

# H

## H/D Exchange

Kandala V. R. Chary and Atul K. Srivastava  
Department of Chemical Sciences, Tata Institute of  
Fundamental Research, Mumbai, India

### Synonyms

[Hydrogen exchange](#)

### Definition

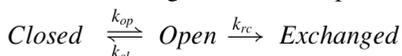
Hydrogen exchange is an interchange of hydrogen atom of a molecule (protein, nucleic acid, etc.) with that of the solvent. It is catalyzed by both acid ( $\text{H}_3\text{O}^+$ ) and base ( $\text{OH}^-$ ).

In general, the hydrogen exchange rate constants ( $k_{rc}$ ) vary with temperature and pH. When  $\log(k_{rc})$  is plotted against pH, it gives rise to a V-shaped curve with a minimum around pH 3.0. In a random coil peptide, the  $k_{rc}$  of an individual amide proton ( $^1\text{H}^{\text{N}}$ ) further depends on the amino acid sequence (due to the inductive and steric effects of neighboring residues). All these effects can be taken together to predict respective  $k_{rc}$  of any given protein sequence.

In folded proteins, the observed  $^1\text{H}^{\text{N}}$  exchange rates ( $k_{obs}$ ) are much smaller than the predicted  $k_{rc}$ , and such retardation in exchange process is expressed in terms of *protection factor*, P ( $k_{rc}/k_{obs}$ ) that can range from a value 1 (for completely unprotected) to  $10^{10}$  or higher for significantly protected one. The slow exchange is due to the exclusion of solvent (and hydrogen exchange catalyst) from the protein core or/and hydrogen bonding.

To exchange, the protected amide hydrogens must transiently expose to the solvent. This opening event may include small fluctuation, cooperative partial unfolding, or even global unfolding of the protein. These states, capable of exchange, represent a higher energy conformation within the native state ensemble, and this very reason makes the hydrogen exchange a powerful technique to study the details of protein native ensemble. The technique is also used to study the partially unfolded intermediates (kinetic and equilibrium), even when their population is too small to be detected by conventional biophysical methods.

The exchange reaction for a protected  $^1\text{H}^{\text{N}}$  can be modeled using a microscopic two-state model:



Under steady state approximation for the rare *Open* state, the observed exchange rate constant can be given

as,  $k_{obs} = \frac{k_{op} k_{rc}}{k_{cl} + k_{rc}}$ . If  $k_{cl} < k_{rc}$ ,  $k_{obs} = k_{op}$ ; this case is called EX1 mechanism. Contrary to that, if  $k_{cl} > k_{rc}$ , it is called EX2 mechanism. In the later case, opening rarely results in exchange and making it a preequilibrium step. Assuming  $k_{op} < k_{cl}$  (a rare population of *Open* state at equilibrium), the observed

exchange rate is given by  $k_{obs} = \frac{k_{op} k_{rc}}{k_{cl}} = K_{op} k_{rc}$ .

$K_{op}$  is the equilibrium constant for the opening event. In the EX2 regime, the free energy change required for the opening event ( $\Delta G_{HX}$ ) can be calculated by  $\Delta G_{HX} = -RT \ln K_{op} = -RT \ln \frac{k_{obs}}{k_{rc}} = RT \ln P$ .

Thus, hydrogen exchange can report on both the kinetic (EX1 regime) and thermodynamic (EX2 regime) aspects of any given protein.

Hydrogen exchange can be detected using different isotopes (<sup>2</sup>H and <sup>3</sup>H). <sup>2</sup>H<sub>2</sub>O is mostly used because of its low cost and nonradioactive nature. The exchange is initiated either by dissolving the lyophilized protein in <sup>2</sup>H<sub>2</sub>O or by rapidly exchanging the protons with deuterium of the solvent. The <sup>2</sup>H incorporated into the molecules are detected by FTIR, UV, NMR, or Mass spectrometry. The latter two techniques are mostly employed to detect hydrogen exchange in biomolecules. The hydrogen exchange data can reveal the details of structure and dynamics of a native protein and partially unfolded intermediates. It is also used to study the protein folding and various kinds of biomolecular interactions and aggregation.

## Cross-References

- ▶ [H/D Exchange](#)
- ▶ [Infrared Spectroscopy of Protein Dynamics: H/D Exchange](#)
- ▶ [Infrared Spectroscopy of Protein Folding, Misfolding and Aggregation](#)
- ▶ [Protein Folding](#)

## H<sup>+</sup>-Lactose Membrane Transport Protein, LacY

Lan Guan<sup>1</sup> and H. Ronald Kaback<sup>2</sup>

<sup>1</sup>Department of Cell Physiology & Molecular Biophysics, Center of Membrane Protein Research, Texas Tech University Health Sciences Center, Lubbock, TX, USA

<sup>2</sup>Departments of Physiology and Microbiology, Immunology & Molecular Genetics, Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA

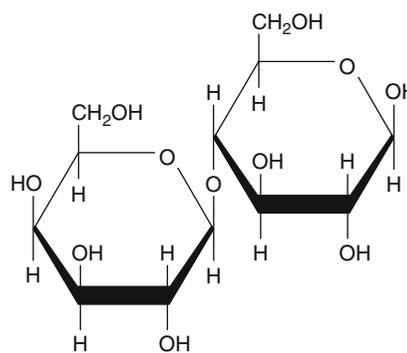
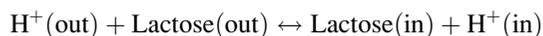
## Synonyms

[Lac carrier protein](#); [Lactose permease](#)

## Definition

LacY (*aka* Lac permease) is the product of the *lacY* gene, the second structural gene in the *lac* operon, which encodes a single polypeptide that is solely

responsible for all the translocation reactions catalyzed by the β-galactoside transport system in *E. coli*. LacY (TCDB 2.A.1.5.1) belongs to the oligosaccharide/H<sup>+</sup> symport subfamily of the major facilitator superfamily (MFS), which contains >40,000 members found ubiquitously in all living organisms. LacY is selective for disaccharides containing a D-galactopyranosyl ring, as well as D-galactose, but has no affinity for D-glucopyranosides or D-glucose. LacY catalyzes the vectorial reversible reaction:

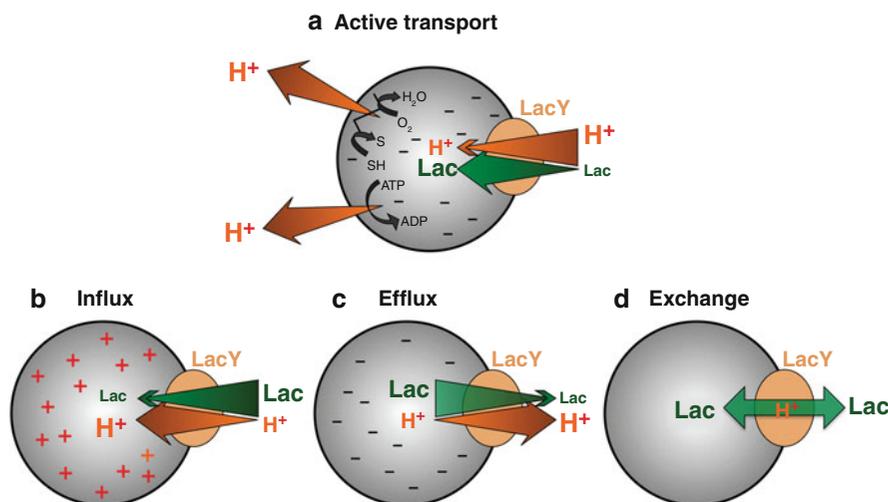


Structure of lactose

## Introduction

The *lacY* gene is the first gene encoding a membrane transport protein to be cloned into a recombinant plasmid and sequenced as well as the first symporter to be purified to apparent homogeneity in a completely functional state and to have a high-resolution crystal structure determined by X-ray diffraction (Kaback et al. 2001; Guan and Kaback 2006). The hydrophobic polypeptide is composed of 417 amino acid residues, has a mass of 46,517 Da, but electrophoreses with an  $M_r$  of ~33 kDa due to excess dodecyl sulfate binding. As a particularly well-studied sugar/H<sup>+</sup> symporter, LacY is a paradigm for exploring the molecular mechanism of solute/cation symport catalyzed by many secondary transport proteins.

LacY is an effective chemiosmotic machine that catalyzes the coupled translocation of an H<sup>+</sup> and a galactopyranoside (galactoside/H<sup>+</sup> symport). Since translocation is obligatorily coupled, sugar accumulation against a concentration gradient is achieved by transduction of the free energy released from the downhill



**H<sup>+</sup>-Lactose Membrane Transport Protein, LacY, Fig. 1** Lactose/H<sup>+</sup> symport. (a) Active transport. The electrochemical H<sup>+</sup> gradient ( $\Delta\tilde{\mu}_{H^+}$ ) across the cytoplasmic membrane of *E. coli* is generated by pumping of H<sup>+</sup> via the respiratory chain or through the hydrolytic activity of F<sub>1</sub>F<sub>0</sub> ATPase. Free energy released from the downhill movement of H<sup>+</sup> catalyzed by LacY (sand color) is coupled to the uphill accumulation of lactose as indicated by the direction of the arrows and font size. (b) Influx

and (c) Efflux. Energetically downhill lactose transport generates  $\Delta\tilde{\mu}_{H^+}$ , the polarity of which depends upon the direction of lactose concentration gradient (influx generates a  $\Delta\tilde{\mu}_{H^+}$  that is interior positive and acid; efflux generates a  $\Delta\tilde{\mu}_{H^+}$  that is interior negative and alkaline). (d) Equilibrium exchange. At equal intra- and extra-cellular lactose concentrations, lactose exchange across the membrane is catalyzed by protonated LacY

movement of H<sup>+</sup> with the electrochemical H<sup>+</sup> gradient ( $\Delta\tilde{\mu}_{H^+}$ ; interior negative and/or alkaline) (Fig. 1a). Conversely, downhill sugar translocation by LacY drives uphill translocation of H<sup>+</sup> with the generation of  $\Delta\tilde{\mu}_{H^+}$ , the polarity of which depends upon the direction of the sugar concentration gradient (Fig. 1b, c). Notably, LacY also catalyzes equilibrium exchange of internal radiolabeled galactoside for external unlabeled sugar substrate or the reverse reaction (i.e., counterflow) in a manner that is completely independent of  $\Delta\tilde{\mu}_{H^+}$  (Fig. 1d). Thus, the primary driving force for the global conformational change involved in transport (alternating access of the binding sites to either side of the membrane) is due to association and dissociation of sugar by protonated LacY. Without a galactoside, LacY does not translocate H<sup>+</sup> in the presence of  $\Delta\tilde{\mu}_{H^+}$ , and unprotonated LacY does not bind galactosides (i.e., binding of both substrates is required for symport).

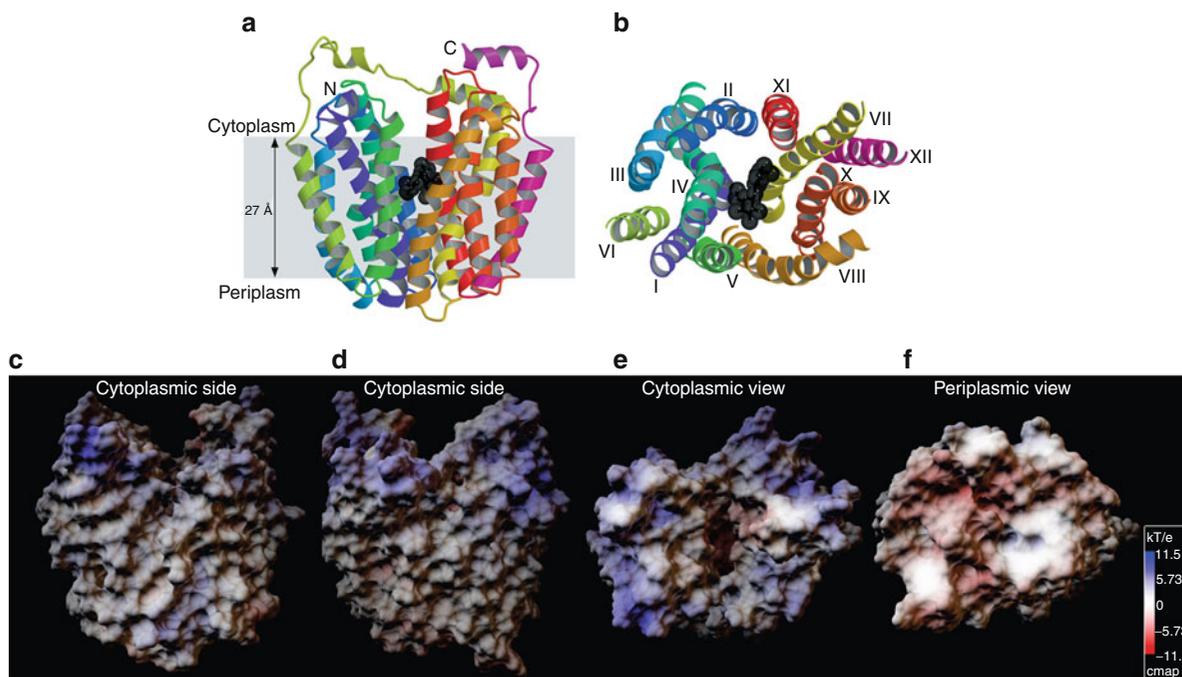
by relatively hydrophilic loops with both N- and C-termini on the cytoplasmic side of the membrane. In the X-ray crystal structures (Abramson et al. 2003; Mirza et al. 2006; Guan et al. 2007; Chaptal et al. 2011), LacY is in an inward-facing conformation. Viewed parallel to the membrane (Fig. 2a, c, d), the molecule is heart-shaped with a cavernous water-filled cleft  $25 \times 15 \text{ \AA}$  and open on the cytoplasmic side only; the periplasmic side is tightly sealed. The largest dimensions of the molecule are  $60 \times 60 \text{ \AA}$ . Viewed normal to the membrane from the cytoplasmic side (Fig. 2b, e), the molecule has a distorted oval shape with dimensions of  $30 \times 60 \text{ \AA}$ . The electrostatic surface potential calculated from the crystal structure reveals a positively charged belt around the periphery of the cytoplasmic opening (Fig. 2c, d, e) (Yousef and Guan 2009) with negatively charged residues distributed preferentially on the periplasmic side of the molecule (Fig. 2f). Thus, LacY follows the positive inside rule.

### X-ray Crystal Structure of the LacY Protein

LacY is an  $\sim 86\%$   $\alpha$ -helical membrane protein and contains 12 mostly irregular transmembrane helices traversing the membrane in zig-zag fashion connected

### Helix Packing and Pseudosymmetry

LacY is organized into two 6-helix bundles connected by a long loop between helices VI and VII (Fig. 2a, b).



**H<sup>+</sup>-Lactose Membrane Transport Protein, LacY, Fig. 2** X-ray crystal structures (PDB, 1pv7). Crystal structures of all LacY molecules determined thus far (PDB, 1pv7, 1pv6, 2cfp, 2cfq, 2v8n, and 2y5y) exhibit an inward-facing conformation. In the *ribbon* representations of LacY (a, b), the 12 transmembrane helices are colored from the N-terminus in *blue* to the C-terminus in *purple*. The lactose homologue thiodigalactoside is shown as *black spheres*. (a) Viewed parallel to the membrane with helices V and VIII in front. The membrane

is represented in *light gray*. (b) Cytoplasmic view normal to the membrane. Helices are labeled with Roman numerals. (c–f) Electrostatic surface potential of LacY molecule. The scale indicates color-coded values of the electrostatic potentials (kT/e). (c) and (d), viewed parallel to the membrane, where helices V/VIII and II/XI are in front, respectively, and the cytoplasmic side is on *top*; (e) and (f), viewed from cytoplasmic and periplasmic sides, respectively

The hydrophilic cavity is lined by helices I, II, IV, and V of the N-terminal domain and helices VII, VIII, X, and XI of the C-terminal domain. Helices III, VI, IX, and XII are largely embedded in the membrane. The cavity-lining helices I, IV, VII, and X are distorted with kinks, thereby providing a structural basis for the large-scale conformation changes observed in LacY. The N- and C-terminal 6-helix domains have the similar topology and exhibit twofold pseudosymmetry. Within each domain, there are two 3-helix repeats with an inverted topology; these symmetrical features may provide a foundation for the alternating access mechanism in LacY.

### Sugar Binding and Specificity

A LacY mutant with a single native Cys residue at position 148 (helix V) binds the monosaccharide

galactose, the disaccharide lactose, or 4-nitrophenyl- $\alpha$ -D-galactopyranoside ( $\alpha$ -NPG) with apparent dissociate constants ( $K_D$ ) of 30 mM, 1 mM, or 20  $\mu$ M, respectively (Sahin-Toth et al. 2000; Sahin-Toth et al. 2001; Guan and Kaback 2004). Galactose is the most specific substrate for LacY with lowest affinity, and various adducts – particularly if they are hydrophobic at the anomeric carbon C-1 of the galactopyranosyl ring, in the  $\alpha$  configuration or both – increase affinity with no effect on specificity. The C<sub>4</sub> OH is unequivocally the most important determinant for specificity, and the C<sub>3</sub>, C<sub>6</sub>, and C<sub>2</sub> OH groups on the galactopyranosyl ring play relatively minor roles in H-bonding.

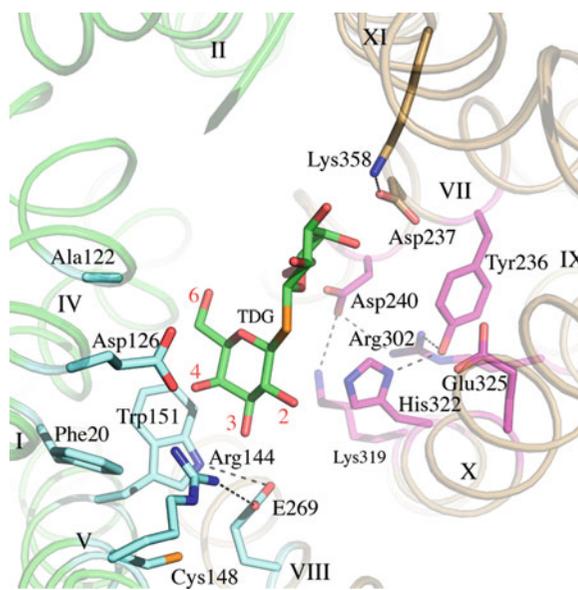
Cys-scanning and site-directed mutagenesis show that only a small number of amino acyl side chains located in transmembrane helices are irreplaceable with respect to active lactose transport (Frillingos et al. 1998). Among them, the charge pair Glu126 (helix IV)/Arg144 (helix V) is essential for galactoside

binding and specificity. An aromatic side chain at position 151 (preferably Trp) in helix V is also critically involved in sugar affinity by stacking hydrophobically with the galactopyranosyl ring and forming an H-bond with Glu269 (helix VIII), another irreplaceable residue also critical for binding and transport. Replacement of Glu126 with Asp or Trp151 with Tyr decreases affinity for galactoside significantly with little effect on apparent affinity for H<sup>+</sup>.

These residues are located at the apex of the cavity approximately halfway through the membrane. Biochemical findings are consistent with the notion that Arg144 (helix V) may interact with the O<sub>3</sub> and O<sub>4</sub> atoms of the galactopyranosyl ring and that Glu126 is in proximity to the O<sub>4</sub>, O<sub>5</sub>, or O<sub>6</sub> atoms of the galactopyranosyl ring. Spectroscopic studies show that Trp151 stacks with the hydrophobic face of the galactopyranosyl ring, and this interaction probably orients the ring so that important H-bonds can be realized. Glu269 in the C-terminal domain is proximal to the O<sub>3</sub> atom of the galactopyranosyl ring, and may form a salt-bridge or an H-bond with Arg144 or Trp151, respectively. In the absence of ligand, Arg144 is displaced from Glu269 and forms a salt bridge with Glu126. Most residues involved in H<sup>+</sup> binding and translocation in the C-terminal domain also play a role in the affinity of LacY for sugar.

### The H<sup>+</sup>-Binding Site

LacY exhibits a remarkably high pK<sub>a</sub> of 10.5 for sugar-binding affinity (Smirnova et al. 2009a). Therefore, at physiological pH, LacY is fully protonated, and galactoside binds to protonated LacY only. The H<sup>+</sup>-binding site is located in the C-terminal helix bundle, aligned parallel to the plane of the membrane at the similar level as the sugar-binding site (Figs. 3 and 4), and exposed to the water-filled cavity. Like the galactosidic sugar, the H<sup>+</sup> may bind and be released directly into the cavity during turnover. In any event, it is apparent that H<sup>+</sup> translocation through LacY does not involve a water-filled channel through the molecule. These structural features may also explain why LacY is able to catalyze lactose/H<sup>+</sup> symport in both directions across the membrane utilizing the same residues. Residues Tyr236 and Asp240 (helix VII), Glu269 (helix VIII), Arg302 (helix IX), and His322



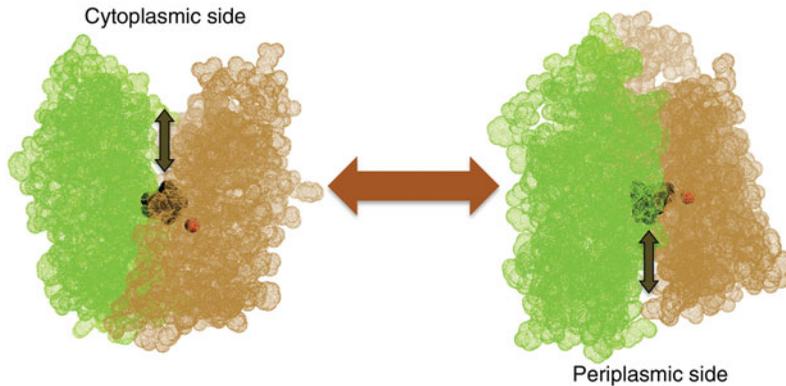
**H<sup>+</sup>-Lactose Membrane Transport Protein, LacY, Fig. 3** The cytoplasmic cavity viewed normal to the membrane with residues important for galactoside and H<sup>+</sup> binding (PDB 1pv7). The N- and C-terminal helices of LacY are colored *green* and *sand*, respectively, and labeled with Roman numerals. Important residues identified by biochemical and biophysical studies for the binding of sugar and/or H<sup>+</sup> are shown in *cyan* and *magenta*, respectively. *Dotted lines* show salt bridge/H-bond interactions. The lactose homologue thiodigalactoside is shown in *green*. The O<sub>2-4</sub> and O<sub>6</sub> atoms on the galactopyranosyl ring are indicated

and Glu325 (helix X) contribute to the H<sup>+</sup> binding and translocation (Kaback et al. 2001), and most are part of a charge-pair/H-bond network (Abramson et al. 2003; Mirza et al. 2006; Guan et al. 2007; Chaptal et al. 2011). Mutation of some of these residues also causes a marked decrease in sugar affinity (Smirnova et al. 2009a).

Mutagenesis studies demonstrate that Asp237 (Helix VII) and Lys358 (helix XI), in the vicinity of the sugar- and H<sup>+</sup>-binding sites, form a salt bridge that is not involved in sugar or H<sup>+</sup> binding but is important for insertion of LacY into the membrane and its stability (Kaback et al. 2001).

### Oligomeric State of LacY

LacY is monomeric in the detergent dodecylmaltoside or in the solvent hexamethylphosphorus triamide, and the rotational diffusion coefficient determined from fluorescence anisotropy with eosinylmaleimide-labeled



**H<sup>+</sup>-Lactose Membrane Transport Protein, LacY, Fig. 4** The alternating access model. An inward-facing crystal structure (PDB, 1pv7) and an outward-facing model of LacY are shown in *dots* colored as *green* or *orange* for the N- or C-terminal helix

bundles, respectively. A sugar substrate is shown as *black spheres*; a *red sphere* indicates positions important for H<sup>+</sup> binding. *Black arrows* indicate binding and release of galactoside and H<sup>+</sup> from cytoplasmic or periplasmic cavities as indicated

permease indicates that the polypeptide reconstitutes as a monomer. Freeze-fracture electron microscopic studies on proteoliposomes reconstituted with purified LacY demonstrate that the protein reconstitutes as a monomer and no change is observed in size or distribution of the particles when a  $\Delta\tilde{\mu}_{\text{H}^+}$  is imposed. Functionally, the initial rate of  $\Delta\tilde{\mu}_{\text{H}^+}$ -driven lactose transport in proteoliposomes reconstituted at very low protein-to-lipid ratios varies linearly with the ratio of permease to phospholipid. An engineered fusion protein containing two LacY molecules covalently linked in tandem has high transport activity and does not exhibit negative dominance with either mutations or chemical modification of either half of the molecule. Finally, dimerization of single-Cys mutants in LacY in the plane of the membrane is a stochastic process. Thus, LacY is both structurally and functionally a monomer (reviewed in Guan and Kaback 2006).

### The Alternating Access Model

As indicated, the sugar- and H<sup>+</sup>-binding sites in LacY are located at the approximate middle of the molecule at the apex of a hydrophilic cavity and inaccessible from the periplasm. However, LacY is a highly dynamic molecule, and sugar/H<sup>+</sup> symport involves a global conformational change in which the cytoplasmic cavity closes with the opening of a complementary periplasmic cavity, thereby allowing alternating access of the sugar- and H<sup>+</sup>-binding sites to either

side of the membrane (Fig. 4). Experimental findings from site-directed alkylation of single-Cys replacement mutants, thiol cross-linking, single molecule fluorescence resonance energy transfer, double electron-resonance spectroscopy, and Trp fluorescence quenching and unquenching provide strong converging evidence that supports the alternating access mechanism (Smirnova et al. 2011). Like other MFS members, the internal pseudosymmetry within each helix bundle may provide the structural foundation for this mechanism. Thus, swapping the topology of two inverted 3-helix repeats in each 6-helix bundle may be the basis for the reciprocal formation of inward- and outward-facing conformers (Radestock and Forrest 2011).

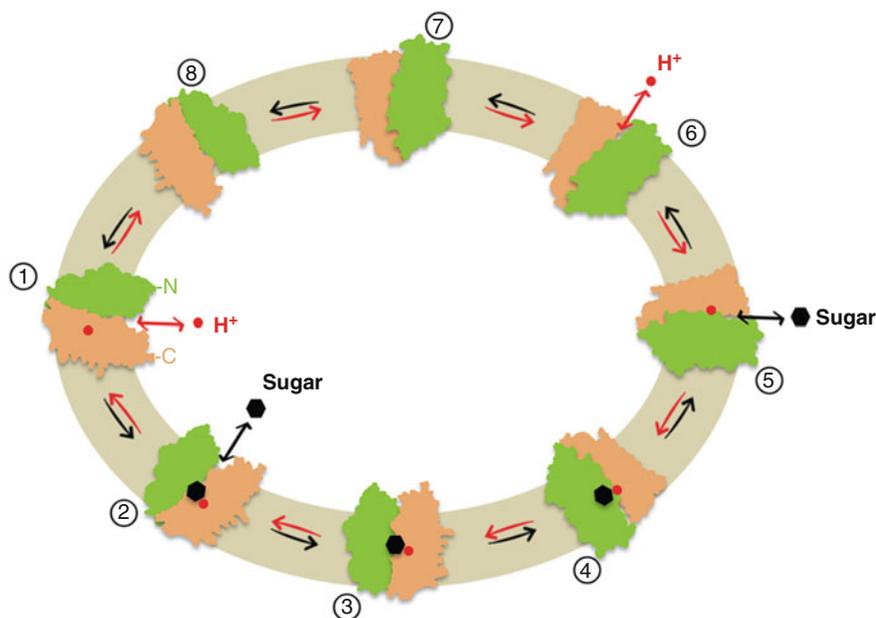
While all X-ray crystal structures obtained thus far reveal the same inward-facing conformation for LacY, results from double electron–electron resonance studies (Smirnova et al. 2007) and other findings reveal that LacY exhibits multiple conformers. In the absence of a sugar substrate, the inward-facing conformation dominates in the membrane and after solubilization in detergent. Sugar binding increases the probability of forming intermediate and outward-facing conformations.

