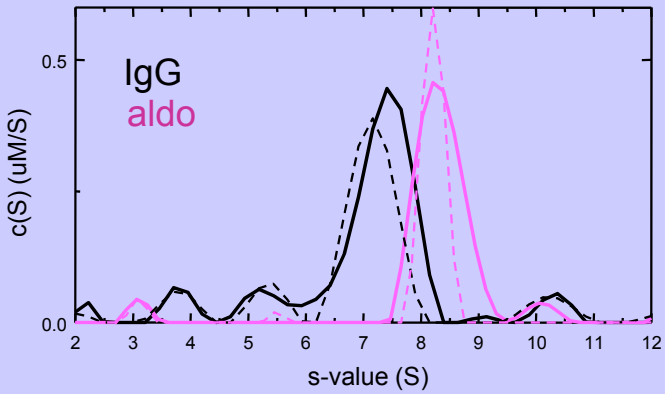


RI **280 nm** **250 nm**

standard c(s) for each λ separately



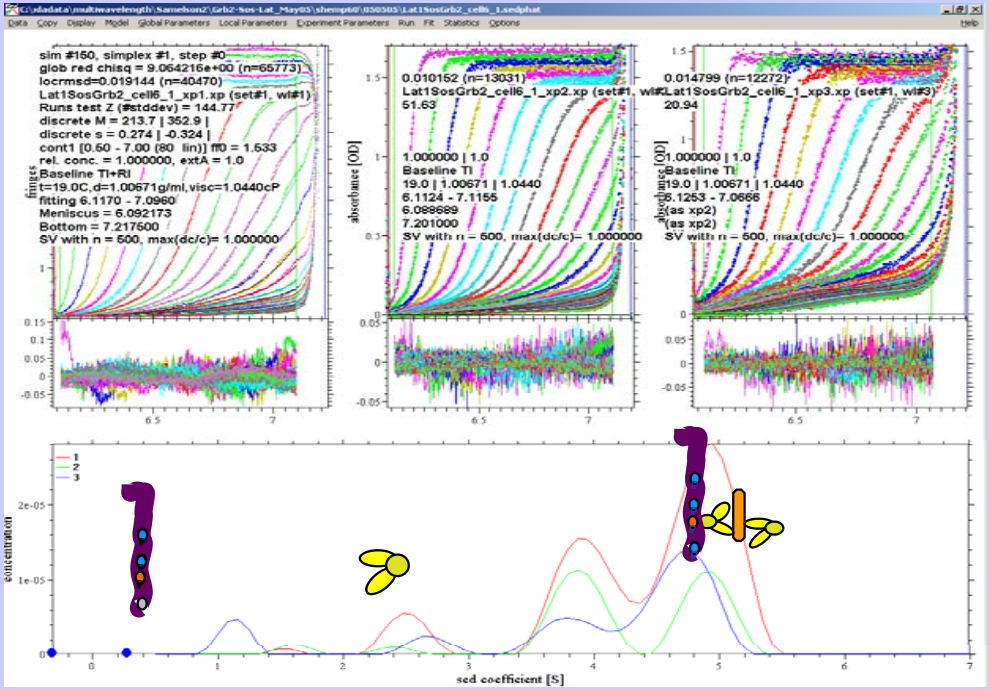
multi-λ $c_k(s)$ from mixture

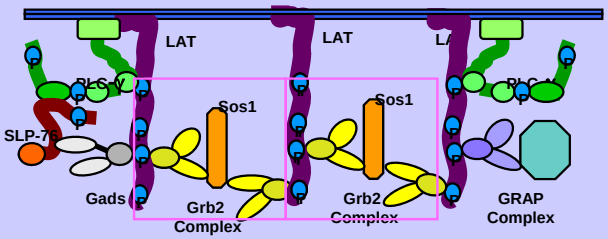


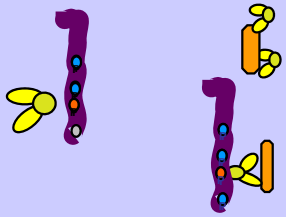
Spectral resolution → hydrodynamic resolution!