### The H<sup>+</sup>-lactose Symport Mechanism

Because the individual steps in the overall translocation cycle cannot be delineated by studying  $\Delta\tilde{\mu}_{\text{H}^+}$ -driven

### H<sup>+</sup>-Lactose Membrane Transport Protein, LacY,

**Fig. 5** The H<sup>+</sup>-lactose symport mechanism. The cartoon representations for the N- and C-terminal domains of LacY molecule are colored as green and sand, respectively. Sugar and H<sup>+</sup> are shown as a black hexagonal and a red sphere, respectively. Sugar and H<sup>+</sup> bind in the middle of the molecule, and LacY moves around the binding sites to release the symported substrates on either side of the membrane as indicated by the arrows



active transport, downhill lactose efflux, equilibrium exchange, and entrance counterflow (Fig. 1) are used to probe the mechanism in right-side-out (RSO) membrane vesicles (Kaczorowski and Kaback 1979). Efflux, exchange, and counterflow with wild-type LacY are explained by the simple kinetic scheme shown in Fig. 5. Lactose efflux down a sugar concentration gradient consists of eight steps (black arrows):

1. Protonation in the inward-facing conformer
2. Binding of a preloaded lactose on the cytoplasmic side of the membrane
3. Formation of a *loaded* intermediate that is closed on both sides
4. Opening of the periplasmic cavity
5. Release of lactose
6. Deprotonation
7. Formation of an *unloaded* intermediate that is closed on both sides
8. Opening of the cytoplasmic cavity

Lactose influx down a sugar concentration gradient occurs in identical fashion as efflux; the reaction starts at step (6) and proceeds via the red arrows around the circle. For equilibrium exchange and counterflow, deprotonation does not occur, and only steps (1–5) are involved. The ordered binding and release of H<sup>+</sup> and galactoside are strongly supported by the behavior of mutants such as E325A that cannot catalyze any

transport reactions that involve net H<sup>+</sup> translocation, but catalyze equilibrium exchange and counterflow as well or better than WT LacY (Carrasco et al. 1986).

Certain enzyme reactions involve H<sup>+</sup> transfer as the rate-limiting step; as a result, these reactions may exhibit a solvent isotope effect when studied in deuterium oxide (D<sub>2</sub>O). In brief, such reactions proceed slower in D<sub>2</sub>O because of differences in the zero-point stretch vibrations of bonds with protium relative to deuterium. With RSO membrane vesicles or proteoliposomes reconstituted with purified LacY, over threefold slowing of the rate of lactose influx or efflux from pH 5.5 to 7.5 is observed, with no effect on exchange, counterflow,  $\Delta\tilde{\mu}_{\text{H}^+}$ -driven active transport or affinity for sugar (Viitanen et al. 1983). These and other observations indicate that reactions involved in H<sup>+</sup> transfer are rate determining when a lactose gradient drives H<sup>+</sup> translocation. Conversely, H<sup>+</sup> transfer is not rate limiting for exchange or counterflow. Thus, in the absence of  $\Delta\tilde{\mu}_{\text{H}^+}$ , the limiting step in the transport mechanism is deprotonation, *step (6) during efflux and step (1) during influx*.

In the presence of  $\Delta\tilde{\mu}_{\text{H}^+}$  (interior negative and/or alkaline), LacY drives accumulation of lactose against a 50- to 100-fold concentration gradient (Robertson et al. 1980). Both  $\Delta\Psi$  (interior negative) and  $\Delta\text{pH}$  (interior alkaline) have quantitatively the same kinetic

effect on transport, a marked decrease in  $K_m$ , as well as the same thermodynamic effect. However, in opposition to the time-honored conjecture, the  $K_D$  of LacY for galactosides on either side of the membrane is similar and unchanged in the absence or presence of  $\Delta\tilde{\mu}_{H^+}$  (Guan and Kaback 2004). In the membrane, an inward-facing conformation of LacY predominates (Smirnova et al. 2011). In the absence of an external galactoside, opening of protonated LacY on the periplasmic side occurs spontaneously, but with very low probability (low frequency). Thus, at any given time, only a small population of molecules in the membrane is open on the periplasmic side statistically. While the sugar-induced opening rate of periplasmic cavity is  $\sim 20 \text{ s}^{-1}$  (Smirnova et al. 2009b), the pre-steady-state conformation may not affect the steady-state transport rate.

Active transport of lactose against a concentration gradient proceeds from step (6) (Fig. 5) via the red arrows in the same manner as lactose influx. Binding of galactoside from the periplasm (5) by protonated LacY (6) results in formation of a high-energy intermediate (4) in which sugar and H<sup>+</sup> are bound and both cavities are closed. In the presence of  $\Delta\tilde{\mu}_{H^+}$ , the intermediate with a bound sugar opens to the cytoplasm (3), sugar dissociates (2) followed by deprotonation (1) and the unloaded LacY returns to the outward-facing conformation (steps 8, 7) and protonates immediately. The turnover number for  $\Delta\tilde{\mu}_{H^+}$ -driven lactose transport by LacY in RSO membrane vesicles or with the purified protein reconstituted into proteoliposomes is  $\sim 20 \text{ s}^{-1}$ . The reaction occurs without a change in  $K_D$  and is totally unaffected by D<sub>2</sub>O. Thus, when there is a driving force on the H<sup>+</sup>, protonation or deprotonation is no longer rate limiting, and either dissociation of sugar or a conformational change becomes the limiting step in the transport reaction. It is particularly noteworthy that a number of experimental findings indicate strongly that  $\Delta\tilde{\mu}_{H^+}$  functions kinetically as a driving force on the H<sup>+</sup> with little or no direct effect on LacY (Guan and Kaback 2006).

## Summary

LacY, a sugar/H<sup>+</sup> symporter that belongs to the major facilitator superfamily, is a particularly well-studied

paradigm for exploring the molecular mechanism of solute/cation symport catalyzed by many secondary transport proteins. X-ray crystal structures show that LacY is organized into two pseudosymmetrical 6-helix bundles surrounding a cavernous cleft open to the cytoplasm only, and within each bundle, there are two repeats with an inverted topology. Positions responsible for binding of galactoside and H<sup>+</sup> are located in the approximate middle of the molecule halfway through the membrane. Structural symmetries allow LacY to open and close periplasmic and cytoplasmic cavities reciprocally, thereby allowing alternating access of the binding sites to either side of the membrane and the coupled translocation of galactoside and H<sup>+</sup> in both directions. An electrochemical H<sup>+</sup> gradient ( $\Delta\tilde{\mu}_{H^+}$ ; interior negative and/or alkaline) dramatically decreases the  $K_m$  for transport by 50- to 100-fold with no effect on  $K_D$  from either side of the membrane. LacY is fully protonated under physiological condition, and the primary driving force for the global conformational change in the alternating-access mechanism is binding of sugar to protonated LacY. In the absence of  $\Delta\tilde{\mu}_{H^+}$ , the limiting step for downhill lactose efflux and influx is H<sup>+</sup> transfer; in the presence of a driving force on the H<sup>+</sup>, dissociation of sugar becomes limiting with a turnover number of  $\sim 20 \text{ s}^{-1}$ .

## Cross-References

- ▶ [Membrane Transport, Energetics and Overview](#)

## References

- Abramson J, Smirnova I, Kasho V, Verner G, Kaback HR, Iwata S. Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*. 2003;301:610–5.
- Carrasco N, Antes LM, Poonian MS, Kaback HR. Lac permease of *Escherichia coli*: histidine-322 and glutamic acid-325 may be components of a charge-relay system. *Biochemistry*. 1986;25:4486–8.
- Chaptal V, Kwon S, Sawaya MR, Guan L, Kaback HR, Abramson J. Crystal structure of lactose permease in complex with an affinity inactivator yields unique insight into sugar recognition. *Proc Natl Acad Sci USA*. 2011;108:9361–6.
- Frillingos S, Sahin-Toth M, Wu J, Kaback HR. Cys-scanning mutagenesis: a novel approach to structure function

- relationships in polytopic membrane proteins. *FASEB J*. 1998;12:1281–99.
- Guan L, Kaback HR. Binding affinity of lactose permease is not altered by the H<sup>+</sup> electrochemical gradient. *Proc Natl Acad Sci USA*. 2004;101:12148–52.
- Guan L, Kaback HR. Lessons from lactose permease. *Annu Rev Biophys Biomol Struct*. 2006;35:67–91.
- Guan L, Mirza O, Verner G, Iwata S, Kaback HR. Structural determination of wild-type lactose permease. *Proc Natl Acad Sci USA*. 2007;104:15294–8.
- Kaback HR, Sahin-Toth M, Weinglass AB. The kamikaze approach to membrane transport. *Nat Rev Mol Cell Biol*. 2001;2:610–20.
- Kaczorowski GJ, Kaback HR. Mechanism of lactose translocation in membrane vesicles from *Escherichia coli*. 1. Effect of pH on efflux, exchange, and counterflow. *Biochemistry*. 1979;18:3691–7.
- Mirza O, Guan L, Verner G, Iwata S, Kaback HR. Structural evidence for induced fit and a mechanism for sugar/H<sup>+</sup> symport in LacY. *EMBO J*. 2006;25:1177–83.
- Radestock S, Forrest LR. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J Mol Biol*. 2011;407:698–715.
- Robertson DE, Kaczorowski GJ, Garcia ML, Kaback HR. Active transport in membrane vesicles from *Escherichia coli*: the electrochemical proton gradient alters the distribution of the *lac* carrier between two different kinetic states. *Biochemistry*. 1980;19:5692–702.
- Sahin-Toth M, Akhooon KM, Runner J, Kaback HR. Ligand recognition by the lactose permease of *Escherichia coli*: specificity and affinity are defined by distinct structural elements of galactopyranosides. *Biochemistry*. 2000;39:5097–103.
- Sahin-Toth M, Lawrence MC, Nishio T, Kaback HR. The C-4 hydroxyl group of galactopyranosides is the major determinant for ligand recognition by the lactose permease of *Escherichia coli*. *Biochemistry*. 2001;40:13015–9.
- Smirnova I, Kasho V, Choe JY, Altenbach C, Hubbell WL, Kaback HR. Sugar binding induces an outward facing conformation of LacY. *Proc Natl Acad Sci USA*. 2007;104:16504–9.
- Smirnova I, Kasho V, Sugihara J, Choe JY, Kaback HR. Residues in the H<sup>+</sup> translocation site define the pKa for sugar binding to LacY. *Biochemistry*. 2009a;48:8852–60.
- Smirnova I, Kasho V, Sugihara J, Kaback HR. Probing of the rates of alternating access in LacY with Trp fluorescence. *Proc Natl Acad Sci USA*. 2009b;106:21561–6.
- Smirnova I, Kasho V, Kaback HR. Lactose permease and the alternating access mechanism. *Biochemistry*. 2011;50:9684–93.
- Viitanen P, Garcia ML, Foster DL, Kaczorowski GJ, Kaback HR. Mechanism of lactose translocation in proteoliposomes reconstituted with lac carrier protein purified from *Escherichia coli*. 2. Deuterium solvent isotope effects. *Biochemistry*. 1983;22:2531–6.
- Yousef MS, Guan L. A 3D structure model of the melibiose permease of *Escherichia coli* represents a distinctive fold for Na<sup>+</sup> symporters. *Proc Natl Acad Sci USA*. 2009;106:15291–6.

## HADDOCK

Alexandre M. J. J. Bonvin, M. van Dijk, E. Karaca, P. L. Kastiritis, Adrien S. J. Melquiond, C. Schmitz and S. J. de Vries

Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, CH Utrecht, The Netherlands

## Definition

HADDOCK (High Ambiguity Driven biomolecular DOCKing) is an information-driven flexible docking approach for the modeling of biomolecular complexes (Dominguez et al. 2003). Docking is defined as the modeling of the structure of a complex based on the known three-dimensional structures of its constituents. HADDOCK distinguishes itself from other docking methods by incorporating a wide variety of experimental and/or bioinformatics data to drive the modeling (Melquiond and Bonvin 2010). This allows concentrating the search to relevant portions of the interaction space using a more sophisticated treatment of conformational flexibility.

Interface regions can be identified by mutagenesis, ► [H/D exchange](#), and chemical modifications (e.g., by cross-linkers or oxidative agents) detected by ► [mass spectrometry](#), ► [nuclear magnetic resonance](#), chemical shift perturbations, and cross-saturation transfer. When experimental data are unavailable or scarce, this information can be supplemented by bioinformatics predictions (de Vries and Bonvin 2008). These diverse information sources typically only identify or predict interfacial regions, but do not define the contacts across an interface. HADDOCK deals with this by implementing them as ambiguous interaction restraints (AIRs) that will force the interfaces to come together without imposing a particular orientation.

HADDOCK can also incorporate classical NMR restraints such as distances from ► [nuclear Overhauser effects](#) and paramagnetic relaxation enhancement measurements, pseudocontact shift, dihedral angles, ► [residual dipolar coupling](#), and diffusion anisotropy restraints, the latter two providing valuable information about the relative orientation of the components in a complex. In addition, symmetry restraints can be

defined in the case of symmetrical homomeric systems. Data derived from low-to-medium resolution techniques such as ► [small angle X-ray scattering](#), ► [cryo-electron microscopy](#), and ion mobility mass spectrometry can provide information about the shape of a complex and be used to drive the docking process (Melquiond et al. 2011).

The HADDOCK protocol, which makes use of the ► [Crystallography and NMR System \(CNS\)](#) package as computational engine, consists of three successive steps: (a) rigid-body energy minimization, (b) semi-flexible refinement in torsion angle space, and (c) final refinement in explicit solvent. By allowing for explicit flexibility during the molecular dynamics refinement, HADDOCK can account for small conformational changes occurring upon binding. Larger and more challenging conformational changes can be dealt with by starting the docking from ensembles of conformations and/or treating the molecules as a collection of domains. The latter approach makes use of the unique multibody docking ability of HADDOCK, which can handle up to six separate domains or molecules at the same time. The selection of the final models is based on a weighted sum of electrostatics, desolvation, and van der Waals energy terms, along with the energetic contribution of the restraints used to drive the docking.

HADDOCK has been extensively applied to a large variety of systems, including protein-protein, protein-nucleic acids, and protein-small molecule docking, and has shown a very strong performance in the blind critical assessment of the prediction of interactions (CAPRI). A considerable number of experimental structures of complexes calculated using HADDOCK have been deposited into the Protein Data Bank (PDB). HADDOCK is available as a web server (<http://haddock.chem.uu.nl/services/HADDOCK>) (de Vries et al. 2010) offering a user-friendly interface to the structural biology community.

## Cross-References

- [CNS \(Crystallography and NMR System\)](#)
- [Double-Mutant Cycle Analysis](#)
- [Electron Microscopy](#)
- [H/D Exchange](#)
- [Ion Mobility Mass Spectrometry – Principles](#)
- [Kinetics: Relaxation Methods](#)

- [Lipid Lateral Diffusion](#)
- [Mass Spectrometry](#)
- [Molecular Dynamics Simulations of Lipids](#)
- [NMR](#)
- [NMR Studies of Macromolecular Interactions – Introduction](#)
- [Protein Complexes](#)
- [Residual Dipolar Coupling](#)
- [X-Ray Scattering of Lipid Membranes](#)

## References

- de Vries SJ, Bonvin AMJJ. How proteins get in touch: interface prediction in the study of biomolecular complexes. *Curr Prot & Pept Sci.* 2008;9:394–406.
- de Vries SJ, van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc.* 2010;5:883–97.
- Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* 2003;125:1731–7.
- Melquiond ASJ, Bonvin AMJJ. Data-driven docking: using external information to spark the biomolecular rendez-vous. In: Zacharias M, editor. *Protein-protein complexes: analysis, modeling and drug design.* London: Imperial College; 2010. p. 183–209.
- Melquiond ASJ, Karaca E, Kastriitis PL, Bonvin AMJJ. Next challenges in protein-protein docking: from proteome to interactome and beyond. *Wiley Interdisciplinary Reviews: Computational Molecular Science.* 2011; Early View doi: 10.1002/wcms.91

---

## Haemagglutinin

- [Haemagglutinin – Computational Studies](#)

---

## Hand-Over-Hand Mechanism

- [Dynein Motility: Mechanism](#)

---

## Harmonic Excitation Light Microscopy (HELM)

- [Structured Illumination Microscopy \(SIM\)](#)

## H-bonding

- ▶ [Ultraviolet Resonance Raman \(UVRR\) Spectroscopy Studies of Structure and Dynamics of Proteins](#)

## h-Channels

- ▶ [HCN Channels: Biophysics and Functional Relevance](#)

## HCN Channels: Biophysics and Functional Relevance

Dario DiFrancesco and Mirko Baruscotti  
Laboratory of Molecular Physiology and Neurobiology, Department of Biosciences, University of Milan, Milano, Italy

### Synonyms

“Funny” channels; f-channels; h-channels; Hyperpolarization-activated channels; q (queer)-channels

### Definition

*Hyperpolarization-activated Cyclic Nucleotide-gated channels* (HCN) are ion channels largely expressed in excitable tissues. In the heart, they generate the spontaneous electrical oscillations of pacemaker cells, while in neurons they are involved in the control of cell excitability.

### Basic Characteristics

#### Cardiac Pacemaking: Properties and Functions of the Pacemaker “Funny” ( $I_f$ ) Current

The heart of a healthy person beats on average 2.5 billion times in a lifetime at a frequency ranging from ~40 bpm (beats per minute, sleep) to ~220 bpm (maximal exercise).

This performance is generated by the activity of spontaneously beating (pacemaker) cells of the sinoatrial node (SAN). The ability to generate repetitive

activity is evident from the electrical behavior of single SAN cells (Fig. 1, upper right).

During the diastolic depolarization (pacemaker) phase, the membrane voltage slowly depolarizes from  $-60/-65$  mV to  $-40$  mV, until reaching the threshold for  $Ca^{2+}$  current-driven initiation of a new action potential. This phase of the action potential is typical of pacing cells and lacks in working muscle. It is generated by the “funny” current, first described over 30 years ago (reviewed by DiFrancesco 1993) and characterized in cardiac pacemaker cells ( $I_f$ ) and neuronal cells ( $I_h$ ).

$I_f$  is a mixed  $Na^+/K^+$  current activating on hyperpolarization from a threshold of  $-40/-50$  mV and a reversal potential of  $-10/-20$  mV; its activation rate increases rapidly at more negative voltages (DiFrancesco 1981). These properties are well designed to initiate the process of diastolic depolarization which explains why  $I_f$  is designated as “pacemaker” current.

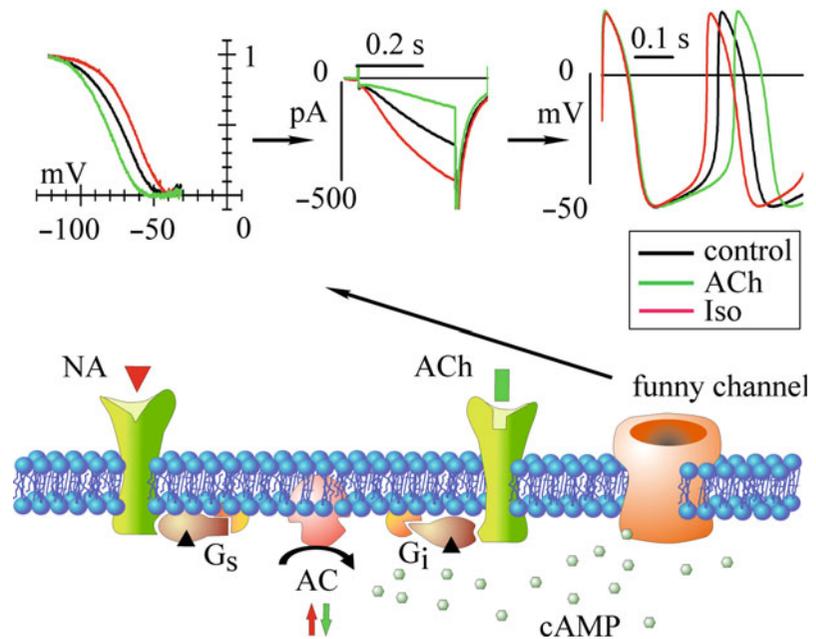
As well as acting as a primary pacemaker generator,  $I_f$  also represents a physiologically relevant target for different homeostatic mechanisms that fine-tune the heart rate in accordance with the metabolic requirement of the organism. Typically,  $I_f$  contributes to autonomic regulation of heart rate; as illustrated in Fig. 1. cAMP is synthesized by adenylate-cyclase, and its levels are finely regulated by a balance between sympathetic and parasympathetic stimuli. Noradrenaline increases cAMP, which binds directly to f-channels (DiFrancesco and Tortora 1991) and causes a rightward shift of the activation curve, thus lowering the voltage hyperpolarization required for channel opening; this in turn increases inward current during diastole and accelerates pacemaking. Acetylcholine has the opposite effect (DiFrancesco 1993).  $I_f$  is also modulated by several hormones and by the interaction with auxiliary or structural proteins, and all these mechanisms act synergistically to fine-tune the current activation range and kinetics and, thus, set the amount of current that can be recruited during cell activity.

#### Properties and Functions of the Neuronal Hyperpolarization-Activated ( $I_h$ ) Current

$I_h$  currents similar to  $I_f$  have been described in several areas of the brain and in peripheral ganglia where they play important functions in cellular activity. In the brain,  $I_h$  has been identified almost ubiquitously (neocortex, hippocampus, basal ganglia, thalamus, brainstem, hypothalamus, retinal cells, olfactory bulb,

### HCN Channels: Biophysics and Functional Relevance,

**Fig. 1** Properties of funny channels and role in the generation of pacemaker activity and autonomic modulation of cardiac rate



cerebellum), while in the peripheral nervous system it has been investigated mostly in the dorsal root ganglion (DRG).

$I_h$  contributes to electrical activity by regulating resting voltage level and input resistance, thus setting cell's excitability and firing rate, and by counteracting excitatory or inhibitory stimuli.

As well as by voltage,  $I_h$  is modulated by intracellular messengers and factors and is sensitive to external stimuli like pH. In sensory neurons, these properties help to mediate perception and transmission of external stimuli;  $I_h$  is expressed in the retina; taste buds; thermal, olfactory, and stretch receptors; inner hair cells of the cochlea; and nociceptors (Robinson and Siegelbaum 2003; Biel et al. 2009).

$I_h$  affects not only the activity of individual cells, but also synaptic communication between neurons and signal processing.  $I_h$  expression in dendritic terminals of hippocampal CA1 neurons increases with the distance from the soma. This is thought to underlie the greater attenuation of temporal summation of distal versus proximal EPSP signals, which normalizes the process of signal integration over the entire length of the dendritic tree (Magee 1999).  $I_h$  plays a role in generating and/or controlling neuronal automaticity in several brain areas including the thalamocortical relay neurons, the cerebellar Purkinje cells, and neurons of entorhinal cortex (Robinson and Siegelbaum

2003). It also exerts important physiological functions in cerebellar Purkinje cells, where it contributes to motor learning, and in prefrontal cortical neurons, where it contributes to working memory (Biel et al. 2009).

### HCN Channels: Molecular Determinants of $I_f/I_h$ Currents

Native f-channels are tetramers coded by subunits of the Hyperpolarization-activated Cyclic Nucleotide-gated (HCN) channel family. In mammals, this family includes four isoforms (HCN1–4) which are variously distributed among excitable tissues. HCN proteins are formed by six transmembrane domains (S1–S6) and intracellular N- and C- termini. The C-termini include two functionally relevant regions: the C-linker, a membrane proximal region of about 80 amino acids, and the Cyclic Nucleotide Binding Domain (CNBD) hosting the binding site for the modulatory second messenger cAMP.

The transmembrane S1–S6 core contains structural elements controlling the permeability, the intrinsic voltage-dependence, and the pharmacological sensitivity of HCN channels. The S4 segment (rich in basic aminoacidic residues) is the voltage sensor element. This segment transfers its voltage-induced movements to the S4-S5 linker which in turn induces the channel gating by increasing/decreasing its

electrostatic interaction with the spatially adjacent C-linker. The S4-S5 linker and the C-linker are thought to contribute to the physical gate whose variable spatial configurations correspond to open/closed pore states. The channel permeability filter is located within the pore (P) region joining the S5 and S6 segments, and is partly determined by the presence in the pore sequence of a GYG triplet common to members of the superfamily of  $K^+$  voltage-dependent and CNG channels to which HCN channels belong. The GYG motif is a known “signature” of  $K^+$ -permeable channels, and it is curious that HCN pores are also permeable to  $Na^+$  ions (DiFrancesco 1981).

The different HCN isoforms have different kinetic and modulatory properties. Typically, HCN4 have the slowest and HCN1 channels the fastest activation kinetics, while HCN2 and HCN3 channels have intermediate rates of activation; the cAMP-induced shift of the activation curve is large (up to 15–20 mV) for HCN2 and HCN4, and small for HCN1 (a few mV at most), while HCN3 does not appear to be cAMP sensitive. Values of kinetic and modulatory parameters can be found elsewhere in the literature (Baruscotti et al. 2005).

Experimental evidence indicates that functional channels can result from both homotetrameric and heterotetrameric assembly. In the presence of different kinetics and cyclic-nucleotide sensitivity of individual isoforms, this feature allows a wide range of properties useful for a variety of cellular functions.

#### Factors Modulating the Activity of HCN Channels

Native  $I_f/I_h$  currents have variable properties in different tissues, and specific kinetic/modulatory features cannot be fully reproduced by heterologous expression of homomeric and/or heteromeric assembly of HCN isoforms. This has led to the idea that HCN channels are part of macromolecular complexes and that auxiliary elements likely affect HCN function; many modulatory factors have indeed been identified (Table 1). This ensures large modulating capability and adaptation to different physiological conditions.

#### Clinical and Pharmacological Relevance of HCN Channels

The search for substances that are able to selectively modify heart rate without side effects has long been an important pharmacological target. This search has identified drugs such as alinidine, zatebradine,

cilobradine, ZD-7288, and ivabradine (“specific heart rate–reducing agents”), which slow heart rate by specific block of the  $I_f$  current. Ivabradine is presently the only commercially available  $I_f$  inhibitor prescribed for the therapy of angina. Ivabradine enters the channels from the intracellular side and binds to a site located within the inner hydrophilic vestibule of the channel (Fig. 2). Block has the peculiar property of being current dependent (Bucchi et al. 2002).

Interest for development of drugs interacting with specific HCN isoforms has increased recently also in view of growing evidence for the contribution of dysfunctional HCN channels to diseases such as epilepsy, inflammation, and neuropathic pain (Dibbens et al. 2010; DiFrancesco et al. 2011; Emery et al. 2011).

Genetic investigation of cardiac rhythm disorders has so far identified six loss-of-function mutations of the HCN4 protein potentially associated with various forms of arrhythmias (Baruscotti et al. 2010; Laish-Farkash et al. 2010; Schweizer et al. 2010).

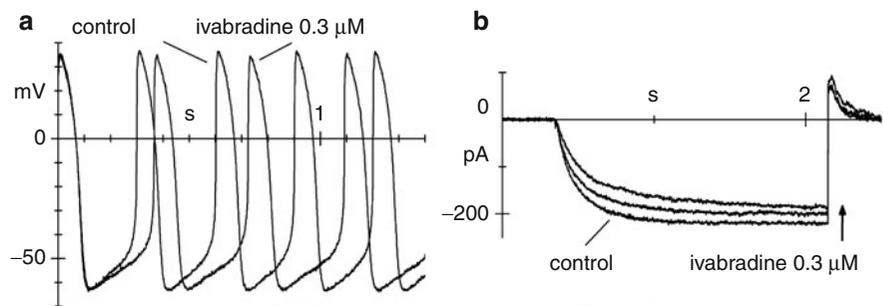
Figure 3 illustrates the effects of the mutation S672R (a) in the HCN4 channel found in an Italian family with asymptomatic bradycardia. The pedigree (b) and the bar graph (c) show that mutated (filled symbols) and wild-type individuals (open symbols) had all rates lower and higher, respectively, than 60 bpm, indicating high degree of co-segregation. Functional studies indicated that mutant channels are activated at more negative voltages than wild-type channels; these change slow the heart rate by decreasing the  $I_f$  current during diastole, and mimic those of mild vagal stimulation (Milanesi et al. 2006).

The pathological consequences of the removal of  $I_f$  have been recently also investigated in vivo in a HCN4 knockout mouse model, and data indicate that removal of  $I_f$  leads to extreme bradycardia and to complete block of the conduction through the AV node (Baruscotti et al. 2011). Knockout mouse models (of HCN1 or HCN2) have also been critically relevant in suggesting a causative role for these isoforms in neuronal diseases such as ataxia, epilepsy, and impaired learning.

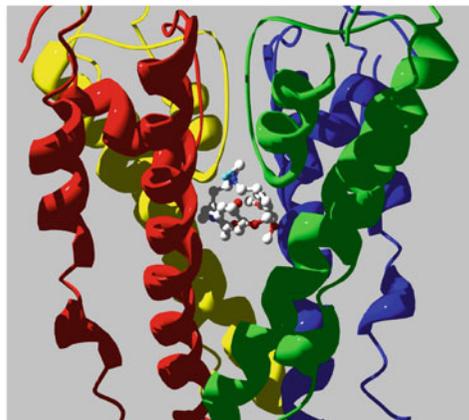
More recent work has identified a mutation in the HCN2 channel associated with recessive generalized epilepsy due to a loss of function mechanism (DiFrancesco et al. 2011). Expression of homomeric mutant channels generates currents characterized by a large negative shift of the activation curve which essentially moves the voltage range for current activation outside the physiological range. This change

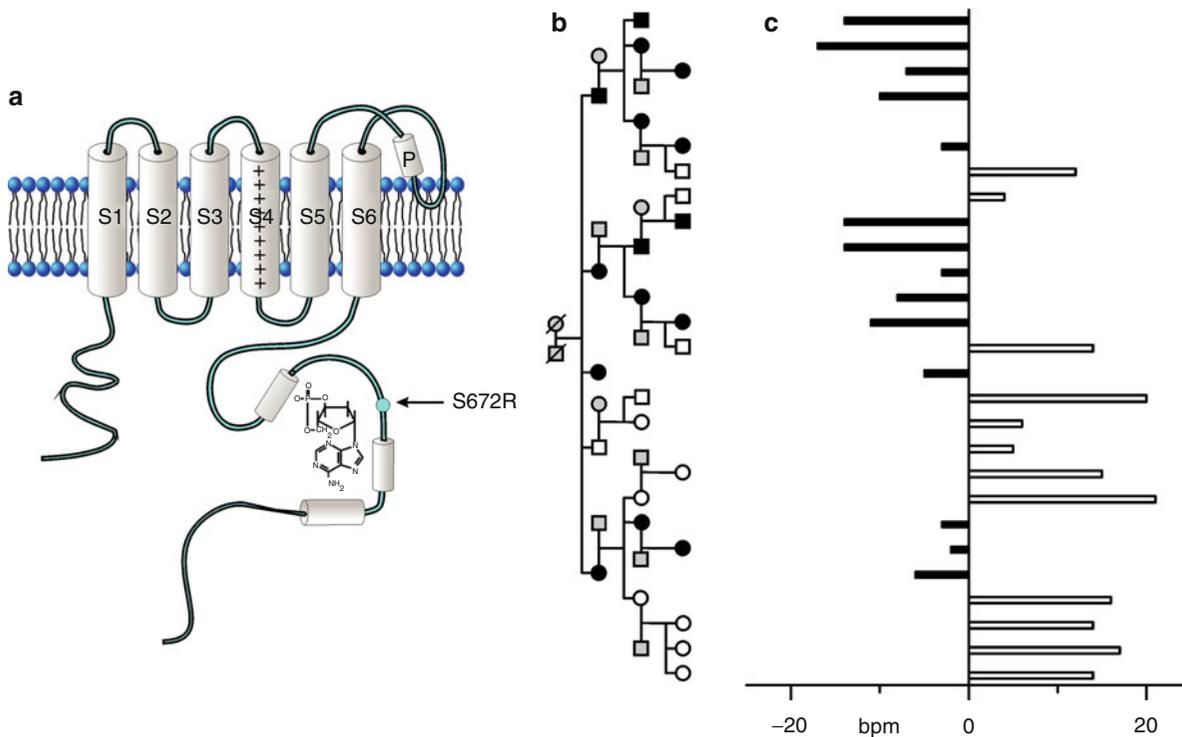
**HCN Channels: Biophysics and Functional Relevance, Table 1** Factors affecting properties and functions of HCN channels

Substance	Mode of action		Target
	Conductance	Kinetics/others	
<i>Proteins</i>			
MiRP1	↑	Altered kinetics	All isoforms
Caveolin 3	↑	Localization	HCN4
		β-AR modulation	HCN4
Filamin A		Slower kinetics	HCN1
TRIP8b	↑/↓	Slower kinetics	HCN1
		Splice-dependent modulation	HCN1, HCN2
		Impairment of cAMP modulation	HCN2, HCN4
Tamalin/Mint2		Formation of multi-protein complex	HCN2
		Regulation of expression in dendrites	HCN2
KCR1	↓		HCN2
<i>Enzymes</i>			
Src kinases	↑	Faster kinetics	HCN2, HCN4
cGMP-dep PKII		Slower kinetics	HCN2
MAP kinase		Shifts right the AC curve	Hippocampal HCN1
Phosphatases calcineurin		Shifts left the AC curve	Hippocampal HCN1
<i>Acidic lipids</i>			
PI(4,5)P2		Shifts right the AC curve	All isoforms
Arachidonic acid/phosphatidic acid		Shifts right the AC curve	HCN2
<i>pH</i>		Shifts left the AC curve	HCN2
		Shifts right the AC curve	HCN1, HCN4



**HCN Channels: Biophysics and Functional Relevance, Fig. 2** Funny channel inhibition caused by ivabradine-induced block causes slowing of heart rate. (a) Slowing of spontaneous rate in a SAN cells by ivabradine; (b) progressive inhibition of  $I_f$  induced by the same dose of ivabradine in a single SAN cell; (c) cartoon showing approximate positioning of an ivabradine molecule within the water-filled cavity underneath the pore of an HCN4 channel





**HCN Channels: Biophysics and Functional Relevance, Fig. 3** S672R mutation of the HCN4 channel associated with bradycardia. (a) Cartoon of one subunit of HCN4 showing the approximate position of the mutation in the CNBD; (b) pedigree

of a family with bradycardia (*filled symbols*: with mutation; *open symbols*: wild-type); (c) corresponding heart rates indicated as variations relative to 60 bpm (From Milanesi et al. 2006)

occurs in homomeric mutant channels but not in heteromeric wild-type mutant constructs, which determines recessive inheritance and leads to strongly increased neuronal excitability, a condition predisposing to epileptogenesis.

## Cross-References

- ▶ [Bioelectricity, Ionic Basis of Membrane Potentials and Propagation of Voltage Signals](#)
- ▶ [Gating of Potassium Channels by Cyclic Nucleotide Binding](#)

## References

- Baruscotti M, Bucchi A, DiFrancesco D. Physiology and pharmacology of the cardiac pacemaker (“funny”) current. *Pharmacol Ther.* 2005;107:59–79.
- Baruscotti M, Bottelli G, Milanesi R, DiFrancesco JC, DiFrancesco D. HCN-related channelopathies. *Pflugers Arch.* 2010;460:405–515.

- Baruscotti M, Bucchi A, Viscomi C, Mandelli G, Consalez G, Gnecci-Rusconi T, Montano N, Casali KR, Micheloni S, Barbuti A, DiFrancesco D. Deep bradycardia and heart block caused by inducible cardiac-specific knockout of the pacemaker channel gene *Hcn4*. *Proc Natl Acad Sci USA.* 2011;108:1705–10.
- Biel M, Wahl-Schott C, Michalakis S, Zong X. Hyperpolarization-activated cation channels: from genes to function. *Physiol Rev.* 2009;89:847–85.
- Bucchi A, Baruscotti M, DiFrancesco D. Current-dependent block of rabbit sino-atrial node I(f) channels by ivabradine. *J Gen Physiol.* 2002;120:1–13.
- Dibbens LM, Reid CA, Hodgson B, Thomas EA, Phillips AM, Gazina E, Cromer BA, Clarke AL, Baram TZ, Scheffer IE, Berkovic SF, Petrou S. Augmented currents of an HCN2 variant in patients with febrile seizure syndromes. *Ann Neurol.* 2010;67:542–6.
- DiFrancesco D. A study of the ionic nature of the pace-maker current in calf Purkinje fibres. *J Physiol (Lond).* 1981;314:377–93.
- DiFrancesco D. Pacemaker mechanisms in cardiac tissue. *Annu Rev Physiol.* 1993;55:451–67.
- DiFrancesco D, Tortora P. Direct activation of cardiac pacemaker channels by intracellular cyclic AMP. *Nature.* 1991;351:145–7.
- DiFrancesco JC, Barbuti A, Milanesi R, Coco S, Bucchi A, Bottelli G, Ferrarese C, Franceschetti S, Terragni B,

- Baruscotti M, DiFrancesco D. Recessive loss-of-function mutation in the pacemaker HCN2 channel causing increased neuronal excitability in a patient with idiopathic generalized epilepsy. *J Neurosci*. 2011;31:17327.
- Emery EC, Young GT, Berrocoso EM, Chen L, McNaughton PA. HCN2 ion channels play a central role in inflammatory and neuropathic pain. *Science*. 2011;333:1462–6.
- Laish-Farkash A, Glikson M, Brass D, Marek-Yagel D, Pras E, Dascal N, Antzelevitch C, Nof E, Reznik H, Eldar M, Luria D. A novel mutation in the HCN4 gene causes symptomatic sinus bradycardia in moroccan jews. *J Cardiovasc Electrophysiol*. 2010;21:1365–72.
- Magee JC. Dendritic  $I_h$  normalizes temporal summation in hippocampal CA1 neurons. *Nat Neurosci*. 1999;2:508–14.
- Milanesi R, Baruscotti M, Gnecci-Ruscone T, DiFrancesco D. Familial sinus bradycardia associated with a mutation in the cardiac pacemaker channel. *N Engl J Med*. 2006;354:151–7.
- Robinson RB, Siegelbaum SA. Hyperpolarization-activated cation currents: from molecules to physiological function. *Annu Rev Physiol*. 2003;65:453–80.
- Schweizer PA, Duhme N, Thomas D, Becker R, Zehelein J, Draguhn A, Bruehl C, Katus HA, Koenen M. cAMP sensitivity of HCN pacemaker channels determines basal heart rate but is not critical for autonomic rate control. *Circ Arrhythm Electrophysiol*. 2010;3:542–52.

---

## Heart

### ► Modeling the Heart

---

## Helicases

- Martin R. Webb<sup>1</sup> and Christopher P. Toseland<sup>2</sup>  
<sup>1</sup>MRC National Institute for Medical Research, London, UK  
<sup>2</sup>Institut für Zelluläre Physiologie, Physiologisches Institut, Ludwig Maximilians Universität, Munich, Germany

## Synonyms

DNA; Molecular motors

## Introduction

Although DNA is most stable as a double-stranded helical structure, this must be unwound and the strands separated transiently for most DNA processing.

The single-stranded intermediates are required for replication (► [Machinery of DNA Replication](#)), repair, recombination, and DNA transfer during conjugation, and, in each of these processes, duplex DNA unwinding is catalyzed by ubiquitous enzymes known as helicases. These proteins cause the destabilization of the hydrogen bonds between the complementary base pairs and the stacking of adjacent bases as it translocates along the DNA. This separation is coupled to hydrolysis of nucleoside triphosphate, most frequently ATP.

A number of enzymes have also been characterized that demonstrate RNA helicase activity (Bleichert and Baserga 2007; Jankowsky 2010). Structurally and functionally, RNA is a diverse molecule that is highly regulated. Helicases function in all aspects of RNA metabolism, such as transcription, editing, rearranging, and mRNA splicing. Specific helicases can also act on RNA–DNA structures, remove proteins from RNA without unwinding, and act as clamps. These roles include regulating transient associations of small noncoding RNAs and ribonucleoprotein complexes to target RNAs. This review mainly focuses on DNA helicases, but the mechanisms and methods described are also applicable to RNA helicases on the whole.

Helicases are motor proteins and as such share some properties with other motor proteins, such as myosins and kinesins (Delagoutte and von Hippel 2002, 2003). They move along a linear track, driven by the hydrolysis of nucleoside triphosphate: this in turn means that some similar techniques are used to investigate these varied types of linear motors. An important difference is that the distance between repeating units of nucleic acid (nucleotides) are much smaller than is the case for the protein track of actin or tubulin. However, some important properties of helicases are common to all motor proteins, including ideas of directionality and processivity, defined as how far along the DNA, or RNA, a helicase moves on average before stalling and dissociating. The movement can generally be defined in terms of steps. The step size is the smallest distance of movement that can be detected, often by single-molecule techniques. The step size may be related to the coupling ratio, which is the distance moved per ATP hydrolyzed: often this is one or a few bases. However, in some cases, the kinetic step size does not correlate with movement during a distinct cycle of ATP hydrolysis. Another property

is the speed of movement along the DNA, which may vary with both nucleic acid structure and on whether the helicase is acting alone or as part of a larger complex with other proteins, as described later.

## Variety and Classification of Helicases

Helicases were first identified in 1976 and classified into families in 1993 (Gorbalenya and Koonin 1993). However, it appears that many putative helicases do not actually unwind nucleic acids, although they do harness the energy of ATP hydrolysis to translocate along the DNA. A more detailed discussion of what constitutes a helicase was given in the review by Singleton et al. 2007. The two largest superfamilies of helicases (SF) are typically monomeric or dimeric: SF-1 (containing some well studied DNA helicases, such as PcrA, Rep, UvrD, RecB, RecD) and SF-2 (containing DNA helicases RecQ, PriA, RecG, UvrB, and almost all RNA helicases) were initially classified on the basis of seven conserved regions of sequence homology. New motifs have been identified, including TxGx, Q-motif, motif 4a, and TRG, which are specific to each superfamily or subfamily. The superfamilies, SF-3 to SF-6, also contain hexameric or double hexameric helicases, which contain fewer conserved motifs. Despite the diversity, there are universal structures across all of the superfamilies, such as the “core domains” that form the tandem RecA-like folds, either within the same polypeptide or between subunits. RecA is the prokaryotic DNA-strand exchange protein required for recombination. These core domains contain residues relating to the “A” and “B” motifs of the “Walker box,” a predictor of nucleoside-5'-triphosphate binding.

A more recent and full classification was given by consideration of the sequence, as well as the oligomeric state, translocation polarity, and DNA substrate (Singleton et al. 2007). With respect to the oligomeric state, helicases can be monomeric, dimeric, or hexameric. The SF1 and SF2 members contain the RecA-like folds on a single polypeptide. Therefore, dimeric helicases from these superfamilies generally contain two functional motors, whereas hexameric helicases contain six individual RecA-like folds. Translocation by helicases can be defined as 3'-5' or a 5'-3', with respect to the strand with the major interaction to the helicase motor elements. SF1, SF2,

and SF6 contain examples of both forms of translocation, whereas all characterized SF3 members are 3'-5' helicases. All members of SF4 and SF5 are 5'-3' helicases. Extra levels of complexity are given by accessory domains. These play roles in targeting motor activity to specific substrates, such as forked junctions. Furthermore, novel enzymatic properties can be produced via the addition of complementary catalytic domains to helicases, allowing the creation of various nucleic acid processing machines. For instance, the cooperation of helicase and nuclease is common. Other domains or subunits may provide recognition of particular sequences, such as occurs with RecBCD, where one subunit recognizes the short *chi* sequence, important for creating a DNA intermediate during recombination (Dixon and Kowalczykowski 1991). Furthermore, helicases may form interactions with other enzymes, such as polymerases, that create a complex for unwinding and processing double-stranded DNA (dsDNA).

## Regulation

Helicases are DNA modifying enzymes and as such, unregulated behavior is deleterious to the cell. One consequence of this is that defects in certain helicases can cause a variety of genetic disorders, as well as leading to a predisposition to cancers, through impairment of DNA repair pathways (van Brabant et al. 2000). Helicases need to function at a defined time in precise places. Each helicase may require a specific DNA structure as substrate and the formation of such structures may trigger helicase activity, for example, free ends or junctions appearing following DNA damage and breaks. Helicases may also have to switch between translocase and helicase activities. The evidence suggests that this is achieved by a change in oligomeric state (e.g., dimerization) and/or by interaction with accessory proteins. Several proteins have been shown to increase helicases activity. An example of this is PcrA, whose helicase activity is greatly enhanced by the replication initiation protein, RepD, which also enables the helicase to be localized at the correct position to take part in asymmetric plasmid replication. The helicase can then unwind complete plasmids of several thousand bases pairs, although a single PcrA on its own is unable to unwind more than a few base pairs. ► [DNA polymerases](#) have also

been shown to activate helicase activity, and many helicases will only work processively as part of larger complexes (Donmez and Patel 2006). Figure 1 shows some of the variety of biological roles of helicases and different DNA structures.

## Mechanism

Helicase mechanisms can be characterized using a variety of bulk solution and single-molecule approaches, some of which are outlined below but described in more detail in reviews: (Donmez and Patel 2006; Lohman et al. 2008; Pyle 2008). These methods have shown directionality, real-time translocation and unwinding rates, along with corresponding ATPase rates for a variety of nucleic acid substrates. Such details lead to an understanding of the dynamics of helicase activity and its coupling to ATP usage, and the mechanism by which ATP is hydrolyzed.

There are several ways of considering the mechanism by which helicases might separate strands and translocate. One way is to consider how changes in the interaction of a protein with DNA could, either actively or passively, act on DNA (Fig. 2). A passive mechanism would have an opportunistic protein waiting at the junction between single-stranded DNA (ssDNA) and dsDNA for a transient opening of the base pairs. When the opening occurs, the helicase advances forward, blocking the reformation of the base pairs. At the other extreme, the ATP-driven translocation of the helicase forces its way into the junction and actively destabilizes the double-strand interactions. For many helicases, there remains uncertainty as to what extent these extremes of mechanisms contribute to the unwinding and translocation of DNA.

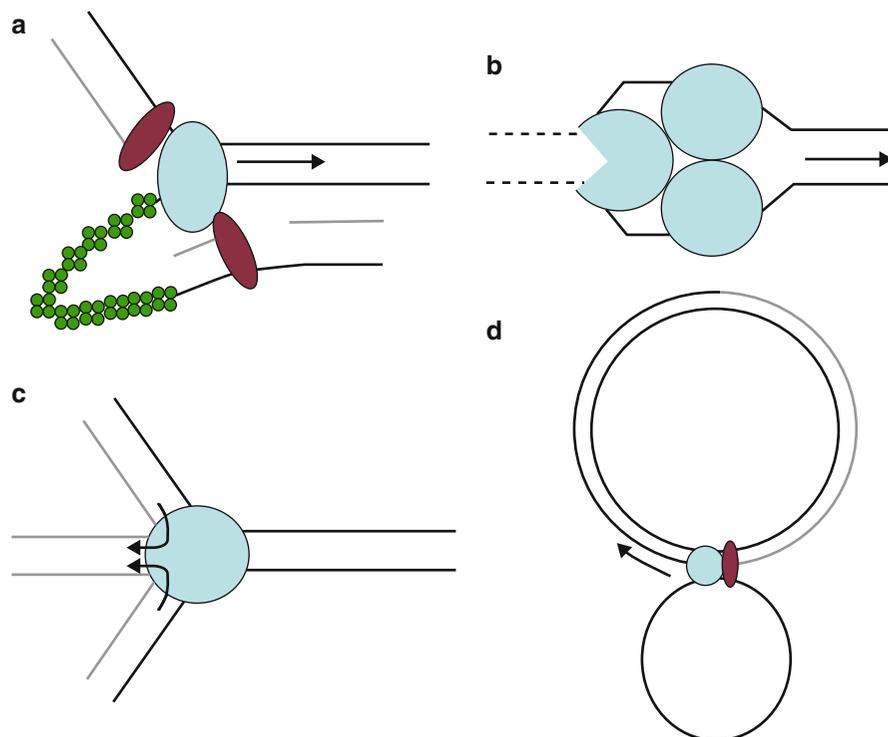
This active mechanism sits well with the observation of wedge domains in some helicase structures that sit at the ssDNA/dsDNA junction and point into the duplex with one strand of ssDNA lying either side of the wedge (Singleton et al. 2007). Detailed information supporting an active mechanism came from the crystal structures of *Bacillus stearothermophilus* PcrA in complex with a partial duplex DNA molecule in two nucleotide states (Dillingham et al. 2001; Singleton et al. 2007). The structures (Fig. 3) revealed a protein divided into four subdomains, of which two, 1A and 2A, are homologous to RecA. These are the core motor domains containing the ATP binding site at their

interface. The ssDNA-binding site can be seen to run along the top of the motor domains in a groove which can accommodate ssDNA. By comparison with the apo structure, a transition of domain 2B upon binding to the partial duplex DNA is observed. The movement makes a new interface for the duplex DNA and may suggest a regulatory function for this domain.

These structures suggested a translocation mechanism, in which domains 1A and 2A were in different conformations depending on whether ATP or ADP was bound (Singleton et al. 2007). The substrate and product structures also showed differences in the contacts made between the protein and ssDNA. The translocation reaction cycle proceeds through a mechanism similar to an inchworm. The core subdomains, 1A and 2A, move relative to one another on ATP binding and hydrolysis. The two domains bind ssDNA bases in pockets that open and close during this conformational change. The binding site is composed of a series of aromatic amino acid residues that form pockets for the bases of the ssDNA. These residues make alternating contacts with individual bases. Initially, the ssDNA is bound tightly by domain 1A. ATP binding causes the pocket in domain 2A to bind the ssDNA, while the pockets in domain 1A release ssDNA. At the same time, the two domains move closer together. Upon ATP hydrolysis, the pockets in 1A bind ssDNA and the pockets in 2A release ssDNA. The domains also move apart, causing the ssDNA to be pulled along the DNA binding channel relative to domain 2A, translocating the DNA. This resets the conformation for another cycle.

Translocation does not require that the dsDNA-binding site is filled. However, if the ssDNA, along which the enzyme is translocating, is attached to a duplex region ahead of the enzyme, the translocation process will remove the duplex, in essence stripping the strands apart, and the one strand is pulled across the protein as described above (Marians 2000). In one sense, DNA unwinding is a biproduct of ssDNA translocation.

Many structures of helicase have now been described, including those of hexameric proteins (Donmez and Patel 2006; Singleton et al. 2007) and, particularly by having structures with different DNA and/or nucleotide present, these have often led to equal insight into the mechanism of action that has been related to solution studies and single-molecule measurements.



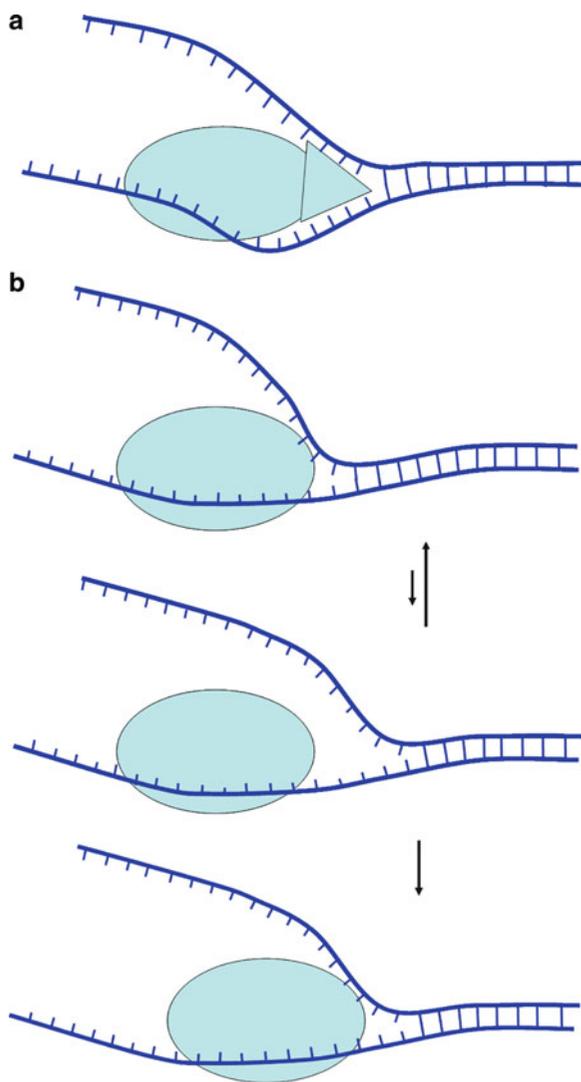
**Helicases, Fig. 1** Cartoon of some the biological roles of DNA helicases and variety of DNA structures. Helicases are shown in *blue* and other proteins in *red*. (a) Helicase and polymerases, involved in bacterial DNA replication. Single-stranded DNA-binding protein (SSB) is in *green*. The *arrows* refer to the helicase movement. Parental DNA is colored *black* and daughter DNA is *grey*. (b) RecBCD helicase nuclease, cutting DNA as it translocates. (c) RecG involved in moving a 4-way junction

during repair of single-strand damage. *Arrows* indicate the pulling of DNA strands as the nascent DNA is unwound from the parental strands. (d) PcrA helicase, with polymerase and initiator protein RepD, perform asymmetric plasmid replication. A potential intermediate is shown, whereby one parent strand is being replicated, while the other remains as a single-stranded loop, awaiting interaction with the second polymerase

## Translocation and Unwinding Measurements in Bulk Solution

One of the most important characteristics of a helicase is the unwinding activity, and there is a wide variety of rates observed from  $<1$  base  $s^{-1}$  to  $>1,000$  bases  $s^{-1}$ . Similarly, there is a wide range of processivities, from helicases that apparently unwind only a few bases to ones that translocate many thousands of bases without dissociation. The speed and processivity are likely to be optimized for their roles in the cell. For example, a replicative helicase will have to function rapidly and processively. A slow speed may delay cell division, while low processivity will lead to replication fork stalling or collapse. On the other hand, a repair helicase may only have to unwind a hundred base pairs at most, so a high processivity is not required.