binding of Grb2 and Sos1 to LAT1pY

Grb2 24 μM
LAT1pY 12 μM
Sos1 12 μM







Nat Struct Mol Biol 2006, 13:798

Applications: Multi-Method Size-And-Shape Analysis

Sedimentation Velocity

$$a_{SVEL}(r,t) \cong C_{SVEL} \iint_{s,D} c(s,D) L(s,D,r,t)\, ds dD$$

$$g^{(1)}(\tau) \cong \gamma_{LS} \iint_{s,D} c(s,D) M(s,D) \exp\left\{-Dq^2\tau\right\} ds dD$$

$$a_{EQ}(r,\omega) \cong C_{EQ} \iint_{s,D} c(s,D) E(M,\omega,r)\, ds dD$$

$$L: \quad Lamm \;\; Eq.\, Sol.$$

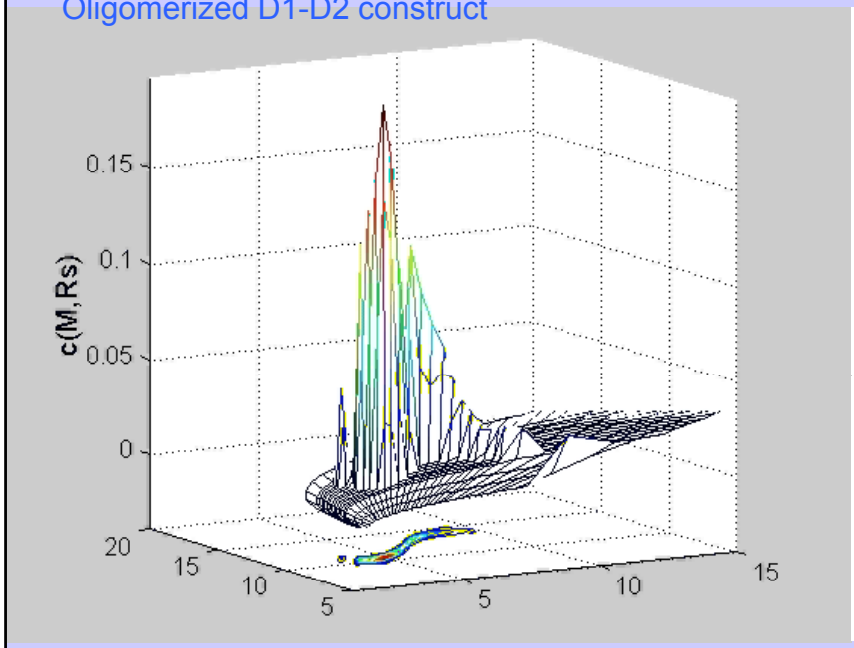
$$M(s,D) = \frac{sRT}{D(1-\bar{v}\rho)}$$

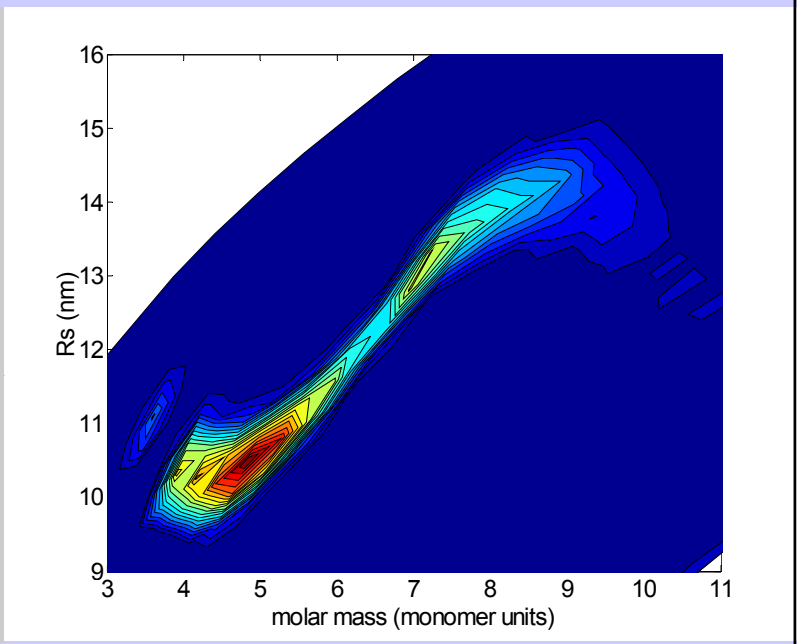
$$E(M,\omega,r) = k_{mb} e^{M(1-\bar{v}\rho)\omega^2 \frac{r^2 - r_0^2}{2RT}}$$