Early methodology to follow unwinding by helicases focused on gel assays, which have the advantage of allowing observation and identification of the DNA product. However, such assays have the disadvantage of a limited time resolution and being discontinuous, that is, each gel lane represents a single snapshot in time. However, a variety of fluorescence-based methods (► [Fluorescence](#)) are now available for these and other measurements of activity (Toseland and Webb 2010; Webb 2007). Fluorescence measurements can report on processes in real time due to the sensitive and rapid response of the fluorescence probe. When combined with a stopped-flow fluorimeter (► [Stopped-Flow Techniques](#)), assays can be performed with millisecond time resolution. One type of measurement is sometimes described as “all or nothing,” in which a fluorophore is located at one



**Helicases, Fig. 2** Cartoon representation of active and passive mechanisms of DNA unwinding. (a) The active mechanism shows a wedge domain at the junction, which aids the disruption of the dsDNA, as indicated by a slight separation of the DNA strands immediately in front of the wedge. In practice, there may be significant protein interactions with dsDNA, as shown in the structure in Fig. 3a. (b) In a passive mechanism, there is transient separation of strands ahead of the helicase, which then translocates into the space vacated and so prevents the DNA reannealing

point on the DNA (► [Fluorescence Labeling of Nucleic Acids](#)): there is little or no change in signal while the translocation approaches the label, but the fluorescence changes when the helicase reaches that point. The label could be a single fluorophore, a fluorophore-quencher pair, or a FRET

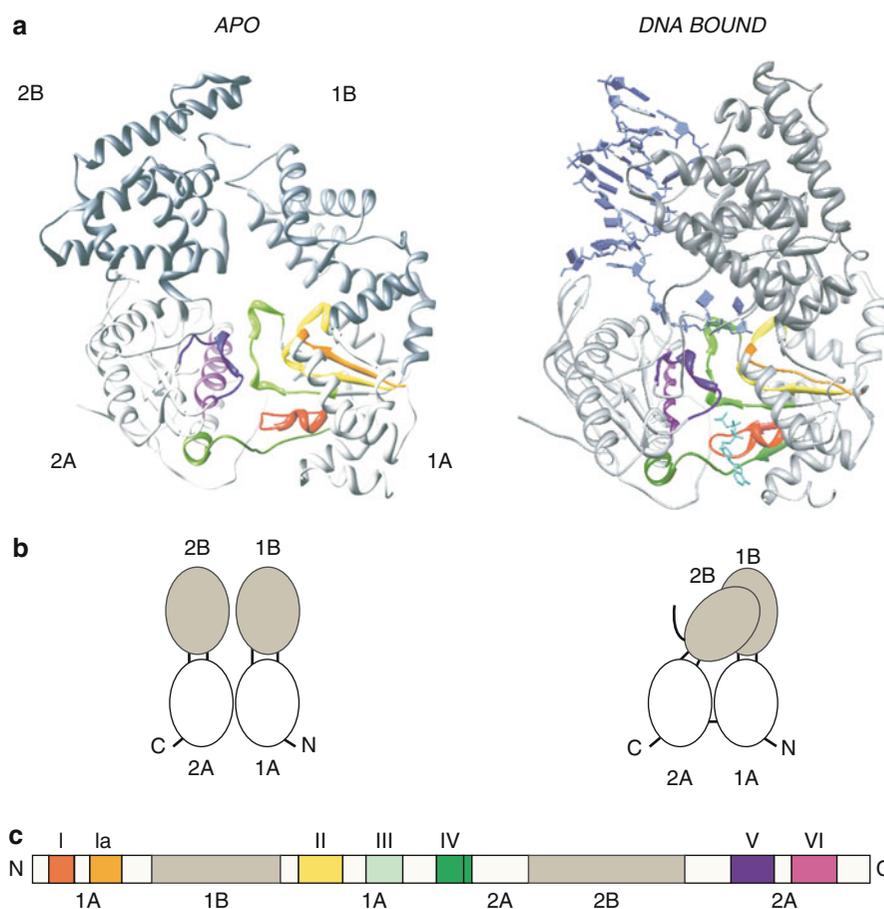
(Förster Resonance Energy Transfer) pair (► [Fluorescence and FRET in Membranes](#)). One of the simplest methods of this type to measure unwinding or translocation over short lengths of DNA is to use oligonucleotides, labeled with a fluorophore(s) at the far end of the strands. Assays of this type are likely to be more successful with DNA that is shorter than ~100 bases. If longer than that, the helicase translocation becomes desynchronized as a consequence of the large number of steps: the helicase molecules become distributed over a wide length of DNA during the translocation.

Alternatively, assays are available using fluorescent probes that differentiate between ssDNA and dsDNA, and these give a continuous change in fluorescence as the unwinding proceeds. There are a variety of fluorescent dyes that bind to dsDNA with enhanced fluorescence. They are released on helicase unwinding, so giving a signal change, allowing the extent of unwinding to be monitored, although the presence of the dye bound to the dsDNA may interfere with helicase activity. A biosensor for ssDNA has been developed based on a single-stranded DNA-binding protein (SSB), which binds to ~70 bases of ssDNA, although this is dependent on conditions (Webb 2010). The biosensor increases in fluorescence when bound to ssDNA, providing a real-time signal during unwinding (Fig. 4). SSB has been shown to interact with a variety of other DNA-binding proteins, including helicases, so that binding and/or rearrangement of SSB on the ssDNA product can potentially modulate the helicase activity. This SSB-based biosensor has been successfully used to measure helicase-catalyzed unwinding of long linear fragments and plasmids, and some of the factors that allow it to be used under a variety of conditions have been discussed (Webb 2010).

## Measurements of ATPase Activity

An assessment of ATPase activity is part of a full characterization of a helicase and an understanding of the dynamics of the system. An immediate question is to what extent is ATP hydrolysis coupled to the translocation: is there a fixed coupling ratio with unidirectional motion or is there a ratio that depends on conditions? The extent of coupling may differ with helicase and conditions, although for many it seems quite fixed with some helicases moving one base per ATP, others moving several.

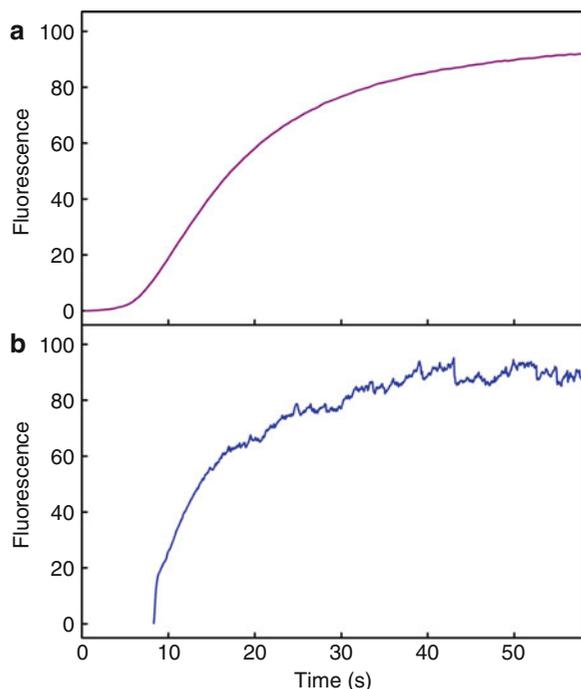
**Helicases, Fig. 3** Structure of a monomeric helicase, PcrA. (a) Crystal structure of the apoprotein and change in structure due to binding DNA and an ATP analogue (Singleton et al. 2007). (b) Cartoon representation of domain movements. (c) Bar representation of domain structure with color code as in panel A



Generally, ATP utilization during unwinding is measured using stopped-flow fluorimetry as the rate of hydrolysis is likely to preclude measurements following manual mixing. This addresses the relationship between ATP hydrolysis and DNA unwinding, answering questions such as how many bases are unwound in a single ATP cycle? It is possible to measure ATP hydrolysis during DNA unwinding using the oligonucleotides and plasmid DNA substrates described above, measuring the products of ATP hydrolysis in real time. One way that achieves this is use of the fluorescent reagentless biosensor, MDCC-PBP (Webb 2007). This has the high sensitivity and time resolution to measure the inorganic phosphate as it is released by the helicase and has been used in varied transient and steady-state kinetic measurements of ATPases and GTPases, such as myosins, kinesins, dyneins, small G proteins, as well as helicases.

The ATPase rate, measured in this way (i.e., during unwinding), can then be compared to the unwinding rate. A coupling ratio is then established to determine the number of DNA bases unwound per ATP. This is an important parameter to determine how the helicase translocates the DNA strand.

A more detailed understanding of the mechanochemical coupling between ATP hydrolysis to DNA unwinding and the relationship between that coupling and protein structural changes requires information about the individual steps of the ATPase cycle. This will aid in the determination of which intermediates are most prevalent, and which steps are likely to be closely associated with the structural transitions. These methods, typically using stopped-flow fluorescence, are well established in the motor protein field, and fluorescent ATP analogues, such as mantATP, have been widely used (Jameson and Eccleston 1997). The fluorescence probe leads to signal changes upon



**Helicases, Fig. 4** Example of continuous fluorescence data for time-resolved measurement of unwinding: DNA unwinding by AddAB helicase. The measurements use a fluorescent SSB, which binds to the ssDNA as it forms (Fili et al. 2010). (a) Bulk solution assay, using a coumarin-SSB. A linear biotinylated 1.5 kb DNA fragment was incubated with streptavidin to give unidirectional unwinding. A nuclease-inactive mutant of AddAB was used in these measurements. (b) The single-molecule equivalent measurement, using Cy3B-SSB. The biotinylated substrate was immobilized on an inert surface. The length of the ssDNA is limited to  $\sim 2$  kb by the extent of the evanescent wave

binding and possibly during conformation changes during the ATP cycle. Helicases typically bind nucleoside triphosphates with the base and phosphate groups buried within the active site. Therefore, fluorophore attachment on the ribose is considerably more benign than modification of the base or triphosphate moieties. A wide variety of adenine nucleotides with ribose-attached fluorophores have been developed, enabling there to be some degree of optimization in terms of fluorophore and length of linker. Helicases may contain tryptophan residues which give a fluorescence signal upon a non-fluorescent ATP binding, thereby preventing the need for non-natural nucleotides. Furthermore, by use of phosphate and/or ADP biosensors (Webb 2007), it is possible to measure the kinetics of product release under the same conditions

as fluorescent ATP binding. In this way, it is possible to measure and compare the different steps of an ATP cycle. Quenched-flow gives similar time resolution to stopped-flow and can be used to measure the cleavage of triphosphate to diphosphate, albeit discontinuously.

## Single-Molecule Methods

Single-molecule methods have been applied widely to motor proteins as they allow real-time observation of movement and mechanics, as well as some of the biochemical processes (Ha and Lohman 2011; Rasnik et al. 2006). Single-molecule measurements using fluorescence, particularly by Total Internal Reflectance Fluorescence Microscopy (Total Internal Reflection Fluorescence Microscopy) (TIRFM), and force techniques, predominantly optical or magnetic tweezers (Magnetic Tweezers; Optical Tweezers), have been applied widely to the study of helicases.

Advantages of the single-molecule approach include the fact that observations of each complex may circumvent inhomogeneity of the substrates. With helicases, the DNA structures may be not be uniform (for example degree of supercoiling) and this would not be apparent in bulk solution studies. Furthermore, helicases unwind DNA with, perhaps, hundreds of discrete biochemical steps. This, in effect, desynchronizes the reaction of individual molecules in bulk solution. The observation of individual unwinding events overcomes this. For example, RecBCD exhibits a pausing behavior after reaching a *chi* sequence (Spies et al. 2003). This activity can only be observed at the single-molecule level because there is a large heterogeneity in the unwinding rates of the helicases population. Such asynchronous behavior means that pauses may not be detected in bulk studies because the events are averaged out over time.

Disadvantages of the single-molecule approach, apart from the requirement for highly sensitive fluorescent probes, include issues related to attachment of the active complex to a surface and the potential effect of that on activity. It is important to assess if “novel” properties result from artifacts in the single-molecule assay due to fluorescent labeling, forces, or attachment. Particularly with TIRFM, the single-molecule data can be compared to the ensemble population

solution measurements. This may be relatively easy with fluorescent assays, but it is harder to compare bulk and single-molecule force measurements.

Single-molecule fluorescence measurements can be applied to helicase unwinding and translocation activity, and conformation changes (Ha and Lohman 2011; Rasnik et al. 2006). TIRFM in combination with an appropriate fluorophore label (or labels) provides a way to observe single events in real time. The surface attachment increases the signal-to-noise ratio as the evanescent wave of this technique excites fluorophores only very close to the surface.

Several types of assays are available, which are suitable to measure unwinding of different lengths of DNA substrate, and, to some extent, these are related to the equivalent assays outlined for bulk measurements. Long, linear DNA substrates (>10,000 bp) can be measured using microfluidic techniques (► [Microfluidics for Single Molecule Detection](#)). One end of the DNA is attached to a surface or a bead held in an optical trap. Hydrodynamic drag (“flow”) stretches the DNA fragments, which can be visualized by an intercalating dye, such as YOYO. The dye is displaced as the helicases unwind the DNA.

Shorter linear DNA substrates (<3,000 bp) can be measured using a Cy3b labeled version of the ssDNA biosensor, described above (Fig. 4). This assay has a resolution of ~70 bp, the binding site of the biosensor. Multiple unwinding events can be visualized as spots of increasing intensity, as the biosensor binds the ssDNA. This assay can be performed with DNA, or helicases, immobilized on a surface. Importantly, the data can be readily compared to bulk measurements using the same type of biosensor.

Single-molecule FRET ► [\(Single-Molecule Fluorescence Resonance Energy Transfer\) \(smFRET\)](#) assays can typically be used with substrates up to ~100 bp, but this approach has the highest potential resolution of <10 bp and potentially single-base. smFRET assays can be used with FRET pairs between the DNA strands, or the protein and DNA. FRET pairs between the protein and DNA are particularly useful as it gives the possibility to measure conformation changes. Attaching fluorophores at several locations on the protein and DNA allows their precise positions to be triangulated and, therefore, determine domain rearrangements.

Force measurements, optical trapping, and magnetic tweezers are used to assay DNA unwinding.

These experiments are important to determine the effect of topology and force upon helicase activity. With magnetic tweezers, DNA is tethered between a surface and a paramagnetic bead. During an unwinding assay, the bead is held under a constant force, so as the duplex is unwound, the DNA lengthens. The height of the bead is monitored using bright-field microscopy at a resolution of 10 bp. Optical tweezers provide the highest resolution, but each molecule is observed one at a time, increasing the time to acquire data from a significant number of complexes. Various assays can be performed, which include, immobilizing a helicase to a surface with DNA attached to a bead. The free end of the DNA can bind to the helicase. While the bead is held in the laser trap, it is possible to measure the force of the helicase unwinding the DNA.

## Summary

Helicases are motor proteins that mostly separate double-stranded DNA, or RNA, while translocating along the duplex. This uses the energy of nucleoside triphosphate hydrolysis (generally ATP). They are involved in most processing of DNA, including replication, recombination, and repair, and of RNA, including editing, rearranging, and splicing. Helicases are investigated using a variety of bulk solution and single-molecule techniques. These studies have revealed key aspects of the helicase functions and mechanisms. Processivity of helicases varies between a few bases to many thousands of bases, and speeds vary from <1 base s<sup>-1</sup> and >1,000 bases s<sup>-1</sup>. Structurally, there is also a wide variety from monomeric proteins to hexameric, although helicases often act as part of larger complexes that include other enzyme function.

## Cross-References

- [ATPase: Overview](#)
- [DNA Polymerase](#)
- [Fluorescence](#)
- [Fluorescence and FRET in Membranes](#)
- [Fluorescence Labeling of Nucleic Acids](#)
- [Machinery of DNA Replication](#)
- [Magnetic Tweezers](#)
- [Microfluidics for Single Molecule Detection](#)

- ▶ [Optical Tweezers](#)
- ▶ [Single-Molecule Fluorescence Resonance Energy Transfer](#)
- ▶ [Stopped-Flow Techniques](#)
- ▶ [Total Internal Reflection Fluorescence Microscopy for Single-Molecule Studies](#)

## References

- Bleichert F, Baserga SJ. The long unwinding road of RNA helicases. *Mol Cell*. 2007;27:339–52.
- Delagoutte E, von Hippel PH. Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I: structures and properties of isolated helicases. *Q Rev Biophys*. 2002;35:431–78.
- Delagoutte E, von Hippel PH. Helicase mechanisms and the coupling of helicases within macromolecular machines. Part II: integration of helicases into cellular processes. *Q Rev Biophys*. 2003;36:1–69.
- Dillingham MS, Soutanas P, Wiley P, Webb MR, Wigley DB. Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc Natl Acad Sci USA*. 2001;98:8381–7.
- Dixon DA, Kowalczykowski SC. Homologous pairing in vitro stimulated by the recombination hotspot, *Chi*. *Cell*. 1991;66:361–71.
- Donmez I, Patel SS. Mechanisms of a ring shaped helicase. *Nucleic Acids Res*. 2006;34:4216–24.
- Fili N, Mashanov G, Toseland CP, Batters C, Wallace MI, Yeeles JTP, Dillingham MS, Webb MR, Molloy JE. Visualizing DNA unwinding by helicases at the single molecule level. *Nucleic Acids Res*. 2010;38:4448–57.
- Gorbalenya AE, Koonin EV. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol*. 1993;3:419–29.
- Ha T, Lohman TM. Single molecule FRET studies of helicases. *Annu Rev Biophys*. 2012;41
- Jameson DM, Eccleston JF. Fluorescent nucleotide analogs: synthesis and applications. *Methods Enzymol*. 1997;278:363–90.
- Jankowsky E. RNA helicases at work: binding and rearranging. *Trends Biochem Sci*. 2010;36(1):19–29.
- Lohman TM, Tomko EJ, Wu CG. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol*. 2008;9:391–401.
- Marians KJ. Crawling and wiggling on DNA: structural insights to the mechanism of DNA unwinding by helicases. *Structure*. 2000;8:R227–35.
- Pyle AM. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys*. 2008;37:317–36.
- Rasnik I, Myong S, Ha T. Unraveling helicase mechanisms one molecule at a time. *Nucleic Acids Res*. 2006;34:4225–31.
- Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem*. 2007;76:23–50.
- Spies M, Bianco PR, Dillingham MS, Handa N, Baskin RJ, Kowalczykowski SC. A molecular throttle: the recombination hotspot *chi* controls DNA translocation by the RecBCD helicase. *Cell*. 2003;114:647–54.
- Toseland CP, Webb MR. Fluorescence tools to measure helicase activity in real time. *Methods*. 2010;51:259–68.
- van Brabant AJ, Stan R, Ellis NA. DNA helicases, genomic instability, and human genetic disease. *Annu Rev Genomics Hum Genet*. 2000;1:409–59.
- Webb MR. Development of fluorescent biosensors for probing the function of motor proteins. *Mol Biosyst*. 2007;3:249–56.
- Webb MR. Fluorescent biosensors to investigate helicase activity. *Methods Mol Biol*. 2010;587:13–27.

---

## Helix Tilt

- ▶ [Transmembrane Helix Orientation and Dynamics](#)

---

## Hemagglutinin – Computational Studies

Marc Willem van der Kamp  
Centre for Computational Chemistry, School of  
Chemistry, University of Bristol, Bristol, UK

## Synonyms

[Haemagglutinin](#)

## Definition

A viral glycoprotein responsible for binding the virus to the cell that is being infected.

## Basic Characteristics

When a virus enters a host, it needs to bind to host cells and subsequently induce fusion of its membrane with the host cell membrane. The hemagglutinin protein on the viral envelope plays a crucial role in this mechanism: it binds to sialic acid analogues on the host cell, inducing endosomal uptake of the virus particle. Many computational studies, typically using ▶ [Molecular Dynamics Simulations](#), have been used to investigate the recognition between different viral hemagglutinins and host cell glycans (Kasson et al. 2009; Newhouse et al. 2009). After binding to the host cell, the virus

needs ensure that its contents, most crucially its genome, enter the host cell cytoplasm. Hemagglutinin plays a crucial role in this process by changing conformation and exposing a previously hidden hydrophobic portion of its peptide chain, the so-called fusion peptide. This peptide inserts itself into the endosomal membrane and eventually causes the viral and endosomal membranes to fuse together. This in turn allows the virus contents to enter the cytoplasm. ► [Molecular Dynamics Simulations](#), including coarse-grained approaches, have been applied to investigate the interaction between “fusion peptides” and membranes (Lague et al. 2005; Kasson et al. 2006; Fuhrmans et al. 2009).

When the virus is replicated in the host cell, hemagglutinin is initially produced as an inactive protein precursor (HA0) that is subsequently cleaved into two active parts, HA1 and HA2, which mediate virus to host cell binding and membrane fusion, respectively. How this proteolytic activation is catalyzed by host proteases is not yet clear. Rungrotmongkol et al. therefore used ► [QM/MM Methods](#) to study the reaction mechanism of the first step of the acylation process by the host protease furin, with the cleavage site of the avian influenza virus subtype H5N1 hemagglutinin as its substrate (Rungrotmongkol et al. 2009). B3LYP/6-31+G\* corrected PM3/CHARMM22 potential energy surfaces were calculated and analyzed in detail, indicating the specific roles of residues in catalysis of this reaction.

## Cross-References

- [Lectins](#)
- [Molecular Dynamics Simulations of Lipids](#)
- [QM/MM Methods](#)

## References

- Fuhrmans M, Knecht V, et al. A single bicontinuous cubic phase induced by fusion peptides. *J Am Chem Soc.* 2009;131(26):9166–7.
- Kasson PM, Kelley NW, et al. Ensemble molecular dynamics yields submillisecond kinetics and intermediates of membrane fusion. *Proc Natl Acad Sci USA.* 2006;103(32):11916–21.
- Kasson PM, Ensign DL, et al. Combining molecular dynamics with Bayesian analysis to predict and evaluate ligand-binding mutations in influenza hemagglutinin. *J Am Chem Soc.* 2009;131(32):11338–40.

- Lague P, Roux B, et al. Molecular dynamics simulations of the influenza hemagglutinin fusion peptide in micelles and bilayers: conformational analysis of peptide and lipids. *J Mol Biol.* 2005;354(5):1129–41.
- Newhouse EI, Xu D, et al. Mechanism of glycan receptor recognition and specificity switch for Avian, Swine, and human adapted influenza virus hemagglutinins: a molecular dynamics perspective. *J Am Chem Soc.* 2009;131(47):17430–42.
- Rungrotmongkol T, Decha P, et al. Combined QM/MM mechanistic study of the acylation process in furin complexed with the H5N1 avian influenza virus hemagglutinin's cleavage site. *Proteins-Struct Funct Bioinform.* 2009;76(1):62–71.

---

## Heme Dioxygenases – Computational Studies

Patrick von Glehn

Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, UK

## Synonyms

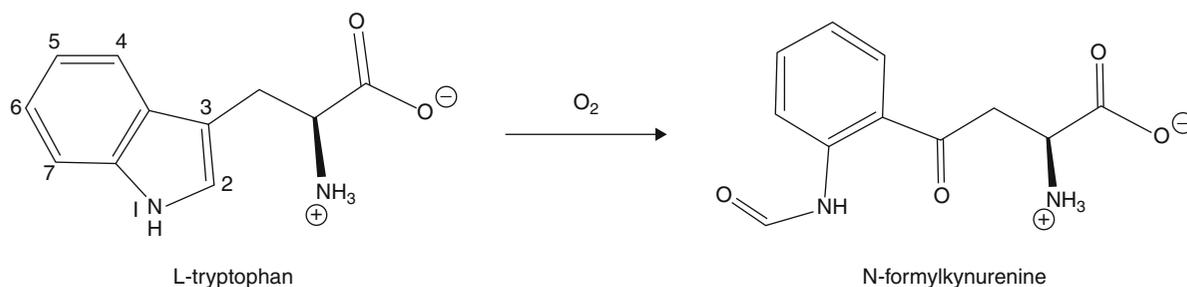
[Indoleamine 2,3-dioxygenase 1 \(IDO\)](#); [Tryptophan 2,3-dioxygenase \(TDO\)](#)

## Definition

Heme dioxygenase enzymes catalyze the conversion of L-Tryptophan to *N*-Formylkynurenine via the insertion of both atoms of heme-bound molecular dioxygen at the C2 and C3 positions of the substrate ([Fig. 1](#)).

One member of the family, indoleamine 2,3-dioxygenase 1 (IDO1) has been implicated in tumor immune escape and is seen as a good target for anticancer drugs (Röhrig et al. 2010). An important criterion for design of IDO1 inhibitors is that they should avoid interaction with another heme dioxygenase, tryptophan 2,3-dioxygenase (TDO). TDO has much higher specificity for L-tryptophan, although its active site is structurally very similar to that of IDO1. Consequently, the elucidation of the substrate binding and reaction mechanisms of these two enzymes has been a topic of considerable interest to which computational investigations have made important contributions.

Early mechanisms involved the abstraction of the indole N-H proton by either an active site base or by



**Heme Dioxygenases – Computational Studies, Fig. 1** Reaction catalyzed by heme dioxygenase enzymes

the heme-bound dioxygen as the initial step (Sono et al. 1996; Sugimoto et al. 2006). This was considered essential because experimental evidence indicated that 1-methyl-tryptophan (1-Me-Trp), which does not have an indole N-H hydrogen, is a competitive inhibitor of both enzymes. Subsequent steps were suggested to be nucleophilic attack on the distal oxygen at the C3 position followed by product formation via either a dioxetane intermediate or a Criegee rearrangement.

A density functional theory (DFT) study on a model system (comprised of a truncated oxy-heme group with imidazole representing the axial histidine ligand and 3-methylindole representing the substrate) was the first computational study of the heme dioxygenase reaction mechanism (Chung et al. 2008). Whereas previous mechanisms had assumed the oxy-heme complex to be in the closed shell ferrous dioxygen state Fe(II)-O<sub>2</sub>, Chung et al. found the open shell singlet state to be significantly lower in energy, which (based on Mulliken spin population analysis) was considered to a ferric superoxide complex Fe(III)-O<sub>2</sub><sup>-</sup>. It should be noted that DFT is known to favor a ferric superoxide description of oxy-heme while higher level multi-configurational calculations have shown a mixture of many different valence bond states, including the two mentioned above, to be a better model (Jensen et al. 2005).

The results of the DFT study showed that mechanisms involving indole N-H proton abstraction by dioxygen are very unfavorable. Newly proposed mechanisms starting with electrophilic addition from the Fe(II)-O<sub>2</sub> closed shell singlet state or radical addition from the Fe(III)-O<sub>2</sub><sup>-</sup> triplet or open shell singlet states gave more reasonable barriers. Addition to the C2 carbon was found to be lower in energy than to the C3 carbon. Following addition, the Criegee rearrangement pathway was shown to be very

unfavorable. Homolytic cleavage of the O-O bond leading to the formation of an epoxide and the high valent Fe(IV) = O species, compound II, was considered to be feasible though higher in energy than the dioxetane pathway. Shortly afterward, compound II was identified by resonance Raman spectroscopy during IDO1 turnover (although not during TDO turnover) (Lewis-Ballester et al. 2009).

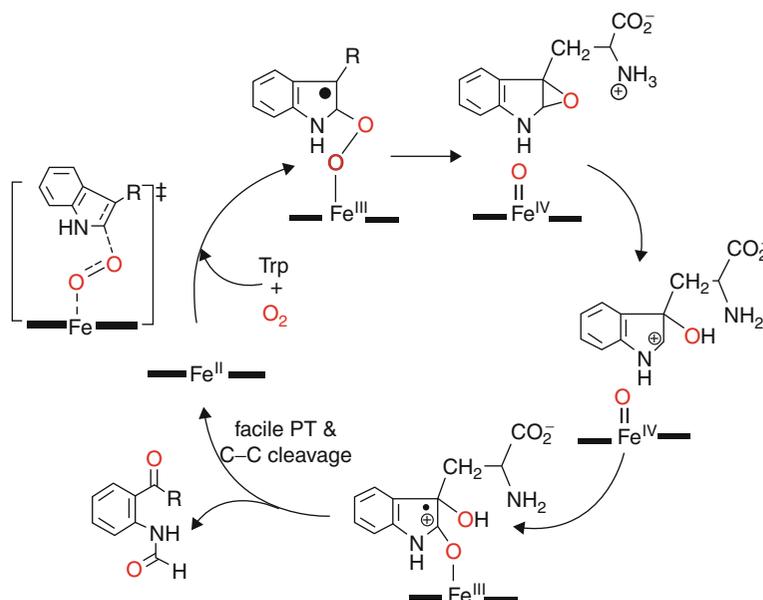
The resolving of crystal structures for human IDO1 and bacterial TDO provided a starting point for hybrid quantum mechanics/molecular mechanics (QM/MM) calculations which have been performed by three groups, one of whom studied both IDO1 and TDO (Capece et al. 2012) while the other two studied TDO only (Guallar and Wallrapp 2010; Chung et al. 2010). Each used different software packages and procedures in their calculations, but the general approach was the same: using molecular mechanics force fields, several nanoseconds of molecular dynamics simulations were run from which structures were chosen as starting points to explore the DFT QM/MM potential energy surfaces (Fig. 2).

These QM/MM studies all concluded that abstraction of the indole N-H proton does not occur because no minimum is observed along any of the proton transfer reaction coordinates investigated. This is in agreement with experimental results that showed that the only candidates for the proposed active site bases (His55 in TDO and Ser167 in IDO1) are not essential to activity and that contrary to previous evidence, 1-Me-Trp is turned over slowly by IDO1 and some TDO mutants (Chauhan et al. 2009). Capece et al. and Chung et al. found the dioxetane pathway to be infeasible.

In all these computational studies, the open shell singlet Fe(III)-O<sub>2</sub><sup>-</sup> state was calculated to be the lowest energy state of the Michaelis complex and radical

### Heme Dioxygenases – Computational Studies,

**Fig. 2** Heme dioxygenase reaction mechanism proposed by Chung et al. based on QM/MM calculations (Reprinted with permission from Chung et al. 2012. Copyright 2012 American Chemical Society)



addition to the C2 carbon was found to occur followed by the formation of an epoxide and compound II. Oxygen addition to the C3 carbon did not result in a minimum on any of the potential energy surfaces. All of the QM/MM studies found epoxide and compound II formation to be the rate-determining step, with energy barriers ranging from 8 to 20 kcal/mol. One notable difference between the studies is that the alkylperoxo species resulting from radical addition was identified as being a transition state by Capece et al., whereas the studies by Chung et al. and Guallar and Wallrapp found the alkylperoxo species to be an intermediate.

In the calculations by Wallrapp and Guallar, following epoxide formation, oxo attack by compound II occurs at the C2 carbon concerted with C2-C3 bond breaking to afford an unstable six-membered ring. The transition state for this step has an energy of around 10 kcal/mol relative to the reactant complex. After the facile opening of this ring, the final step is the breaking of the Fe-O bond, during which a change in spin state from singlet to triplet is thought to occur.

Chung et al. calculated pathways in both the singlet and triplet states for the first two steps and found that the surfaces cross before the transition state to compound II and epoxide formation where the triplet surface becomes lower in energy. The rest of the reaction profile is given on the triplet surface only,

although how or where a change in spin state may occur is not discussed. Regiospecific epoxide ring opening is assisted by transfer of a proton from the substrate NH<sub>3</sub><sup>+</sup> group on to the epoxide oxygen, with a transition state with energy of around -9 kcal/mol relative to the reactant complex (much more favorable than the third step in the calculations by Wallrapp and Guallar). The compound II oxygen then attacks at the unsaturated C2 position followed by the final step, which involves proton transfer back to the Tryptophan NH<sub>3</sub><sup>+</sup> group, concerted with C2-C3 and Fe-O bond cleavage.

The calculations of Capece et al. show that IDO1 and TDO follow the same reaction mechanism, which aside from the concerted versus stepwise nature of the epoxide and compound II formation is qualitatively the same as that proposed by Chung et al. for TDO.

Consequently, the experimentally observed differences in substrate selectivity between the heme dioxygenases may have their origins in differences in substrate binding rather than reaction mechanism. Substrate binding and selectivity in IDO1 and TDO have also been investigated with classical molecular dynamics simulations (Capece et al. 2010).

### Cross-References

► [QM/MM Methods](#)

## References

- Capece L, Arrah M, Roitberg AE, Yeh SR, Marti MA, Estrin DA. Substrate stereo-specificity in tryptophan dioxygenase and indoleamine 2,3-dioxygenase. *Proteins*. 2010;78(14):2961–72.
- Capece L, Lewis-Ballester A, Yeh SR, Estrin DA, Marti MA. Complete reaction mechanism of indoleamine 2,3-dioxygenase as revealed by QM/MM simulations. *J Phys Chem B*. 2012;116(4):1401–13.
- Chauhan N, Thackray SJ, Rafice SA, Eaton G, Lee M, Efimov I, Basran J, Jenkins PR, Mowat CG, Chapman SK, Raven EL. Reassessment of the reaction mechanism in the heme dioxygenases. *J Am Chem Soc*. 2009;131(12):4186–7.
- Chung LW, Li X, Sugimoto H, Shiro Y, Morokuma K. Density functional theory study on a missing piece in understanding of heme chemistry: the reaction mechanism for indoleamine 2,3-dioxygenase and tryptophan 2,3-dioxygenase. *J Am Chem Soc*. 2008;130(37):12299–309.
- Chung LW, Li X, Sugimoto H, Shiro Y, Morokuma K. ONIOM study on a missing piece in our understanding of heme chemistry: bacterial tryptophan 2,3-dioxygenase with dual oxidants. *J Am Chem Soc*. 2010;132(34):11993–12005.
- Guallar V, Wallrapp FH. QM/MM methods: looking inside heme proteins biochemistry. *Biophys Chem*. 2010;149(1–2):1–11.
- Jensen KP, Roos BO, Ryde U. O<sub>2</sub>-binding to heme: electronic structure and spectrum of oxyheme, studied by multiconfigurational methods. *J Inorg Biochem*. 2005;99(1):45–54.
- Lewis-Ballester A, Batabyal D, Egawa T, Lu C, Lin Y, Marti M, Capece L, Estrin DA, Yeh SR. Evidence for a ferryl intermediate in a heme-based dioxygenase. *PNAS*. 2009;106(41):17371–6.
- Röhrig UF, Awad L, Grosdidier A, Larrieu P, Stroobant V, Colau D, Cerundolo V, Simpson AJG, Vogel P, Van den Eynde BJ, Zoete V, Michielin O. Rational Design of Indoleamine 2,3-Dioxygenase Inhibitors. *J Med Chem*. 2010;53(3):1172–89.
- Sono M, Roach MP, Coulter ED, Dawson JH. Heme-Containing Oxygenases *Chemical Reviews*. 1996;96(7):2841–88.
- Sugimoto H, Oda S, Otsuki T, Hino T, Yoshida T, Shiro Y. Crystal structure of human indoleamine 2,3-dioxygenase: catalytic mechanism of O<sub>2</sub> incorporation by a heme-containing dioxygenase. 2006;103(8):2611–6.

---

## Heme Peroxidases

Emma L. Raven  
Department of Chemistry, University of Leicester,  
Leicester, UK

## Synonyms

Peroxidases

## Definition

Heme peroxidases catalyze the hydrogen peroxide-dependent oxidation of a variety of different substrates by means of oxidized Compound I and Compound II intermediates.

## Basic Characteristics

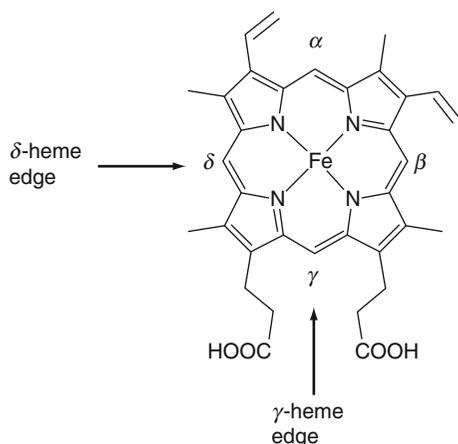
Heme peroxidases have a long history and are widely distributed in biological systems. Horseradish peroxidase (HRP), which is one of the most famous members of the family, was first implicated as an oxidizing species in horseradish roots as early as 1810 (Planche 1810) and in the years to follow the colorful spectra of its various reaction intermediates played a key part in the development of rapid kinetics methods (reviewed in Kresge et al. 2004). Entire books have been written on heme peroxidases, exploring their reactivity, properties, and biological function (Everse et al. 1991; Dunford 2010). Peroxidases have been characterized which utilize both *c*-type hemes (bacterial diheme peroxidases) and redox-active selenocysteine residues (glutathione peroxidases). A separate group of vanadium-containing peroxidases are also known which catalyze the H<sub>2</sub>O<sub>2</sub>-dependent oxidation of halides (Dunford 2010). This article describes the structure and mechanism of the most common and widely distributed peroxidases – those which possess *b*-type hemes.

## Heme Structure

Like the globins and the *b*-type ► [cytochromes](#), heme peroxidases contain heme *b* as a prosthetic group, [Fig. 1](#), which (unlike the cytochromes *c*) is bound to the protein through non-covalent interactions. In the resting state, this heme group is in the ferric (Fe<sup>III</sup>) form and is either five-coordinated, to an axial histidine residue, or six-coordinated with a weakly bound water molecule or hydroxide as the sixth ligand.

## Structure

Structures for a number of peroxidase enzymes have been solved (reviewed recently in Poulos 2010). Despite attracting much of the early attention, its heavily glycosylated structure meant that HRP proved



**Heme Peroxidases, Fig. 1** The structure of heme *b*

very difficult to crystallize, and a structure did not emerge until fairly recently (Gajhede et al. 1997). Cytochrome *c* peroxidase (CcP), on the other hand, proved to be much more obliging and crystallizes readily. The structure of CcP was thus the first to emerge, and for this reason served as a benchmark for other peroxidases. Since then, structures for numerous other peroxidases have been solved, including, amongst others, manganese peroxidase (MnP), chloroperoxidase, ascorbate peroxidase (APX), and lignin peroxidase (Poulos 2010).

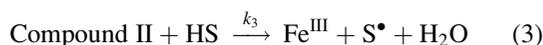
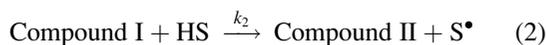
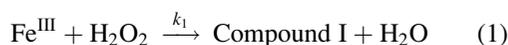
Within the heme active site, there is high sequence homology across all peroxidase structures. The active site of cytochrome *c* peroxidase is shown in Fig. 2. There is a “proximal” histidine ligand (His175) which is hydrogen bonded to Asp and Trp residues on the proximal side of the heme. On the distal side, there is a “distal” histidine residue together with distal Arg and Trp residues. In most other peroxidases, such as HRP, the proximal and distal Trp residues are replaced with Phe, Fig. 2. This replacement of Trp for Phe has consequences for the Compound I intermediate, as explained below.

The heme active sites in the peroxidases are closely related to those in the globins, which have a similar proximal/distal histidine structure but which are generally more hydrophobic than in the peroxidases. There has been ongoing interest in the factors that control O<sub>2</sub> binding and peroxidase activity in these proteins (see, e.g., Ozaki et al. 2001). In fact, the relationships and differences in reactivity across all heme proteins – including the ► [cytochrome P450s](#),

globins, and other heme oxygenases – has attracted considerable attention (Ortiz de Montellano 1987; Dawson 1988; Karlin 2010).

## Mechanism

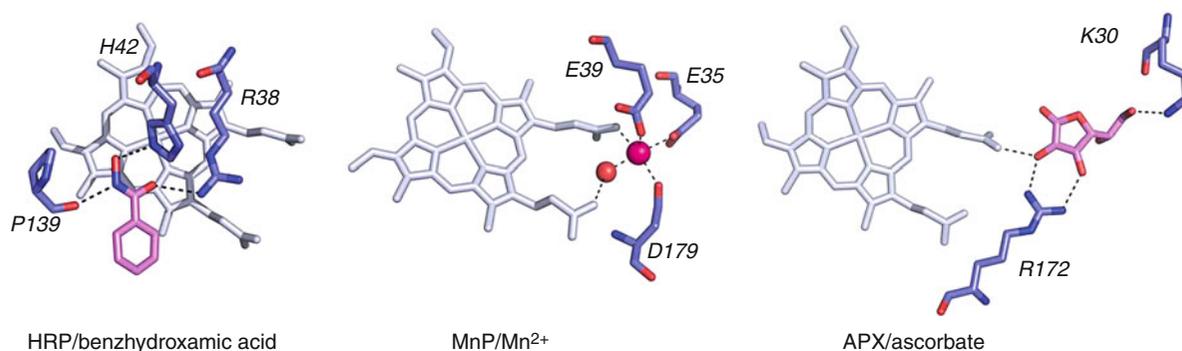
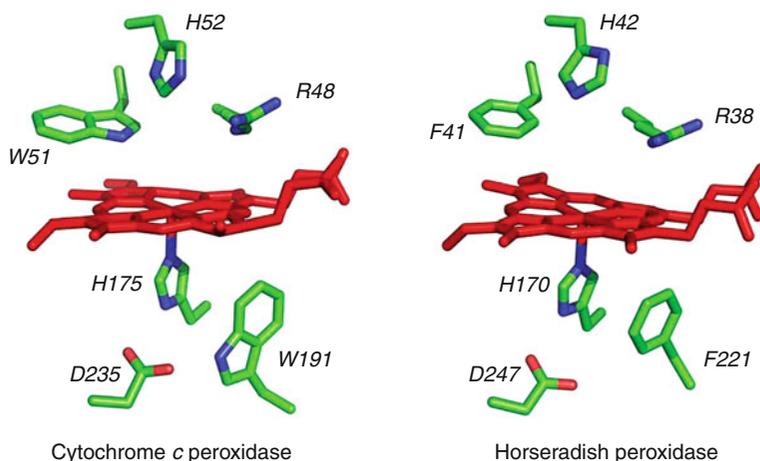
The reaction mechanism has been well-studied across a large number of heme peroxidases (Dunford 2010). The reaction proceeds by means of initial reaction of the ferric enzyme with H<sub>2</sub>O<sub>2</sub> to form a two-equivalent oxidized heme intermediate, known as Compound I (Eq. 1) and releasing one mole of water. Compound I is formally oxidized by two-electrons (compared to the ferric state), and in most cases contains a ferryl heme (Fe(IV) = O) and a porphyrin π-cation radical (see below). This is followed by two, sequential single electron reductions of Compound I by substrate, returning the enzyme to the resting, ferric state through a second intermediate, known as Compound II, which is formed from one-electron reduction of Compound I, Eqs. 2–3. The substrate varies across peroxidases, but here is represented as HS where each HS donates an electron and a proton to Compound I, thus providing the protons necessary for release of the second mole of water from Compound II.



In HRP the formation of Compound I is accompanied by a color change of the ferric protein from red/brown to green, with the green color characteristic of a Compound I species that contains both ferryl heme (Fe(IV) = O) and a porphyrin π-cation radical. In many peroxidases, this Compound I species is observable by rapid ► [stopped-flow methods](#). CcP, famously, does not use a porphyrin pi cation radical in its Compound I, but instead uses protein radical located on Trp191 (Sivaraja et al. 1989). This was one of the first examples of a stable protein radical in any protein. Using mutagenesis, much attention has focused on the role of various active site residues on the reaction mechanism, in particular the reaction with H<sub>2</sub>O<sub>2</sub> and the intermediates involved (Erman and Vitello 2002).

**Heme Peroxidases,**

**Fig. 2** Comparison of the active site structures of CcP and HRP



**Heme Peroxidases, Fig. 3** The two substrate binding locations observed in heme peroxidases. The  $\delta$ -heme edge is typically used by aromatic substrates, as typified in the HRP/benzhydroxamic acid complex (*left*, Henriksen et al. 1998).

Substrate binding is also possible at the  $\gamma$ -heme edge, shown here for the MnP/Mn<sup>2+</sup> (*centre*, Sundaramoorthy et al. 1994) and ascorbate peroxidase/ascorbate (*right*, Sharp et al. 2003) complexes.

**Substrate Binding**

Most commonly, the substrate for peroxidase enzymes is a small organic substrate. Identification of the substrate binding sites in heme peroxidases lagged a long way behind the elucidation of structures and mechanism as described above. The first structures that were solved for a peroxidase-substrate complex were for the CcP/cytochrome *c* and MnP/Mn<sup>2+</sup> complexes (reviewed recently Gumiero et al. 2010), but these enzymes have somewhat atypical substrates (i.e., not a small organic molecule). Across all peroxidases (and with the exception of CcP, with its anomalous substrate), it has emerged from structural analyses that there are two binding locations for the substrate – one close to the so-called  $\delta$ -heme edge of the heme group, which is typically used for aromatic substrates, and one close to the  $\gamma$ -heme edge, **Figs. 1 and 3**

(reviewed in Gumiero et al. 2010) see also below for figure legend. This ability to accommodate different kinds of substrates at different locations might well account for the tremendous versatility of these enzymes.

**Cross-References**

- ▶ [Bacterial Globins](#)
- ▶ [Cytochromes](#)
- ▶ [Hemes](#)

**References**

- Dawson JH. Probing structure-function relations in heme-containing oxygenases and peroxidases. *Science*. 1988;240:433–9.

- Dunford HB. Peroxidases and catalases: biochemistry, biophysics, biotechnology and physiology. 2nd ed. Chichester: Wiley; 2010.
- Erman JE, Vitello LB. Yeast cytochrome *c* peroxidase: mechanistic studies via protein engineering. *Biochim Biophys Acta*. 2002;1597:193–220.
- Everse J, Everse KE, Grisham MB. Peroxidases in chemistry and biology. Boca Raton: CRC Press; 1991.
- Gajhede M, Schuller DJ, Henricksen A, Smith AT, Poulos TL. Crystal structure of horseradish peroxidase C at 2.15 angstrom resolution. *Nat Struct Biol*. 1997;4:1032–8.
- Gumiero A, Murphy EJ, Metcalfe CL, Moody PC, Raven EL. An analysis of substrate binding interactions in the heme peroxidase enzymes: a structural perspective. *Arch Biochem Biophys*. 2010;500(1):13–20.
- Henriksen A, Schuller DJ, Meno K, Welinder KG, Smith AT, Gajhede M. Structural interactions between horseradish peroxidase C and the substrate benzhydroxamic acid determined by X-ray crystallography. *Biochemistry*. 1998;37:8054–60.
- Karlin KD. Bioinorganic chemistry: model offers intermediate insight. *Nature*. 2010;463:168–9.
- Kresge N, Simoni RD, Hill RL. Britton chance: Olympian and developer of stop-flow methods. *J Biol Chem*. 2004;279:e10.
- Ortiz de Montellano PR. Control of the catalytic activity of prosthetic heme by the structure of hemoproteins. *Acc Chem Res*. 1987;20:289–94.
- Ozaki S-I, Roach MP, Matsui T, Watanabe Y. Investigations of the roles of the distal heme environment and the proximal heme iron ligand in peroxide activation by heme enzymes via molecular engineering of myoglobin. *Acc Chem Res*. 2001;34:818–25.
- Planche LA. *Bull Pharm*. 1810;2:578–80.
- Poulos TL. Thirty years of heme peroxidase structural biology. *Arch Biochem Biophys*. 2010;500(1):3–12.
- Sharp KH, Mewies M, Moody PCE, Raven EL. Crystal structure of the ascorbate peroxidase-ascorbate complex. *Nat Struct Biol*. 2003;10:303–7.
- Sivaraja M, Goodin DB, Smith M, Hoffman BM. Identification by ENDOR of Trp191 as the free-radical site in cytochrome *c* peroxidase compound ES. *Science*. 1989;245:738–40.
- Sundaramoorthy M, Kishi K, Gold MH, Poulos TL. The crystal structure of manganese peroxidase from *Phanerochaete cryosporium* at 2.06 Å resolution. *J Biol Chem*. 1994;269:32759–67.

---

## Hemes

Brian R. Gibney  
Department of Chemistry, Brooklyn College – The  
City University of New York, Brooklyn, NY, USA

## Synonyms

Hemin; Metalloporphyrin; Porphyrin; Protoheme;  
Protoporphyrin IX

## Definition

Heme, the ferrous iron complex of protoporphyrin IX, is a common prosthetic group in proteins including the electron transfer ► [cytochromes](#), the respiratory complexes, and the oxygen carrier hemoglobin.

## Basic Characteristics

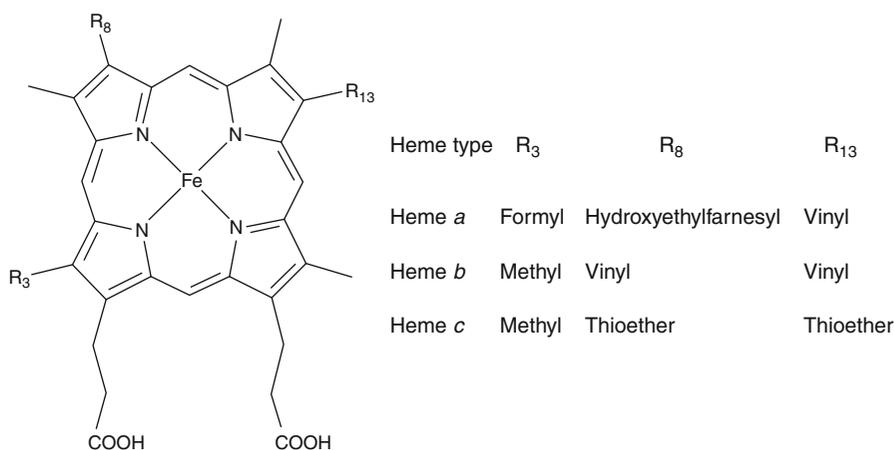
### Introduction

Iron porphyrins, including iron (protoporphyrin IX) or heme *b*, are one of the most versatile and visible classes of redox cofactors utilized in biochemistry and biophysics. Naturally occurring from archaeobacteria to mammals, heme proteins carry out diverse biochemical tasks such as electron transfer, substrate oxidation, metal-ion storage, ligand sensing/transport, and gene regulation. Thus, heme proteins are critical components in numerous biological processes including hormone and steroid biosynthesis, drug metabolism, aerobic respiration, and even programmed cell death. The dysfunction of natural heme proteins leads to a myriad of human diseases including sickle cell anemia, diabetes, and Alzheimer's and coronary artery disease.

### Chemical Structure of Hemes

The basic structure of a heme is the tetrapyrrole macrocycle with dimensions of about 10Å × 10Å × 3Å (Dolphin 1978). The aromatic nature of the porphyrin encourages a planar conformation of the macrocycle and leads to significant absorption bands in the visible region. The most intense band, the Soret or  $\gamma$ -band, typically has a molar extinction coefficient value greater than 100,000 M<sup>-1</sup> cm<sup>-1</sup> which makes it ideal for optical and ► [resonance Raman](#) spectroscopies. The center of the porphyrin contains two imine nitrogens (–C=N–C–) and two amine nitrogens (–C–NH–C–), the latter of which are deprotonated for iron incorporation. The four nitrogens of the porphyrin dianion serve as equatorial ligands to the bound iron. Heme iron commonly exists in either the +2 (ferrous) or +3 (ferric) oxidation states, the latter of which is active in X-band electron paramagnetic resonance (EPR) spectroscopy. Thus, the heme structure shown in [Fig. 1](#) is formally neutral as an Fe(II) complex, and a monocation as the Fe(III) complex. Ferric hemes can be oxidized by one electron to the ferryl state, Fe(IV),

**Hemes, Fig. 1** Chemical structure of biological hemes



with a further one-electron oxidation being centered on the porphyrin macrocycle, formally generating an Fe (IV) (porphyrin  $\pi$ -cation radical) complex.

Figure 1 also shows how the peripheral architecture of the macrocycle varies between biological heme types. Heme *b*, or protoheme, is the most common chemical structure observed in biology and is the prosthetic group found in the oxygen transport globins, drug metabolizing ► [cytochromes P450](#), and the electron transfer cytochromes *b* (Antonini and Brunori 1971; Ortiz de Montellano 1995). The periphery of heme *b* contains a methyl group at R<sub>3</sub>, and vinyl groups at positions R<sub>8</sub> and R<sub>13</sub>. This protoporphyrin IX macrocycle is itself biosynthesized from 5-aminolevulinic acid (ALA) via six consecutive enzymatic steps that are universal to all living organisms. The final step in heme *b* biosynthesis is the insertion of the iron by the enzyme ferrochelatase. Once assembled, heme *b* binds to proteins via one or two axial ligand(s) to the heme iron, hydrogen bonding interactions between the propionic acid groups and local amino acids, and hydrophobic interactions between the macrocycle and the hydrophobic core of the protein. While the full range of heme protein biological functions are observed in the subset of heme proteins containing just heme *b*, biology utilizes several other porphyrin architectures in heme proteins.

The second most common biological heme structure is heme *c* which is found in cytochromes *c* (cyts *c*) and in cytochrome *f* of the ► [cytochrome \*b\*<sub>6</sub>\*f\* complex](#) (Scott and Mauk 1996). Heme *c* differs from heme *b* in that the porphyrin periphery is covalently linked at positions R<sub>8</sub> and/or R<sub>13</sub> by thioether bonds to cysteine residues. These cysteines are often observed

in a Cys-Xaa-Xaa-Cys-His sequence motif in cyt *c* in which Xaa represents any amino acid and the histidine serves as one of the axial ligands to the bound iron. The thioether linkages have been found to be critical to cyt *c* structure and function as their removal often results in heme dissociation with concomitant protein unfolding. The biosynthesis of heme *c* from its heme *b* precursor occurs by one of three distinct enzymatic pathways in biology. The existence of *c*-type cytochromes has produced many hypotheses concerning the advantages of covalently attaching the heme to the protein scaffold. The thioether linkages in cyts *c* may be an artifact of evolution, may guide protein folding, may allow for high heme to protein ratios which may facilitate their electron transfer function, or may be present to enforce the coordination of weak ligands, e.g., methionine to ferric heme, by increasing their effective concentration.

While heme *b* and heme *c* constitute the majority of hemes in biological systems, other hemes including hemes *a*, *d*, *o*, siroheme, coproporphyrin, and P460 are also observed. Heme *a*, structure in Fig. 1, is vital to aerobic respiration in humans where it is solely found in the active site of the integral membrane protein complex, cytochrome *c* oxidase (CcO). Heme *a* is biosynthetically derived from heme *b* in two enzymatic steps. In the first step, heme *b* is transformed into heme *o* by converting the vinyl group at R<sub>8</sub> into a hydroxyethylfarnesyl group. In the second step, the methyl group at R<sub>3</sub> in heme *o* is oxidized to a formyl group to generate heme *a*. The hydroxyethylfarnesyl group of heme *a* may perform one or more of the following roles in terminal oxidase function: acting as a lipophilic anchor, modulating electron transfer

pathways, functioning in the translocation of protons and aiding in exogenous ligand exchange reactions. The formyl group of heme *a* side chain raises its redox potential relative to heme *b* or *o*, making it a better electron acceptor.

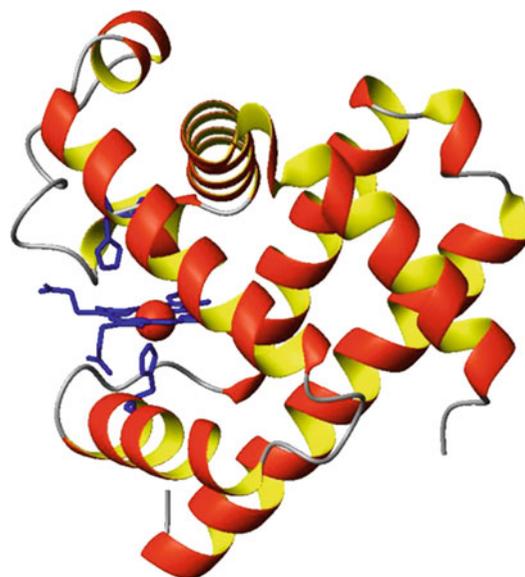
### Heme Proteins

Approximately 5% of all structurally characterized proteins in the RCSB Protein Data Bank contain heme (Reedy and Gibney, 2004). As mentioned earlier, the hemes interact with proteins via heme iron axial ligation, propionic acid hydrogen bonding, macrocycle hydrophobic interactions, and covalent linkage to the protein in cyts *c*. The prevalent protein secondary structure around hemes is the alpha helix, but there are instances where hemes are bound to beta-sheets. The heme iron may be bound by amino acids in the protein or by exogenous ligands and the protein may include additional hemes or other cofactors.

Heme iron is known to bind a wide range of endogenous and exogenous ligands. The heme iron is usually five- or six-coordinate in the protein scaffold. Histidine and methionine are the most common amino acid ligands to heme iron, with examples of lysine, asparagine, cysteinate, and tyrosinate ligation known. The heme coordination motif together with the constellation of amino acids surrounding the heme helps establish the chemical reactivity of the heme and thus its biochemical function.