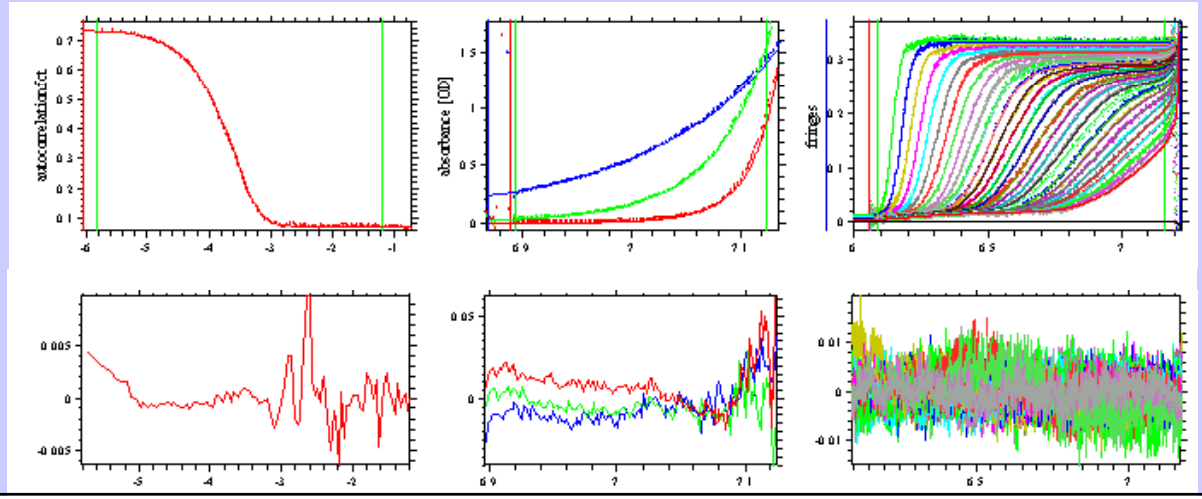
calculate as

$$\underset{C_e, c_{i,j}}{\text{Min}} \left\{ \sum_e \sum_{n_e} \frac{1}{\sigma_{n_e}} \left[a_{n_e}^{(e)}(x_{n_e}^{(e)}) - C_e \sum_{i,j} c_{ij} K^{(e)}(s_i, D_j, x_{n_e}^{(e)}) \right]^2 + \alpha \sum_{i,j} c_{ij} \ln c_{ij} \right\}$$

Oligomerized D1-D2 construct



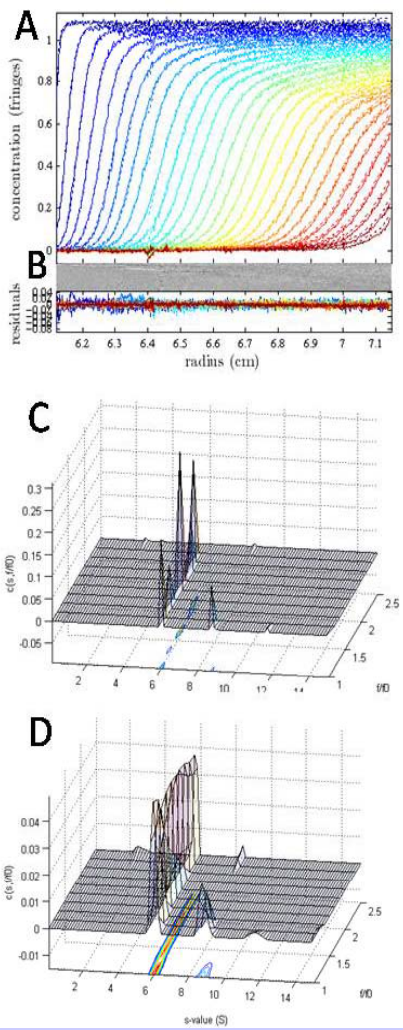




One of Many Problems: Solution With Fewest Discrete Species?

Sedimentation Velocity

IgG



data fit as size-and-shape distribution

$$a(r,t) = \int_{s_{\min}}^{s_{\max}} \int_{f_{\min}}^{f_{\max}} c\left(s, f_r\right) \chi_1\left(s, f_r, r, t\right) ds df_r$$

no regularization
→ lack of resolution of diffusion dimension justifies scaling law
→ but may not be strictly fulfilled

with Tikhonov regularization
→ broad distribution, incompatible with single species
→ "Bayesian prior" regularization method sometimes too rough
→ prior assumption of few discrete species often reasonable
→ susceptibility to systematic errors?