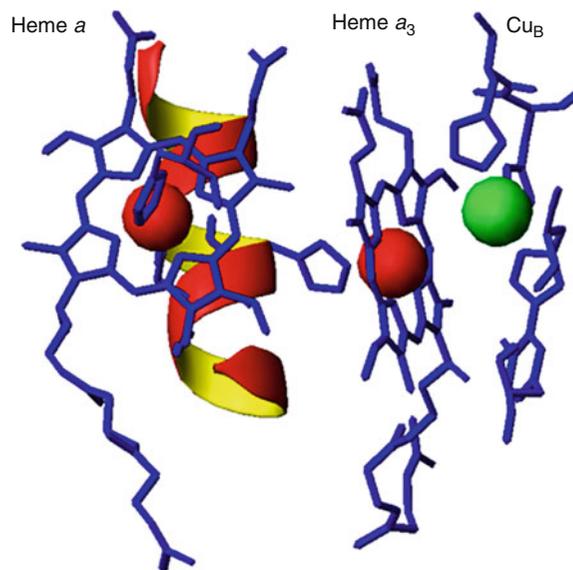
The function of hemes that are six-coordinate in both oxidation states, Fe(III) and Fe(II), is often electron transfer. An iron coordination motif that is invariant upon iron oxidation/reduction results in lower reorganization energies and more rapid electron transfer kinetics. The presence of six-coordinate also serves to limit the ability of exogenous ligands to bind the heme that would alter the heme reduction potential, or driving force for electron transfer. Conversely, the open coordination site in a five-coordinate heme protein facilitates the binding of an exogenous ligand.

Numerous five-coordinate heme proteins bind an exogenous ligand that is critical to their biological function. These include heme proteins that sense, transport, and chemically modify exogenous ligands. Heme proteins involved in sensing small molecules include the O<sub>2</sub>-sensor FixL and the NO-sensor soluble guanylyl cyclase. The CO-sensor protein CooA is six-coordinate, but transient dissociation of the ligating



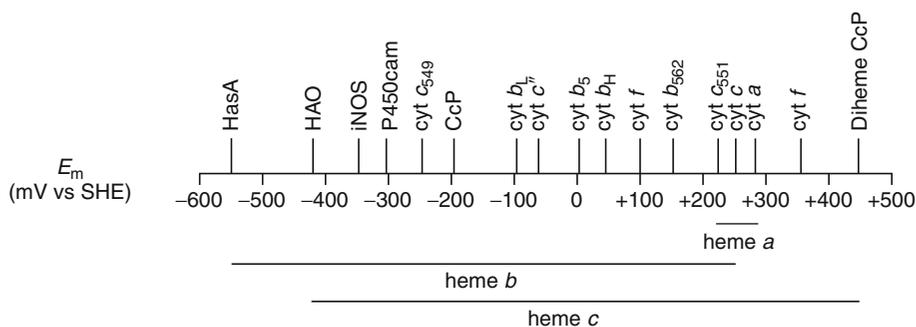
Myoglobin

**Hemes, Fig. 2** Structure of myoglobin



**Hemes, Fig. 3** Active site structure of cytochrome *c* oxidase

N-terminus provides the five-coordinate heme that binds CO. Five-coordinate heme proteins are also used for the transport of small molecules as exemplified by the O<sub>2</sub>-transporter hemoglobin and the NO-transporter nitrophorin. The bound exogenous ligand may also be used for chemical catalysis. For instance, heme peroxidases oxidize organic substrates using hydrogen peroxide and the drug metabolizing



**Hemes, Fig. 4** Midpoint reduction potential range of natural heme proteins (values)

cytochromes P450 catalyze the oxidation of organic substrates by  $O_2$  (Dunford 1999). In addition to these diatomic gases, five-coordinate heme proteins are observed to bind water (hydroxide), ammonia, cyanide, thiocyanide, azide, acetate, and sulfate.

Heme proteins may contain a single heme, multiple hemes, or hemes in conjunction with other cofactors. Figure 2 shows the structure of myoglobin, the prototypical example of a water-soluble *b*-type heme protein. As noted above, the alpha helical secondary structure and histidine coordination observed in myoglobin are common to many heme proteins. Multiheme proteins typically function as electron transfer conduits to buried active sites, e.g., hydroxylamine oxidoreductase. Hemes are also found in proteins that contain other cofactors including iron-sulfur clusters and copper ions. In the case of cytochrome *c* oxidase, the heterobimetallic active site shown in Fig. 3 is composed of heme  $a_3$  and  $Cu_B$  with the adjacent heme *a* serving an electron transfer role.

### Heme Reduction Potentials

The midpoint reduction potential value,  $E_m$ , of the iron in a heme protein establishes its driving force for electron transfer and is integrally related to its protein stability, protein folding pathway, and kinetics. In addition, the  $E_m$  value represents the *relative* affinity of the protein for heme in the two oxidation states. Heme proteins with large negative  $E_m$  values bind Fe(III) heme tighter than Fe(II) heme while heme proteins with large positive  $E_m$  values bind Fe(II) heme tighter than Fe(III) heme. Figure 4 shows that the Fe(III)/Fe(II) midpoint reduction potentials for natural heme proteins span an 1,100 mV range. The extremes are heme acquisition system A (HasA) at  $-550$  mV and diheme cytochrome *c* peroxidase at  $+450$  mV.

The reduction potential range of heme *a* in proteins is rather limited, while both heme *b* proteins and heme *c* proteins span a considerable range of reduction potential values.

### Cross-References

- ▶ [Catalases and Catalase-Peroxidases](#)
- ▶ [Cytochrome \*c\* Oxidase \(Complex IV\)](#)
- ▶ [Cytochromes](#)
- ▶ [Electron Transfer Cofactors](#)
- ▶ [Electron Transport Chains in Bacteria](#)
- ▶ [Mitochondrial Electron Transport](#)
- ▶ [Peroxidases](#)
- ▶ [Photosynthetic Electron Transport](#)
- ▶ [Proton-Coupled Electron Transfer](#)
- ▶ [Redox Potential](#)

### References

- Antonini E, Brunori M. Hemoglobin and myoglobin in their reactions with ligands. Amsterdam: North-Holland; 1971.
- Dolphin D. The porphyrins. New York: Academic; 1978.
- Dunford HB. Heme peroxidases. New York: Wiley; 1999.
- Ortiz de Montellano PR. Cytochrome P450: structure, function, and mechanism. 2nd ed. New York: Plenum; 1995.
- Reedy CJ, Gibney BR. Heme protein assemblies. Chem Rev. 2004;104:589–604.
- Scott RA, Mauk AG. Cytochrome *c* – a multidisciplinary approach. Sausalito: University Science; 1996.

### Hemin

- ▶ [Hemes](#)

---

## Hemoglobin-Like Proteins in Bacteria

- ▶ [Bacterial Globins](#)

---

## Hemoproteins

- ▶ [Cytochromes](#)

---

## Heptahelical Receptors

- ▶ [Dynamics of Helix 8 in GPCR Function](#)
- ▶ [G Protein–Coupled Receptor Activation Based on X-ray Structural Studies](#)

---

## Heterocorrelation

- ▶ [Infrared Spectroscopy: Data Analysis](#)

---

## Heteromeric Versus Homomeric Association of Protein Complexes

Jacqueline M. Matthews  
School of Molecular Bioscience, The University of  
Sydney, New South Wales, Sydney, Australia

### Synonyms

[Protein–protein interactions](#); [Subunit composition of protein complexes](#); [Superfamily heterodimers](#)

### Definition

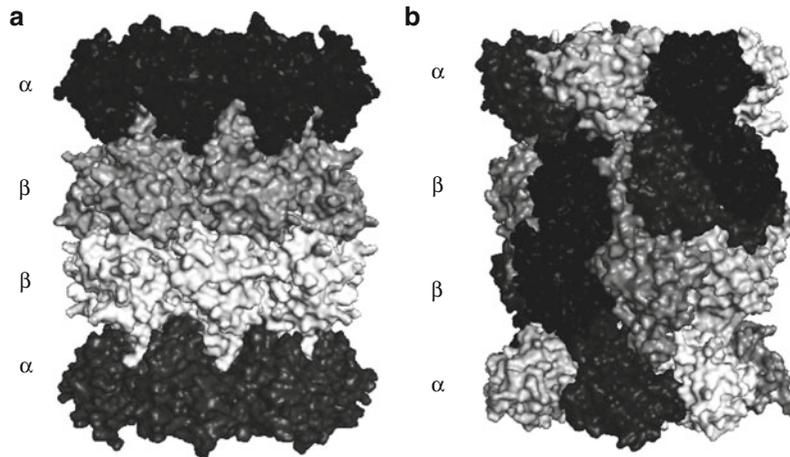
Many proteins form homodimers or other higher-order homo-oligomers (Marianayagam et al. 2004; Matthews 2012) and/or interact with other proteins to effect biological responses. Whereas prokaryotic multiprotein complexes tend to have a simple composition, the equivalent assembly in eukaryotes can be far

more complex. For example, the catalytic core units of proteasomes are made up of four seven-membered rings, comprising two alpha subunit and two beta subunit rings (Fig. 1). In bacteria and archaea, there is only one type of alpha and one type of beta subunit, but in eukaryotes, there are seven different types each of alpha or beta subunits with each ring containing all seven different subunits (Darwin 2009; Pearce et al. 2008). Whereas the central core of voltage-gated sodium ion channels in bacteria generally comprises four identical subunits, that in bacteria comprises a single-chain protein with four homologous domains (Charalambous and Wallace 2011). Proteins that regulate gene expression, which in prokaryotes are often homodimers or oligomers, tend to form multiprotein complexes in eukaryotes (Beckett 2001).

### Basic Characteristics

Differences in protein complex composition are reflected in the sizes of the genomes and the complexities of the organisms involved. Prokaryotic genomes tend to be small compared to eukaryotic genomes, whereas eukaryotic genomes can support the development and maintenance of highly complex organisms. In part, the expanded sizes of eukaryotic genomes are likely to have been caused by genome duplication events. Over time, identical copies of genes acquire mutations allowing them to diverge in terms of both sequence and function to form paralogs. With the occurrence of additional genome duplication events, families of related proteins are gradually formed (Pereira-Leal et al. 2007). Thus, protein that formed homodimers in parent organisms could have evolved families of related proteins that can form homomeric and/or heteromeric interactions with other family members. These “superfamily heterodimers” include cell surface receptors, enzyme complexes, transcription factors, and ion channels. Proteins that have numerous interaction partners tend to be of increased importance to the viability of an organism (Aragues et al. 2007), and many of these superfamily heterodimers are biologically important.

Different combinations and permutations of subunits in complexes tend to have different activities, such as transcription factor complexes targeting different DNA sequences or recruiting different cofactors. The exchange of a single component from the same



**Heteromeric Versus Homomeric Association of Protein Complexes, Fig. 1** Complexity of subunit composition in proteasomes and proteasome-like complexes. (a) The archaeal 20S proteasome contains a single type of alpha ( $\alpha$ ) and a single type of beta ( $\beta$ ) subunit arranged in seven-membered rings (PDB

accession code 1PMA). Here, the different rings are shown in different colors. (b) The eukaryotic 20S proteasome contains seven different types of  $\alpha$  and  $\beta$  subunits (PDB accession code 2NZJ). Each ring comprises one copy of each  $\alpha$  subunit, or one copy of each  $\beta$  subunit

family can transform a transcription complex from one that activates expression of a target gene to one that switches off expression (e.g., Jing et al. 2008). An ability to use regulatory proteins in a combinatorial fashion is generally thought to be a mechanism through which complexity in higher eukaryotes is achieved (e.g., Lange et al. 2011). By assembling multiprotein complexes with different activities, tight regulation of cellular events can be achieved through nonlinear responses to changes in protein concentration (Boulanger et al. 2003; Hansen et al. 2008).

Prokaryotes are still able to maintain high levels of regulation over the activity of multiprotein complexes through mechanisms including the regulation of protein oligomerization through posttranslational modification, small ligand binding, and/or DNA binding (Beckett 2001).

## Cross-References

- ▶ [Mass Spectrometry: Mapping Large Stable Protein Complexes](#)
- ▶ [Protein Structure](#)
- ▶ [Proteins: Relationship Among Divergence of Sequence, Structure, and Function](#)
- ▶ [RNA Polymerases and Transcription](#)

## References

- Aragues R, Sali A, Bonet J, Marti-Renom MA, Oliva B. Characterization of protein hubs by inferring interacting motifs from protein interactions. *PLoS Comp Biol.* 2007;3:1761–71.
- Beckett D. Regulated assembly of transcription factors and control of transcription initiation. *J Mol Biol.* 2001;314:335–52.
- Boulanger MJ, Chow D-C, Brevnova EE, Garcia KC. Hexameric structure and assembly of the interleukin-6/IL-6  $\alpha$ -receptor/gp130 complex. *Science.* 2003;300:2101–4.
- Charalambous K, Wallace BA. NaChBac: the long lost sodium channel ancestor. *Biochemistry.* 2011;50:6742–52.
- Darwin KH. Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis. *Nat Rev Microbiol.* 2009;7:485–91.
- Hansen G, Hercus TR, McClure BJ, Stomski FC, Dottore M, Powell J, Ramshaw H, Woodcock JM, Xu Y, Guthridge M, McKinstry WJ, Lopez AF, Parker MW. The structure of the GM-CSF receptor complex reveals a distinct mode of cytokine receptor activation. *Cell.* 2008;134:496–507.
- Jing H, Vakoc CR, Ying L, Mandat S, Wang H, Zheng X, Blobel GA. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol Cell.* 2008;29:232–42.
- Lange M, Demajo S, Jain P, Di Croce L. Combinatorial assembly and function of chromatin regulatory complexes. *Epigenomics.* 2011;3:567–80.
- Marianayagam NJ, Sunde M, Matthews JM. The power of two: protein dimerization in biology. *Trends Biochem Sci.* 2004;29:618–25.
- Matthews JM, editor. *Protein dimerization and oligomerization in biology.* Austin: Landes Bioscience/Springer; 2012.

Pearce MJ, Mintseris J, Ferreyra J, Gygi SP, Darwin KH. Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science*. 2008;322:1104–7.

Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol*. 2007;8:R51.

---

## Heteronuclear Multidimensional Nuclear Magnetic Resonance Spectroscopy

- ▶ [Multidimensional NMR Spectroscopy](#)
- ▶ [Structure Determination by NMR: Overview](#)

---

## Hidden Markov Modeling in Single-Molecule Biophysics

Mario R. Blanco<sup>1</sup>, Alexander E. Johnson-Buck<sup>2</sup> and Nils G. Walter<sup>2</sup>

<sup>1</sup>Department of Cellular and Molecular Biology, The University of Michigan, Ann Arbor, MI, USA

<sup>2</sup>Department of Chemistry, The University of Michigan, Ann Arbor, MI, USA

### Synonyms

[HMM](#); [Markov model](#)

### Definition

A hidden Markov model (HMM) is a probabilistic model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states.

### Introduction

The development of ▶ [single-molecule spectroscopy](#) has allowed for the investigation of a variety of biological questions previously inaccessible by ensemble techniques. The strength of single-molecule tools comes from the high-resolution data extracted from such experiments. Proper interpretation of these data requires efficient, unbiased analysis routines that are able to distinguish relevant signals from the

intrinsically noisy measurements. The hidden Markov model, a statistical algorithm initially developed for speech recognition, has been adapted for the analysis of a variety of single-molecule signals. In this article, we will give a general introduction to the theoretical basis of hidden Markov modeling and the various single-molecule techniques in which they have been co-opted for signal analysis.

### Basic Characteristics

Signals collected from single-molecule experiments can be described as a series of discrete states governed by an underlying physical property of the molecule(s) being interrogated. The discrete states are often obscured (hidden) by noise that is inherent to the experimental technique, making the identification and characterization of these states difficult. The hidden states are no longer efficiently detected by visual inspection or simple algorithms and doing so can introduce bias and an incomplete characterization of the underlying behavior(s). Probabilistic maximum-likelihood algorithms, like a hidden Markov model (HMM), have become the preferred method of analysis; they provide a mathematically derived routine that limits the possibility of user bias as well as providing a theoretical framework with which to interpret the quantitative results extracted from single-molecule experiments. A HMM describes a stochastic progression through a series of discrete states, where the likelihood of the next event in a series of observations can be predicted upon knowledge of the immediately preceding event and does not depend on knowledge of any of the prior events; that is, the process is Markovian (Fraser 2008). The main model assumptions of a HMM are as follows:

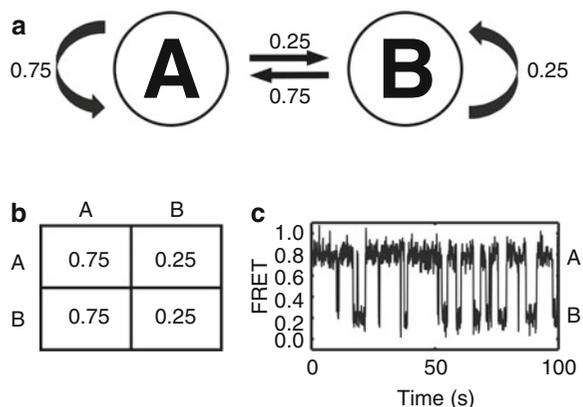
1. Given the current state, the probability of the current observation is independent of states and observations at all earlier times.
2. Given the current state, the probability of the next observed state occurring is independent of earlier states. More simply put, the future does not depend on the past.

HMMs are well suited for single-molecule analysis because of their ability to find discrete states, usually stable, biologically relevant conformations, within noisy time series data, and to reliably find the most probable path through these states. As a molecule

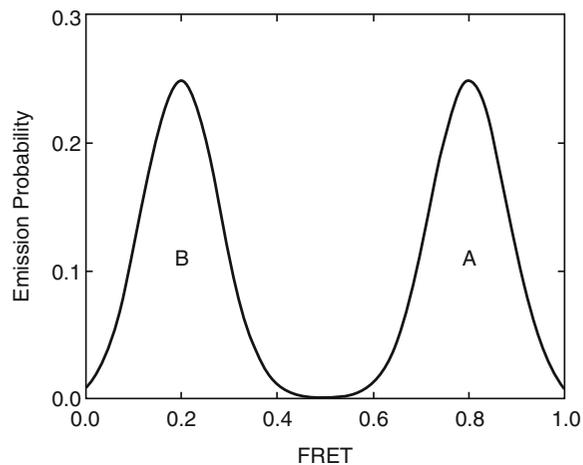
transitions from one stable conformation to another it is often the case that the process is Markovian and therefore governed by single exponential kinetics. Through the iterative optimization of the HMM parameters – the probability matrices of transition, emission, and initiation – a model is derived that best approximates the data. The transition probability matrix describes the probability of any one state changing to any other state or staying in the same state in the subsequent time step (Fig. 1). The emission probability distribution contains the probabilities of a specific signal value being emitted by each discrete state. Calculating an emission probability often requires an assumption of the noise in the system, usually shot noise that is efficiently approximated by simple Gaussian distributions (Fig. 2). The initiation probability matrix gives the probabilities of starting at each of the possible discrete states.

During the evaluation of single-molecule data one does not usually know what the idealized values of the states *A* and *B* or their probability matrices are. To begin the analysis, a Markov model is estimated with a given number of states and a probability transition matrix. The model is then optimized by determining the parameters that yield the maximum likelihood for the given trace. If the number of states has been determined, then the appropriate number of states can be entered and the rest of the parameters optimized. It is often the case that the number of states is also unknown, but can be determined by optimization of various Markov models with differing numbers of states and the results compared through a criterion such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) (Blanco and Walter 2010). Due to the iterative process required to find the most probable model parameters and sequence of states, HMMs are computationally expensive. The availability of a family of algorithms whose complexity scales only linearly with the length of the trajectory makes it possible to apply HMMs to time series that reach biologically relevant long time scales. The algorithms most commonly used include:

1. The forward algorithm, which calculates the conditional probability of being in a state *s* at time *t* given all of the observations up to that time. It also calculates the conditional probability of each observation given previous observations. Using these terms it calculates the probability of the entire data



**Hidden Markov Modeling in Single-Molecule Biophysics, Fig. 1** Simple hidden Markov model (a) An example of a hidden Markov model with two states with the transition probabilities presented above individual arrows to represent the likelihood of transitioning from one state to another or staying in the same state within the next time step. (b) The transition probability matrix of the Markov model from (a). (c) A simulated single-molecule FRET trajectory using the two-state model with states *A* (FRET = 0.8) and *B* (FRET = 0.2)



**Hidden Markov Modeling in Single-Molecule Biophysics, Fig. 2** Emission probabilities for the Markov process. The emission probability distributions of the two states *A* and *B* from the simulated single-molecule FRET trajectory in Fig. 1 are plotted. Here the emission probabilities are calculated by assuming Gaussian noise distributions around the discrete mean FRET values ( $A = 0.8$ ,  $B = 0.2$ ), modeling the shot noise of the signal collection instrumentation

2. The Viterbi algorithm, used when one needs to estimate a sequence of states from a sequence

of observations. It finds the most probable state sequence.

3. The Baum–Welch algorithm (forward-backward algorithm) calculates, given a given sequence of observations and an initial set of model parameters, in a single pass based on the forward algorithm a new set of parameters that has higher likelihood of being correct. Running many iterations of the Baum–Welch algorithm yields a sequence that approaches a local maximum of the likelihood.

Although HMMs provide an excellent tool for analyzing single-molecule data it must be noted that not all data fulfill the assumptions of a Markov process. For example, the changes of biomolecules can exhibit time-dependent transition probabilities due to molecular memory effects. In these cases HMMs can still approximate the data but care must be taken in the interpretation of the results. For an in-depth discussion of the mathematical foundations underlying HMM, we refer the reader to Rabiner (1989) and Fraser (2008).

## Applications of Hidden Markov Models in Single-Molecule Biophysics

The general framework of HMMs has been adapted to a variety of single-molecule techniques. Due to the differing character of the signals acquired from these techniques there is not a single HMM that can be utilized for the analysis of all types of data collection. However, HMMs have been modified and improved in the various fields to better model the data and noise of each particular field.

### Ion Channel Recordings

Among some of the earliest single-molecule experiments came from the electrophysiology field, where the action potential across a single ion channel can be recorded over time with the use of ► [patch clamp techniques](#). These techniques allow for the direct measurement of ionic currents through a single channel protein molecule. The amplitude of the signal collected describes the permeability of ions through the channel and the change in this permeability can be recorded in real time. HMMs can approximate the open and closed states of these pores effectively. This HMM implementation assumes the underlying signal is a Markov process whose noise is assumed to be Gaussian. QuB

(available at [http://www.qub.buffalo.edu/wiki/index.php/Main\\_Page](http://www.qub.buffalo.edu/wiki/index.php/Main_Page)) is a readily available software package routinely used for the analysis of ion channel recordings (Qin et al. 2000).

### Fluorescence Microscopy

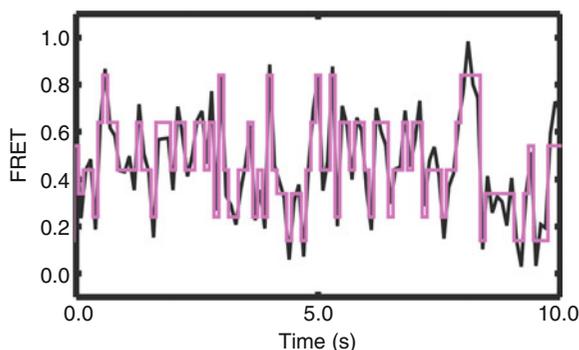
Single-molecule fluorescence microscopy has become one of the most popular single-molecule techniques due to the wide range of biomolecules that can be studied with this technique. HMMs have been adapted for various types of fluorescence microscopy to better model the different forms of intensity time traces collected.

#### Single-Molecule Fluorescence Resonance Energy Transfer (smFRET)

► [smFRET](#) can provide a real-time view of dynamics of biomolecules ranging from small catalytic RNAs (ribozymes) to large RNA-protein complexes such as the ribosome. The distance-dependent interaction between two fluorophores can report on intra- and intermolecular conformational changes. The application of HMM to the analysis of smFRET data became more accessible after the release of HaMMY, a user-friendly analysis software specifically designed for the analysis of smFRET data (McKinney et al. 2006). Previously, less sophisticated algorithms such as thresholding techniques were used that could often not handle complicated trajectories. The adaptation of HMM to smFRET has been advantageous for the field and the tool continues to be developed for the analysis of systems with more complicated behaviors. An example of the power of HMM analysis of smFRET trajectories is the trajectory of a single pre-mRNA molecule imaged under splicing conditions (Fig. 3) (Blanco and Walter 2010). The large number of states and rapid kinetics of transitions makes this type of trace difficult to analyze without the use of HMMs. In addition to HaMMY, programs such as QuB and vbFRET (Bronson et al. 2010) (available at <http://vbfret.sourceforge.net>) are available for smFRET analysis with advantages for more complex trajectories (Blanco and Walter 2010).

#### Switchable FRET

Switchable FRET is a combination of two techniques: smFRET and photoswitching, the reversible activation and deactivation of fluorophores commonly used in super-resolution imaging techniques such as



### Hidden Markov Modeling in Single-Molecule Biophysics,

**Fig. 3** A complex smFRET trajectory analyzed with HMM. HMM can be utilized to characterize the dynamics of single pre-mRNA molecules during splicing that exhibit rapid kinetics and a large number of states without the need for smoothing which would eliminate these small rapid conformational changes. In black is the raw FRET trajectory, and in magenta the five-state HMM idealized fit. In this particular case, the QuB ion channel analysis software was utilized for the HMM analysis

► **stochastic optical reconstruction microscopy (STORM)**. This technique utilizes multiple donor–acceptor fluorophore pairs to sequentially probe and obtain multiple distances within a single molecule. Traditional HMMs for smFRET cannot incorporate the stochastic photoswitching of the acceptor dye and therefore a linked hidden Markov model was developed where FRET and donor–acceptor stoichiometry are tracked (Uphoff et al. 2010). The linked HMM allows for the proper identification of states and determination of their transitions.

### Multi-fluorophore Bleaching

Some single-molecule experiments utilize multiple fluorophores to help determine the number of subunits assembled in a particular biological complex. The sudden drops in intensity when single fluorophores each undergo an irreversible ► **photobleaching** event can be used as a measure of the number of particles present. HMMs have been developed for the unbiased determination of the discrete steps in the fluorescence intensity traces. Up to 30 fluorophores can be reliably detected through the use of HMMs (Messina et al. 2006). The ability to reliably visualize and count the number of single molecules present has applications ranging from self-assembly of biomolecules to tracking the assembly of *trans* factors in multi-component systems.

### Molecular Motor Step Size

HMMs have been developed for the case of fluorescent measurements of molecular motors. An HMM variant, the variable-stepsize HMM (vsHMM) where the position of the motor is modeled as a large number of states, has been developed to more accurately track the movement of these motors as a result of their reaction cycle. This model differs in that it allows for an arbitrary distribution of step sizes that allows it to run as a robust algorithm with little user input. The algorithm has also been extended to the variable-stepsize integrating-detector HMM (VSI-HMM) which serves to better model the variation in signals during data acquisition such as random baseline fluctuations. Together, these HMMs have been utilized to characterize the movement of a myosin motor both in vitro and in vivo (Syed et al. 2010).

### Single-Particle Tracking

► **Single-particle tracking** can be used to extract modes of diffusion for a single molecule. In the case of biomolecules diffusing through a cell this approach can provide insight into the regions of localized activity, concentration gradients, or sites of modification. A single-particle track, with certain assumptions, can be modeled with a two-state HMM. The two-state HMM is optimized through the diffusion coefficients of the states and the rates of transition between them. It has been shown that this HMM is sufficient to extract multiple states of diffusion within a single trajectory in a practical manner (Das et al. 2009).

### Tethered Particle Microscopy (TPM)

► **TPM** experiments use light microscopy to measure the position of a bead tethered to a microscope slide via a polymer to infer the behavior of the polymer. An example is the use of DNA as the tether to measure DNA folding/unfolding dynamics and the effects of DNA-binding proteins on those dynamics. HMM analysis algorithms have been used to track the subtle changes in bead position and determine the relevant changes from those induced by Brownian motion by incorporating factors for the diffusive motion of the bead (Beausang et al. 2007). This approach has allowed for data analysis without the need for filtering. Another improvement to the HMM algorithms used for these experiments introduced factors to account for the nonlinear extension of DNA, allowing for a more accurate, quantitative assessment of the kinetics.

## Limitations of Hidden Markov Models

HMMs can be a powerful tool for the unbiased analysis of single-molecule data, but the resulting models need to be carefully inspected. As noted previously, not all biological processes studied under single-molecule conditions fulfill the Markov property, thus violating one of the assumptions of HMMs. This can affect the ability of the model to fully recapitulate the underlying behaviors. Additionally, as can be seen by the various adaptations of HMMs, it is necessary to define the right set of parameters to properly simulate the noise present in the system as well as the discrete number of states of interest. Finally, as with any fitting technique, when using HMM a rigorous test for model selection is required. This testing is often complicated by the lack of a clear and decisive way of selecting the proper number of states for the model. Although several methods have been presented for an unbiased approach at state selection, there is little consensus regarding which is best.

## Cross-References

- ▶ [Patch-Clamp Recording of Single Channel Activity: Acquisition and Analysis](#)
- ▶ [Photoactivated Localization Microscopy \(PALM\)](#)
- ▶ [Single Fluorophore Photobleaching](#)
- ▶ [Single-Molecule Fluorescence Resonance Energy Transfer](#)
- ▶ [Single-Molecule Spectroscopy](#)
- ▶ [Single-Particle Tracking](#)
- ▶ [Stochastic Optical Reconstruction Microscopy](#)
- ▶ [Tethered Particle Microscopy](#)

## References

- Beausang JF, et al. DNA looping kinetics analyzed using diffusive hidden Markov model. *Biophys J.* 2007;92(8):L64–6.
- Blanco M, Walter NG. Analysis of complex single molecule 419FRET time trajectories. *Method Enzymol.* 2010;472:153–78.
- Bronson JE, et al. Graphical models for inferring single molecule dynamics. *BMC Bioinform.* 2010;11(Suppl 8):S2.
- Das R, Cairo CW, Coombs D. A hidden Markov model for single particle tracks quantifies dynamic interactions between LFA-1 and the actin cytoskeleton. *PLoS Comput Biol.* 2009;5(11):e1000556.

- Fraser AM. Hidden Markov models and dynamical systems. Philadelphia: SIAM; 2008.
- McKinney SA, Joo C, Ha T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J.* 2006;91(5):1941–51.
- Messina TC, et al. Hidden Markov model analysis of multichromophore photobleaching. *J Phys Chem B.* 2006;110(33):16366–76.
- Qin F, Auerbach A, Sachs F. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys J.* 2000;79(4):1915–27.
- Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE.* 1989;77:257–86.
- Syed S, et al. Improved hidden Markov models for molecular motors, part 2: extensions and application to experimental data. *Biophys J.* 2010;99(11):3696–703.
- Uphoff S, et al. Monitoring multiple distances within a single molecule using switchable FRET. *Nat Meth.* 2010;7(10):831–6.

---

## Hierarchically Structured Lipid Systems

Chandrashekhar V. Kulkarni  
Biological and Soft Systems, Cavendish Laboratory,  
University of Cambridge, Cambridge, UK

## Synonyms

[Lipid phase equilibria](#); [Lipid self-organization](#); [Lipid superstructures](#); [Multiscale structural ordering of lipids](#)

## Definition

Amphiphilic lipids self-assemble into various thermodynamically stable nanostructures in the presence of water, which can be kinetically stabilized into hierarchically ordered lipid systems exhibiting multiple structural length scales.

## Introduction

Amphiphilic molecules such as lipids have the inherent tendency of self-assembling in an aqueous environment. The hydrophobic effect acts upon minimizing interactions between the water and hydrophobic parts of lipid molecules, thus having a prime contribution

toward their structural ordering. Lipid polymorphism spans thermodynamically stable and metastable forms of simple micelles as well as complex, yet well-ordered, one-, two- and three-dimensional assemblies (Seddon 1990; Hyde et al. 1997). They typically display intermediate behavior between that of solids and liquids, and also, their dimensions lie in the range of  $\sim 2.5$ – $25$  nm; therefore they are also referred as liquid crystalline (LC) nanostructures or lyotropic (phases induced by the addition of solvent) LC phases. Biological evidence of lipid self-assembly includes the plasma membrane, highly convoluted forms of ER (endoplasmic reticulum) membranes, inner nuclear membrane, mitochondrial inner membrane, chloroplast thylakoid membranes, tubules, and liposomes. However, most of the biological architectures are comprised of several molecular constituents rather than just one or a few simplistic lipid structural makeups. In other words, above biological entities would not be purely “self-assembled” and “thermodynamically stable” lipid systems, but may well be “kinetically stable” and/or “metastable” ensembles, which are affected by various physiological and physicochemical parameters. It is also possible to design many more complex morphologies by artificially modulating the basic lipid assembly at various hierarchical orders, and this is what is mainly focused on in the following sections.

## Lipid Phase Behavior

Amphiphilic lipid molecules usually have slightly more hydrophobic character compared to that of simple detergents and tend to form *type 2* or “inverse” aggregates where the interfacial curvature is toward the aqueous region. The lipid molecular chain splay increases with increasing temperature (pressure has the opposite effect), which makes the molecule more inverse conical, diverting away from cylindrical shape. This in turn increases the mean curvature (the average of the two principal curvatures) of a phase, resulting in phases with greater (negative) curvature. The formation of phases is also governed by other important physicochemical factors such as the type of lipid, degree of unsaturation, head group area, lipid chain length, type of solvent, composition of solvent, volume fraction of lipid in a corresponding mixture of lipids, addition of other additives (like buffers, salts, other amphiphilic molecules), and pH. Some of these

factors have been reviewed in the literature (Seddon and Templer 1995; Kulkarni et al. 2011b). The most simple lyotropic system is with two components, i.e., a pure lipid plus water, whose behavior is usually studied at different temperatures and represented on a temperature–composition phase diagram, as shown in Fig. 1. The monoelaidin–water system (Kulkarni 2011c), taken here as an example, shows three bicontinuous cubic phases,  $Ia3d$ ,  $Pn3m$ , and  $Im3m$ , whose architecture is based on mathematical minimal surfaces G, D, and P, respectively. These intricate cubic phase assemblies exhibit finely designed and ordered topology where an individual phase consists of a continuous lipid bilayer enclosing two interwoven but discrete aqueous channels. The numbers, dimensions, and angles at which the aqueous channels meet at single points in the unit cell of the cubic phases differ among the phases, which allows a certain degree of tunability with respect to their dimensions (Kulkarni et al. 2011b).

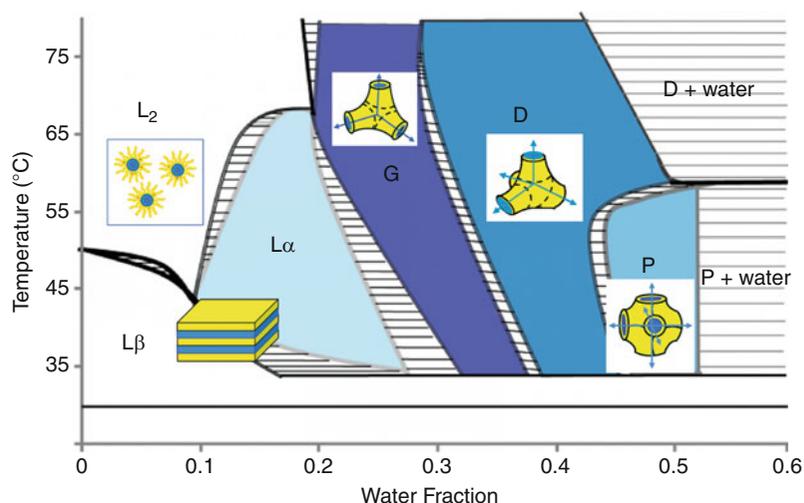
The lamellar phase is quite common in many lipid systems and is an indispensable structural constituent of biological membranes. This phase consists of one-dimensional stacking of planar lipid bilayers, whose behavior may differ in the degree of chain fluidity, head group ordering, tilting and interdigitation of monolayers, etc. The aforementioned parameters have vital significance in the functioning of biomembranes; another consequence is that they allow the observation of polymorphism in lamellar phases. Monoelaidin system (Fig. 1) shows two polymorphs: fluid ( $L_\alpha$ ) and gel ( $L_\beta$ ) lamellar phases; however, more are also possible at lower temperatures and hydration (not shown) (Kulkarni 2011c). The inverse hexagonal  $H_{II}$  phase, comprising of long water channels enclosed by lipidic cylinders, is also common among many lipid systems. Other LC nanostructures such as micellar cubic –  $Fd3m$ , disordered bicontinuous – sponge and disk-shaped bicelles are observed when additional chemical or biological component/s is/are added to the binary lipid–water system (Kulkarni et al. 2011b).

## Hierarchical Structural Ordering of Lipids

Lyotropic LC phases are known for their applications in various biotechnological fields, ranging from drug delivery to membrane protein crystallization (Kulkarni et al. 2011b). However, direct use of bulk LC phases is

### Hierarchically Structured Lipid Systems,

**Fig. 1** Binary phase diagram of the monoelaidin–water system displaying lamellar (fluid lamellar,  $L_\alpha$ , and lamellar gel,  $L_\beta$ ) phases; bicontinuous cubic phases –  $G$ ,  $D$ , and  $P$  phases – also indicated by corresponding crystallographic space groups as,  $Ia3d$ ,  $Pn3m$ , and  $Im3m$ ; and the inverse micellar fluid isotropic phase ( $L_2$ ). (Figure modified from Kulkarni et al. 2010b)

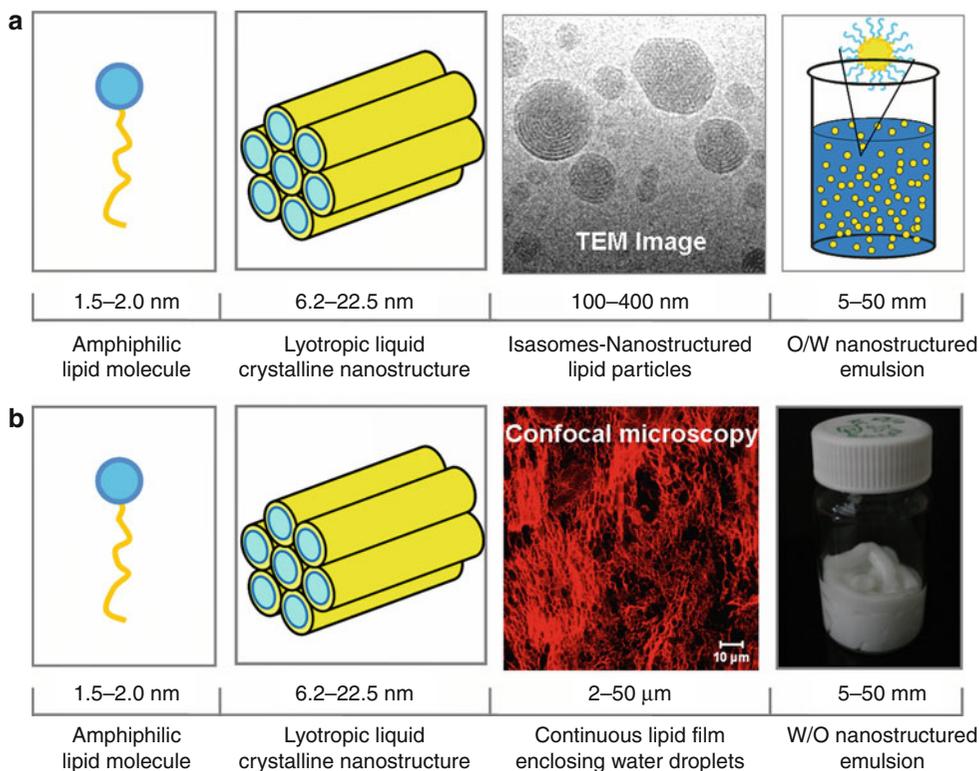


hampered due to their high viscosity and variable domain consistency. To overcome these problems, they are usually transformed or kinetically stabilized into other forms and/or hierarchical structures using some external stabilizer. Depending on the method of preparation and amount of dispersed phase, LC nanostructures reorganize into oil-in-water (O/W) or water-in-oil (W/O) emulsions (Fig. 2). The hydrophobic phase (lipid or lipid+oil) itself displays a wide range of nanostructures; when dispersed in water in the form of particles, they are generically known as “Isasomes” (Yaghmur et al. 2005), i.e., internally self-assembled *somes*, i.e., particles. More specifically, the particles with hexagonal and cubic internal structures are termed, respectively, “hexosomes” and “cubosomes,” the latter of which was the first such emulsion, demonstrated more than two decades ago (Larsson 1989). The reverse type of emulsion, where a hydrophobic phase (with an LC nanostructure) forms a continuous wall-like architecture enclosing water droplets (up to ~95% by wt), is called a W/O nanostructured emulsion, and more importantly, it does not require any external stabilizer (Kulkarni et al. 2010).

Recent research advances have extended the limits of internal nanostructures of Isasomes beyond the cubic ( $Pn3m$ ,  $Im3m$ , or  $Fd3m$ ) and hexagonal ( $H_2$ ) phases so as to include the microemulsion (Yaghmur et al. 2005), lamellar (Battaglia et al. 2007), and even sponge phases (Angelov et al. 2011); the resulting Isasomes are called emulsified microemulsions (EME), lamellarsomes, and spongisomes (name given here), respectively. Sometimes, the vesicles and

liposomes were seen to coexist and/or enclose the Isasomes. Here, note that the lamellarsomes and liposomes are not one and the same, as the interior of liposomes is usually filled with water whereas lamellar phase is enclosed by the lamellarsomes. Being similar in size as liposomes, the Isasomes are comprised of LC phases thereby providing considerably higher volume fraction of hydrophobic environment. The latter is already equipped with well-defined nanostructural units facilitating loading of poorly water soluble and amphiphilic functional molecules. The same dispersion allows loading of highly water-soluble molecules in the continuous aqueous medium. Increased diffusive path length provided by the internal nanostructures of Isasomes thus demonstrate an enhanced capacity for sustained release of the loaded components, which can be further improved by increasing the viscosity by adding hydrogelling agent in the aqueous region (Kulkarni and Glatter 2012). In this manner, the Isasomes provide certain advantages over vesicle/liposomal structures and conventional (nonstructured oil phase) emulsions; the latter are most common for detergent-based amphiphiles.

There are several other types of hierarchically ordered forms of lipids being developed and investigated by researchers for various applications, including poly-high internal phase emulsions (poly-HIPE); bicontinuous interfacially jammed emulsion gels (bijels); polymerized LC phases; lipid nanoparticles (LNP), i.e., liposomes loaded with functional molecules; lipid nanocapsules; etc. Self-assembled monolayers (SAM), Langmuir-Blodgett films, and



**Hierarchically Structured Lipid Systems, Fig. 2** Hierarchical ordering of lipids in (a) O/W and (b) W/O nanostructured emulsions indicated with approximate length scales. The LC nanostructure shown here is “hexagonal,” but it

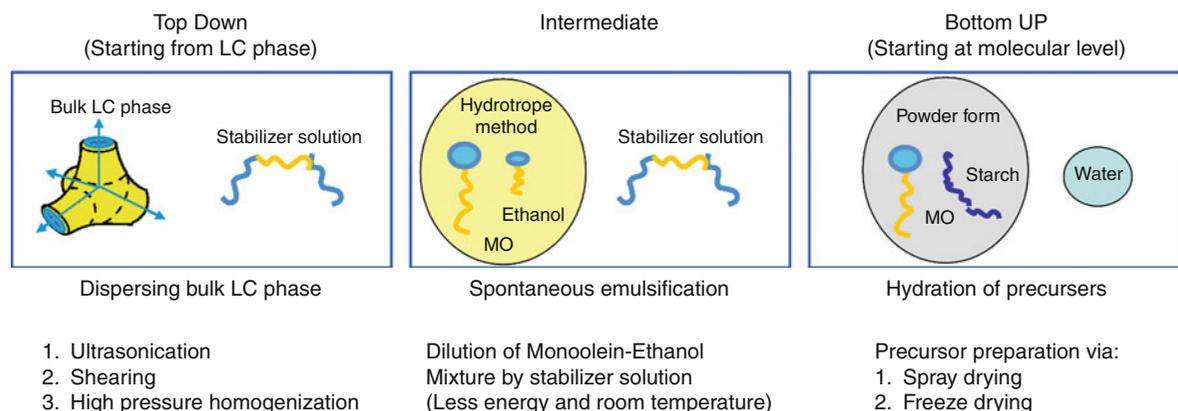
► **supported lipid bilayers** (SLB) are some simple, commonly employed model lipid assemblies. Solid lipid nanoparticles (SLN) developed in the early 1990s, and their successive variants, nanostructured lipid carriers (NLC), prepared by the high-pressure homogenization technique, are also under investigation, particularly for drug delivery applications (Muller et al. 2002). Nonetheless, structured emulsions, formed primarily from simple monoglycerides, have emerged to be quite successful lipid-based materials that are utilized for various commercial applications (Yaghmur and Glatter 2009).

### Isasomes: Nanostructured Lipid Particles

Isasomes are kinetically stabilized O/W emulsions comprising internally nanostructured submicron-sized (typically 200–400 nm) lipid particles dispersed in excess water (typically 80–95%) as a continuous

medium. Some groups define the Isasomes differently, e.g., nonlamellar lipid nanoparticles (Barauskas et al. 2005), self-assembled lipid nanocarriers, nanostructured lipid particles, or colloidal lipid nanostructures. The primary components of Isasomes are lipid, stabilizer, and water. Lipids that form cubic and/or hexagonal phases at room temperature are usually required for Isasome preparation; however, monoolein (MO) and monolinolein (ML) seem to be popular in their pure or commercial derivatives (Dimodan-U/J<sup>®</sup> (DU) and glycerol monooleate (GMO), respectively), while phytantriol (PT) and monoelaidin (ME) have also been used in recent years. Boyd and coworkers (2009) have compiled a list of lipids and lipid mixtures forming nonlamellar phases and their potential utility for the preparation of hierarchical structures. The literature on various lipid phase diagrams is also indexed by Koynova and coworkers (Koynova and Caffrey 2002). A range of polymers, surfactants, food hydrocolloids, and nanoparticles are utilized as emulsion

can be easily tuned into other forms such as cubic, fluid isotropic, and sponge. Parts of the figure are obtained from (Yaghmur et al. 2005) and (Kulkarni et al. 2010a)



**Hierarchically Structured Lipid Systems, Fig. 3** Different approaches for the preparation of Isasomes, especially cubosomes

stabilizers, which can be broadly categorized into the following categories:

### 1. *Polymers and Polymeric Surfactants*

The amphiphilic triblock copolymer, Pluronic<sup>®</sup> F127 (PEO<sub>99</sub>-PPO<sub>67</sub>-PEO<sub>99</sub>), has been most widely used for structured emulsions such as Isasomes; nevertheless, polysaccharide and polyethylene glycol (PEG)-based stabilizers, e.g., Tween, poloxamers, and PEGylated lipids, are also used, with varying levels of success (Yaghmur and Glatter 2009).

### 2. *Solid Particles (Pickering Emulsion Stabilizers)*

Surfactant-based stabilizers are prone to affect the internal nanostructure of Isasomes; moreover, they can penetrate inside the lipid particles to disrupt the nanostructured interior. Instead of surfactants, some solid particles are exploited to obtain “surfactant-free emulsions” which are also known as Ramsden-Pickering or Pickering emulsions. An assortment of solid particles utilized for Pickering emulsions include latex particles, metal oxides, carbon and sulfate particles, polymeric rods and grafted polymer brushes, silica nano- or microcolloids, and clay platelets (Kulkarni and Glatter 2012). The latter two are best suited for stabilizing Isasomes because they allow the emulsion stability to be controlled by changing one or more of the following parameters: charge, pH, and temperature; although it is also possible to alter conventional parameters, such as concentration and particle size and shape. The overall emulsion stability varies inversely with the particle size, as the smaller particles pack more efficiently. The particle size ranges from nanometer to micrometer, stabilizing up to millimeter-sized

droplets, which is rather unusual for surfactant-based systems (Kulkarni and Glatter 2012).

### 3. *Biomolecules*

Emulsions are also known to be stabilized by proteins, such as casein, albumin, and lactoglobulin (Dickinson 2009; Yaghmur and Glatter 2009). Protein-stabilized food emulsions are sometimes called Pickering emulsions, even though most proteins used for this purpose are also amphiphilic. Other biomolecules including bile salts, ► cellulose and derivatives, modified forms of starch, etc., are also being explored for stabilizing structured emulsions (Kulkarni and Glatter 2012).

Various emulsifying systems and stabilizers including lipids, surfactants, proteins, hydrocolloids, and nanoparticles and the underlying principles have been recently reviewed (Hunter et al. 2008; Dickinson 2009; Yaghmur and Glatter 2009).

A variety of methods have been developed for the preparation of Isasomes, requiring high or low energy input for their crude or fine production. There are two major approaches: one starting from the ready-made LC phases (“top down”), whereas the other starts from the molecular level (“bottom up”), as shown schematically in Fig. 3. In the first, the LC phases are dispersed via high energy techniques such as ultrasonication, high-pressure homogenization, or shearing in a couette-like device, whereas the second utilizes spray or freeze-drying to make dry precursors of lipid and starch or dextran to form Isasomes, simply by subsequent hydration. The latter is more useful for large-scale production, mainly for the food industry.

An intermediate “hydrotrope”-based approach was also reported in the literature, by which Isasomes are

**Hierarchically Structured Lipid Systems, Table 1** Physicochemical parameters for modulation of the structure and properties of Isasome systems. Compilation is based on the literature to date, more specifically from the following references (Barauskas et al. 2005; Boyd et al. 2009; Yaghmur and Glatter 2009; Kulkarni et al. 2011a, b; Kulkarni and Glatter 2012)

No.	Parameter	Change in the Isasome system
1	Concentration of dispersed phase or continuous phase (usually water)	Viscosity, amount of Isasomes, sometimes the size of Isasomes
2	Composition of hydrophobic phase (e.g., lipid to oil ratio)	Internal nanostructure type and lattice parameters of LC phase, size of Isasomes
3	Type of hydrophobic additive (oil type – tetradecane or limonene, other lipid, glycerol, biomolecules, etc.)	Internal nanostructure type and lattice parameters of LC phase, stability of Isasomes
4	Type of primary lipid that forms (internal) LC nanostructure	Internal nanostructure type and/or lattice parameters of LC phase
5	Concentration of stabilizer	Internal nanostructure type and lattice parameters of LC phase, size of Isasomes
6	Type of stabilizer (polymeric surfactants, solid nanoparticles, clay, proteins, etc.)	Internal nanostructure type and lattice parameters of LC phase, size of Isasomes, mechanism of stability of Isasomes
7	Hydrogel formation in continuous phase	Viscosity, (arresting) particle dynamics, sometimes internal nanostructure, stability of Isasomes
8	Charged components	Type of internal nanostructure, stabilization/destabilization of Isasomes
9	pH	Type of internal nanostructure, stability of Isasomes
10	Temperature	Internal nanostructure type and lattice parameters of LC phase, stability of Isasomes, size of Isasomes
11	Pressure	Internal nanostructure type and lattice parameters of LC phase, stability of Isasomes, size of Isasomes
12	Method of preparation (ultrasonication, shearing–varying shear rate)	Stability of Isasomes, size of Isasomes, mechanism of stability of Isasomes
13	Intermixing two differently nanostructured Isasomes	Type of internal nanostructure, kinetics of LC phase formation/disappearance, size of Isasomes

prepared via spontaneous emulsification upon blending of two solutions, i.e., lipid plus hydrotrope and stabilizer solution. This method requires less energy compared to others; moreover, it allows Isasome production at room temperature. Some other methods with varying degrees of success include mechanical stirring during the hydration of lipid plus stabilizer film and microfluidization-based methods as revealed by Yaghmur and coworkers (Yaghmur and Glatter 2009) and Boyd and coworkers (2009).

For laboratory scale Isasome production, the ultrasonication method is very convenient, primarily for lower concentrations of dispersed phase (~5–30% Isasomes in the total dispersion); however, for concentrations above ~30%, the shearing method works better, by which one can produce highly concentrated Isasome dispersions (up to ca. 70% of dispersed phase) (Kulkarni and Glatter 2012). The shearing technique can be easily scaled up for the continuous production of Isasome-like systems. Recently, it was also shown that by simply

optimizing shear conditions, one can create W/O nanostructured emulsions (Fig. 2b) without requiring any external stabilizer (Kulkarni et al. 2010).

Isasomes allow a great degree of tunability at every level of structural hierarchy, ranging from the lowest, molecular level, through an intermediate, nanostructural level, to the particle size, at the highest level. Various physical parameters, chemical additives, and/or preparation methods can induce such morphological and behavioral tuning (Kulkarni and Glatter 2012), some of which are listed in Table 1.

Hierarchical systems like Isasomes have been characterized using a range of different techniques; some of them are compiled in Table 2. Small-angle X-ray scattering, cryo-transmission electron microscopy (cryo-TEM), and dynamic light scattering were found to be most accepted among the others, which provide morphological information whereas spectroscopic techniques reveal the physicochemical modulation of Isasome systems.

**Hierarchically Structured Lipid Systems, Table 2** Various instrumental techniques used for characterization of lipid-based systems

No.	Characterization technique	Information (most important/feasible) gained
1	Small-angle X-ray scattering (SAXS)	Internal nanostructure type, coexistence and lattice parameters of LC phase, morphology of Isasomes, time-resolved studies – kinetic events, e.g., phase transitions or dynamic changes in particle morphology, after calibration for intensity – amount of dispersed phase
2	Cryo-TEM	Internal nanostructure type, coexistence and lattice parameters of LC phase, size and morphology of Isasomes
3	Dynamic light scattering (DLS)	Isasome size, size distribution, kinetic events, e.g., stability or aggregation phenomena
4	Optical light microscopy (polarized and cross polarized)	LC nanostructure of bulk systems, especially to distinguish isotropic and anisotropic phases, texture studies, e.g., between lamellar and hexagonal phases, kinetics events such as formation and phase transitions to some extent
5	Cryo-field emission scanning electron microscopy (cryo-FESEM)	3-D information on the particle structure
6	Environmental scanning electron microscopy (SEM)	Particle morphology under hydrated conditions
7	Atomic force microscopy (AFM)	Particle morphology
8	Confocal laser scanning microscopy (CLSM)	Particle morphology
9	Small-angle neutron scattering (SANS)	Information about internal nanostructure, morphology of Isasomes, some time-resolved studies
10	Spectroscopic techniques	Physicochemical modulation of Isasome systems (e.g., effect of additives, existing components and physical parameters), diffusion studies
11	Rheological techniques	Viscosity, viscoelasticity, texture/consistency

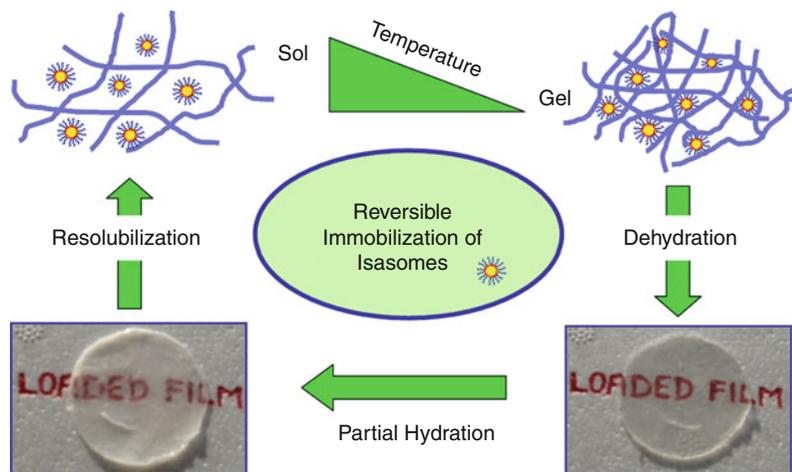
The bulk LC phases and their hierarchical structures have similarities at the nanostructural level; however, the latter are more advantageous from an applications perspective. Isasomes are easy to handle, being relatively low viscosity materials, using conventional industrial designs customized for liquid handling. Use of a dispersed form of LC phases drops the overall cost, by utilizing less bulk material. The bulk LC phase is usually oil continuous; in contrast, Isasomes are internally oil continuous but externally water continuous, typically containing a large excess (80–95%) of water. This morphological and chemical combination facilitates their loading with hydrophobic and hydrophilic as well as amphiphilic functional molecules. These nanostructured assemblies are well-known vehicles for drug delivery as well as for the controlled release and uptake of active molecules. They find further applications in the fields of pharmaceuticals, foods, cosmetics, and bioadhesives given the possibility of their customization from food grade or biocompatible components (Yaghmur and Glatter 2009).

### Further Advances in Lipid Structural Hierarchy

Isasomes can be embedded in hydrogels formed via addition of hydrogelling agents in the dispersion medium. The fluidity of the medium, and hence the particle dynamics, can be altered via temperature-induced solgel transitions or by altering the concentration of hydrogelling agent (Kulkarni and Glatter 2012). Dried thin films made from these hydrogels are capable of immobilizing the Isasomes and also drying them into fluid isotropic ( $L_2$ ) nanostructures. On the other hand, the Isasomes can be regenerated by solubilizing such films; more interestingly, the Isasome nanostructure and size remains practically the same, before and after loading/unloading in the films. This opens up new types of applications where Isasomes need to be stored for a long time or transported somewhere, along with functional molecules. Isasome-loaded gels and films can potentially be used as a fabrication material in the pharmaceuticals and medicinal fields, where sustained release of components is required.

### Hierarchically Structured Lipid Systems,

**Fig. 4** Thermoreversible hydrogel can be used to entrap the Isasomes, which can be further immobilized into dried films. The Isasomes can be restored by rehydration of the loaded films. In this manner, it is possible to store and transport the Isasome systems before their regeneration for specific application (Figure adopted from Kulkarni et al. 2011a)



Similar to the Isasome systems, the stability and viscosity of W/O-nanostructured emulsions can also be tuned by forming the hydrogel in its aqueous reservoirs/droplets (Fig. 4).

Loading of lipid structures with proteins and other biomolecules is also known, e.g., proteoliposomes, proteocubosomes, or lipid bilayer nanodisks (proteins with bicelles), which may sometimes lead to nanostructural transitions (Kulkarni et al. 2011b). Isasomes with dissimilar nanostructures, charges, or functional molecules can be intermixed or transformed into fascinating layer-by-layer assemblies which have great potential in diverse areas. There are enormous opportunities to further develop, explore, and utilize new and novel types of lipid-based hierarchical structures.

### Summary

Lipids self-assemble into thermodynamically stable liquid crystalline phases of various sizes, shapes, and properties, and some of them also mimic biological structures. However, in order to ease the exploration and increase their applicability, it is customary to organize these lipid phases into higher-order architectures exhibiting multiple length scales. Accordingly, the O/W emulsions in the particulate form called “Isasomes” and W/O type of emulsions in the continuous-film form called “nanostructured emulsions” appear to be very successful options. The essential components for the preparation of Isasomes,

technical methodology of their preparation, characterization techniques, and their tunability have been described in the current entry. Some more hierarchical level structures designed from lipidic systems are reported in the final section. Research into the exploration and applications of lipid-based systems is likely to continue to develop as a lively biotechnological area.

### Cross-References

- ▶ [Atomic Force Microscopy](#)
- ▶ [Atomic Force Microscopy \(AFM\) for Topography and Recognition Imaging at Single Molecule Level](#)
- ▶ [Cellulose](#)
- ▶ [Dynamic Light Scattering](#)
- ▶ [Electron Microscopy](#)
- ▶ [Functional Roles of Lipids in Membranes](#)
- ▶ [Glycerolipids: Chemistry](#)
- ▶ [Lipid Mesophases for Crystallizing Membrane Proteins](#)
- ▶ [Lipid Organization, Aggregation, and Self-assembly](#)
- ▶ [Membrane Fluidity](#)
- ▶ [Membrane Proteins: Structure and Organization](#)
- ▶ [Neutron Scattering of Membranes](#)
- ▶ [Phase Transitions and Phase Behavior of Lipids](#)
- ▶ [Phospholipid Bilayer Nanodiscs: Application to Single-Molecule Measurements](#)
- ▶ [Supported Lipid Bilayers](#)
- ▶ [Vibrational Spectroscopy, A Short History of](#)
- ▶ [X-Ray Scattering of Lipid Membranes](#)

## References

- Angelov B, Angelova A, et al. SAXS investigation of a cubic to a sponge (L3) phase transition in self-assembled lipid nanocarriers. *Phys Chem Chem Phys*. 2011;13:3073–81.
- Barauskas J, Johnsson M, et al. Self-assembled lipid superstructures: beyond vesicles and liposomes. *Nano Lett*. 2005;5(8):1615–19.
- Battaglia G, Tomas S, et al. Lamellarsomes: metastable polymeric multilamellar aggregates. *Soft Matter*. 2007;3(4):470–5.
- Boyd BJ, Dong Y-D, et al. Nonlamellar liquid crystalline nanostructured particles: advances in materials and structure determination. *J Liposome Res*. 2009;19(1):12–28.
- Dickinson E. Hydrocolloids as emulsifiers and emulsion stabilizers. *Food Hydrocoll*. 2009;23(6):1473–82.
- Hunter TN, Pugh RJ, et al. The role of particles in stabilising foams and emulsions. *Adv Colloid Interface Sci*. 2008;137(2):57–81.
- Hyde ST, Andersson S, et al. The language of shape: the role of curvature in condensed matter: physics, chemistry and biology. Amsterdam: Elsevier; 1997.
- Koynova R, Caffrey M. An index of lipid phase diagrams. *Chem Phys Lipids*. 2002;115(1–2):107–219.
- Kulkarni CV, Mezzenga R, et al. Water-in-oil nanostructured emulsions: towards the structural hierarchy of liquid crystalline materials. *Soft Matter*. 2010a;6(21):5615–24.
- Kulkarni CV, Tang TY, et al. Engineering bicontinuous cubic structures at the nanoscale—the role of chain splay. *Soft Matter*. 2010b;6(21):3191–94.
- Kulkarni CV, Tomšič M, et al. Immobilization of nanostructured lipid particles in polysaccharide films. *Langmuir*. 2011a;27(15):9541–50.
- Kulkarni CV, Wachter W, et al. Monoolein: a magic lipid? *Phys Chem Chem Phys (PCCP)*. 2011b;13:3004–21.
- Kulkarni CV. Nanostructural studies on monoolein-water systems at low temperatures. *Langmuir*. 2011c;27(19):11790–800.
- Kulkarni CV, Glatter O. Hierarchically organized systems based on liquid crystalline phases. In: Garti N, Mezzenga R, editors. *Self-Assembled Supramolecular Architectures: Lyotropic Liquid Crystals*, John Wiley & Sons, NJ, USA; 2012, Ch. 6.
- Larsson K. Cubic lipid-water phases: structures and biomembrane aspects. *J Phys Chem*. 1989;93(21):7304–14.
- Muller RH, Radtke M, et al. Nanostructured lipid matrices for improved microencapsulation of drugs. *Int J Pharm*. 2002;242(1–2):121–8.
- Seddon JM. Structure of the inverted hexagonal (HII) phase, and non-lamellar phase transitions of lipids. *Biochim Biophys Acta (BBA) Rev Biomembr*. 1990;1031(1):1–69.
- Seddon JM, Templer RH. Polymorphism of lipid-water systems. In: Lipowsky R, Sackmann E, editors. *Handbook of biological physics*, vol. 1. Amsterdam: Elsevier; 1995. p. 97–160.
- Yagmur A, Glatter O. Characterization and potential applications of nanostructured aqueous dispersions. *Adv Colloid Interface Sci*. 2009;147–148:333–42.
- Yagmur A, de Campo L, et al. Emulsified microemulsions and oil-containing liquid crystalline phases. *Langmuir*. 2005;21(2):569–77.

## High-Pressure NMR

Kazuyuki Akasaka

Institute of Advanced Technology, High Pressure Protein Research Center, Kinki University, Kinokawa, Wakayama Prefecture, Japan

## Synonyms

[Variable-pressure NMR](#)

## Definition

Pressure is a thermodynamic variable that acts on volume of the system without significantly perturbing its thermal energy, and therefore is more directly coupled to the “structure” of a biomacromolecule in solution than temperature. The volume of a biomacromolecule (meaning partial volume in solution), notably of a water-soluble protein, is intimately and dynamically coupled to its hydration, which brings about a dynamic variability in conformation under closely physiological conditions. Thus high-pressure (or variable-pressure) NMR, combining pressure perturbation with the high-resolution one- and multidimensional NMR spectroscopy, enables exploration of details of protein and other macromolecular conformational variation in solution in a much wider conformational space (from within the folded N to the fully unfolded U) than explored by conventional NMR spectroscopy. For naturally occurring proteins, this possibility is substantiated by the two practical rules: (1) (marginal balance) A variety of conformers exist between N and U in a narrow range of free energy (small  $\Delta G^0$ ) often within several kcal/mol and (2) (volume theorem) Under normal conditions, the partial molar volume of a folded protein decreases ( $\Delta V < 0$ ) with the loss of the conformational order, that is, with the progress of hydration and unfolding (Akasaka 2006).

Technically, high-pressure NMR originated in two different approaches: one pressurizing the NMR detection coil itself in an autoclave and the other pressurizing only the sample solution confined in a pressure-resistive cell. For biological macromolecular studies, the latter method combined with

a commercial NMR spectrometer is chosen for its high magnetic field homogeneity and versatility in pulse sequence. Typically, a range of pressure of 0.1~ a few hundred MPa has been used, the highest pressure being limited so far by the pressure-resistivity of the sample cell at about 400 MPa (Akasaka and Yamada 2001). Various materials have been tested for the cell, including plastics, glass, quartz, sapphire, and zirconia. Handmade quartz cells (<1 mm in inner diameter) have been used successfully, but because of its difficulty in manufacturing, handling, and limited sample space, ceramic cells (1–3 mm in inner diameter) made of zirconia are beginning to be employed.

Upon applying pressure  $p$  to protein solutions at constant temperature, all NMR parameters including chemical shift, signal intensity, NOE, line width, spin relaxation, and even J-coupling are subject to change, as the equilibrium population of a lower volume conformer ( $\Delta V < 0$ ) increases by  $\exp(-p\Delta V/RT)$ . Chemical shifts are particularly sensitive to pressure, representing the intrinsic fluctuations of the folded conformer N, including transitions to alternatively folded ones N'. With increasing pressure, signal intensities may start to decrease for certain residues, meaning local unfolding into an intermediate conformer I, and, at still higher pressure, even total unfolding into conformer U. All the above changes are essentially reversible with pressure and reflect a small-to-large amplitude fluctuations occurring in a time range of ps ~ ks, difficult to be detected unless "amplified" by pressure. High-pressure NMR shares a common target with H/D exchange and relaxation dispersion NMR methods in "high-energy" conformers, but is unique in that it can directly detect signals from "high-energy" conformers, allowing detailed analysis of their structures and stability.

## Cross-References

- ▶ [H/D Exchange](#)
- ▶ [J Coupling](#)
- ▶ [Nuclear Overhauser Effect](#)
- ▶ [Relaxation Dispersion](#)

## References

Akasaka K. Probing conformational fluctuation of proteins by pressure perturbation. *Chem Rev.* 108;2006:1814–35.

Thematic Issue "Protein Dynamics and Folding" (Guest Editor: A. Joshua Wand).

Akasaka K, Yamada H. On-line cell high pressure nuclear magnetic resonance technique: application to protein studies. In: James TL, et al., editors. *Nuclear magnetic resonance of biological macromolecules part A, Methods in enzymology*, vol. 338. Academic; 2001. pp. 134–58.

---

## High-Speed Atomic Force Microscopy (AFM)

Toshio Ando

Department of Physics and Bio-AFM Frontier Research Center, Kanazawa University, Kanazawa, Ishikawa, Japan

### Synonyms

[Fast scanning force microscopy](#)

### Definition

High-speed AFM (HS-AFM) is a type of atomic force microscopy (AFM) that, unlike conventional AFM, can take an image very quickly and with relatively low imaging force, therefore allowing for visualizing the dynamic behavior of biomolecules.

### Introduction

AFM can take high-resolution images of objects in various environments. In particular, its in-liquid imaging capability is useful for biological research. However, the imaging rate of conventional AFM is too low to capture dynamic processes of biomolecules (DNA, proteins). HS-AFM has been developed to overcome this problem by optimizing all devices contained in the instrument for fast scanning and by newly developing fast and precise control techniques (Ando et al. 2001, 2008). The imaging rate of the current HS-AFM has reached 15–25 frames per second (fps) for a scan range of  $\sim 240 \times 240$  nm<sup>2</sup> with 100 scan lines. By further reducing the scan range, the imaging rate can exceed 50 fps. Importantly,

high-speed imaging with small cantilevers applies less force during its scans than conventional AFM and therefore, is less invasive. Thus, one can directly image functioning biomolecules in action at high spatiotemporal resolution, without disturbing their biological functions (Kodera et al. 2010; Shibata et al. 2010; Uchihashi et al. 2011). The resulting molecular movies provide much information on both structure and dynamics of a functioning molecule, which significantly facilitates a comprehensive understanding of the functional mechanism.

### Imaging Rate and Feedback Bandwidth

When a sample is scanned over an area of  $W \times W$  with scan velocity  $V_s$  in the x-direction and  $N$  scan lines, the image acquisition time  $T$  is given by  $T = 2WN/V_s$ . Assuming that the sample surface is characterized with a single spatial frequency  $1/\lambda$ , then the feedback scan is executed in the z-direction with frequency  $f = V_s/\lambda$  to trace the sample surface. The feedback bandwidth  $f_B$  of the microscope should be equal to or higher than  $f$ , and thus, we obtain the relationship  $T \geq 2WN/(\lambda f_B)$ . In the current implementation of HS-AFM,  $f_B \sim 110$  kHz, achieving the aforementioned imaging rate.

### Devices for HS-AFM

The tapping mode HS-AFM instrument developed at the Ando laboratory (Fig. 1) is commercially available worldwide from the Research Institute of Biomolecule Metrology, Co. Ltd. (Tshukuba, Japan).

*Cantilevers and related devices:* The size of rectangular cantilevers for high-speed imaging is small: 6–10- $\mu\text{m}$  long, 2- $\mu\text{m}$  wide, and 90-nm thick. They have resonant frequencies  $f_c = 2\text{--}3.5$  MHz in air and  $f_c = 0.6\text{--}1.2$  MHz in water, a spring constant  $k_c = 0.1\text{--}0.2$  N/m, and a quality factor  $Q_c = 2\text{--}3$  in water. Small cantilevers (BL-AC10DS-A2, Olympus:  $f_c$  in air 1.5 MHz,  $f_c$  in water 600 kHz,  $k_c \sim 0.1$  N/m) are commercially available (Atomic Force F&E GmbH, Mannheim, Germany; Asylum Research, Santa Barbara, USA; Bruker, Calle Tecate, USA). Alternatively, small cantilevers with a sharp tip made by electron beam deposition are also available. Although not yet commercialized, small cantilevers

will soon be available also from NanoWorld AG (Neuchâtel, Switzerland). The deflection of a small cantilever is measured by an optical beam deflection (OBD) detector equipped with a long-working distance objective lens, a quarter-wavelength plate, and a polarization beam splitter (Fig. 1) (Ando et al. 2001). The small cantilevers grant 10–20-times higher detection sensitivity than conventional cantilevers. The oscillation amplitude is measured by a peak-hold method (at every half cycle of oscillation) (Ando et al. 2001) or a Fourier method (at every full cycle of oscillation) (Ando et al. 2008).

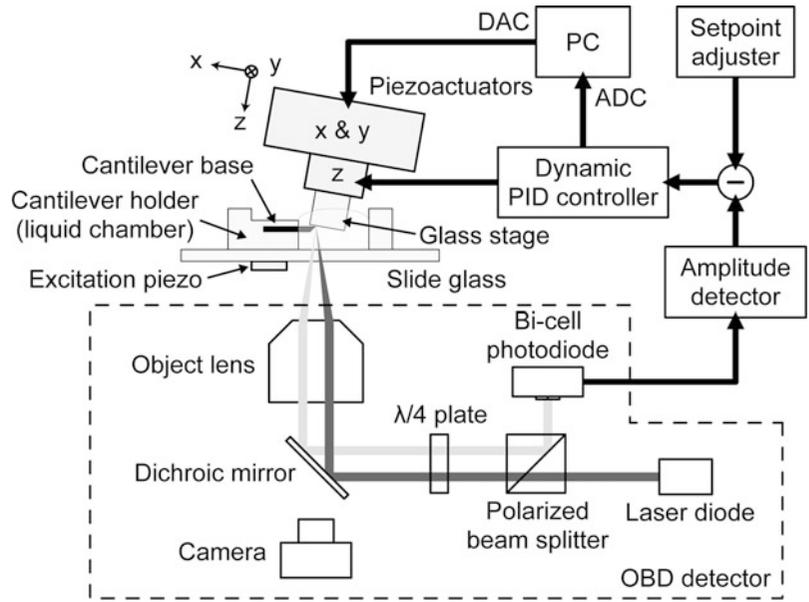
*Scanner:* The maximum displacements of the high-speed scanner are 1–2  $\mu\text{m}$  in the x-direction,  $\sim 3$   $\mu\text{m}$  in the y-direction, and  $\sim 1$   $\mu\text{m}$  in the z-direction. The displacements are large enough for imaging biomolecules. The z-scanner can move at 200 kHz without a practical delay, while the x-scanner can move at  $>10$  kHz in a triangle waveform. The generation of unwanted vibrations is effectively damped by analogue feedback Q-control of the z-scanner and by digital feed forward Q-control (inverse compensation) of the x-scanner (Ando et al. 2008).

*Dynamic PID controller:* This controller allows one to use an amplitude set point  $A_s$  (cantilever oscillation amplitude to be maintained constant during imaging) very close to the free oscillation amplitude  $A_0$  (i.e.,  $A_s \sim 0.9 \times A_0$ ) (Kodera et al. 2006). This large set point minimizes the tapping force exerted by the oscillating tip on the sample. In conventional PID controllers, this large set point often results in complete detachment of the cantilever tip from the sample surface at steep downhill regions of the sample (i.e., “parachuting”), resulting both in an inability to image until parachuting ceases and in a significant reduction of the feedback bandwidth.

*Drift compensator:* The cantilever is excited by a piezoactuator placed at (or near) the cantilever holder (Fig. 1). To minimize the tapping force,  $A_0 \sim 1\text{--}2$  nm is used. Therefore, the difference between  $A_0$  and  $A_s$  is very small ( $\sim 0.1\text{--}0.2$  nm). However,  $A_0$  easily changes with time by  $>0.2$  nm due to drift in the excitation power, and therefore, stable imaging is difficult under this condition. This problem was solved by maintaining the constant average amplitude of a second harmonic oscillation of the cantilever (Kodera et al. 2006). This drift compensator works effectively, and thus enables stable low-invasive and high-speed imaging when operating together with the dynamic

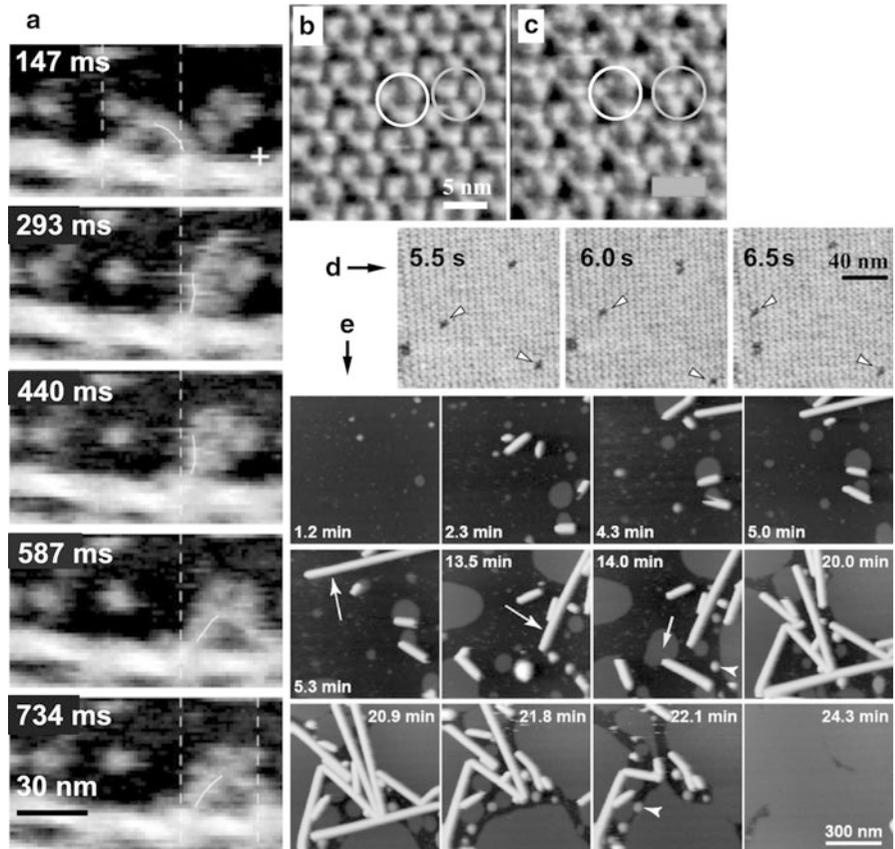
**High-Speed Atomic Force Microscopy (AFM),**

**Fig. 1** Schematic of an HS-AFM instrument. Unlike conventional AFM, the cantilever and sample are arranged upside-down so that the cantilever tip points upward to the sample



**High-Speed Atomic Force Microscopy (AFM),**

**Fig. 2** AFM images of biomolecules captured by HS-AFM (a) Myosin V (tail-truncated myosin V) walking along an actin filament (Modified from Fig. 1d of Kodera et al. 2010). (b, c) bR under *dark* (b) and *light* illumination (c) (Modified from Fig. 1a of Shibata et al. 2010). The *white circles* indicate trimers and the *gray circles* indicate triads of nearest neighbor bR molecules each belonging to adjacent trimers. (d) Anisotropic diffusion of single point defects in 2D crystal of streptavidin; (e) Process of planar lipid bilayer formation after placing liposomes on a mica surface (Figure 8 of Giocondi et al. 2010)



PID controller, even under the conditions of  $A_0 \sim 1$  nm and  $A_s = 0.9\text{--}0.95 \times A_0$ .

## HS-AFM Imaging of Biomolecules

*Substrate surface:* To observe dynamically acting protein molecules under physiological conditions, the substrate surface must have an appropriate affinity for the molecules. To observe dynamic interactions between different proteins, selective protein attachment to a surface is required. Thus, we have to prepare an appropriate substrate surface depending on the dynamic biomolecular event to be visualized. As substrate, bare mica surfaces, planar lipid bilayer surfaces, and the surfaces of two-dimensional crystals of streptavidin formed on a biotin-containing lipid bilayer have been used for dynamic AFM imaging (Yamamoto et al. 2010).

*Examples of dynamic imaging:* Myosin V walking along an actin filament has been visualized (Fig. 2a) (Kodera et al. 2010). For this observation, partially biotinylated actin filaments were immobilized via streptavidin on a biotin-containing lipid bilayer surface. To facilitate weak sideways adsorption of myosin V on the surface, a small fraction of a positively charged lipid was included in the bilayer. The observed velocity is similar to that measured by fluorescence microscopy, indicating very small effects of the tip-sample and surface-sample interactions on motor activity. The high-resolution movies not only provide corroborative evidence for previously proposed or demonstrated molecular behaviors, including a lever-arm swing, but also reveal more detailed behaviors of the molecules. Likewise, the photo-activated dynamic structural change of bacteriorhodopsin (bR) in the purple membrane has been visualized by HS-AFM (Shibata et al. 2010). A large structural change of bR is observed at the cytoplasmic surface; each bR molecule moves outward from the center of a trimer by  $\sim 0.8$  nm upon light illumination (Fig. 2b, c). The following dynamic processes have also been visualized by HS-AFM: association and dissociation of bR trimers at the free interfaces of the crystalline phase (Yamashita et al. 2009); anisotropic diffusion of single point defects in 2D crystals of streptavidin (Fig. 2d) (Yamamoto et al. 2008); amyloid-like fibril formation by lithostathine (Milhiet et al. 2010); formation of planar lipid bilayers

from liposomes (Fig. 2e) (Giocondi et al. 2010); and counterclockwise propagation of conformational states in  $F_1$ -ATPase ring without  $\gamma$  subunit (Uchihashi et al. 2011).

## Cross-References

- ▶ [Atomic Force Microscopy](#)
- ▶ [Atomic Force Microscopy \(AFM\) for Topography and Recognition Imaging at Single Molecule Level](#)
- ▶ [Phospholipid Bilayer Nanodiscs: Application to Single-Molecule Measurements](#)

## References

- Ando T, Kodera N, Takai E, Maruyama D, Saito K, Toda A. A high-speed atomic force microscope for studying biological macromolecules. *Proc Natl Acad Sci USA*. 2001;98:12468–72.
- Ando T, Uchihashi T, Fukuma T. High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. *Prog Surf Sci*. 2008;83:337–437.
- Giocondi MC, Yamamoto D, Lesniewska E, Milhiet PE, Ando T, Le Grimmelc C. Surface topography of membrane domains. *Biochim Biophys Acta -Biomembranes* 2010;1798:703–18.
- Kodera N, Sakashita M, Ando T. Dynamic proportional-integral-differential controller for high-speed atomic force microscopy. *Rev Sci Instrum*. 2006;77:083704 (7 pp).
- Kodera N, Yamamoto D, Ishikawa R, Ando T. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature*. 2010;468:72–6.
- Milhiet PE, Yamamoto D, Berthoumieu O, Dosset P, Le Grimmelc C, Verdier JM, Marchal S, Ando T. Deciphering the structure, growth and assembly of amyloid-like fibrils using high-speed atomic force microscopy. *PLoS One*. 2010;5:e13240 (8 pp).
- Shibata M, Yamashita H, Uchihashi T, Kandori H, Ando T. High-speed atomic force microscopy shows dynamic molecular processes in photo-activated bacteriorhodopsin. *Nat Nanotechnol*. 2010;5:208–12.
- Uchihashi T, Iino R, Ando T, Noji H. High-speed atomic force microscopy reveals rotary catalysis of rotor-less  $F_1$ -ATPase. *Science* 2011;333:755–8.
- Yamamoto D, Uchihashi T, Kodera N, Ando T. Anisotropic diffusion of point defects in two-dimensional crystal of streptavidin observed by high-speed atomic force microscopy. *Nanotechnology*. 2008;19:384009 (9 pp).
- Yamamoto D, Uchihashi T, Kodera N, Yamashita H, Nishikori S, Ogura T, Shibata M, Ando T. High-speed atomic force microscopy techniques for observing dynamic biomolecular processes. *Methods Enzymol*. 2010;475(B): 541–64.
- Yamashita H, Voitchovsky K, Uchihashi T, Antoranz Contera S, Ryan JF, Ando T. Dynamics of bacteriorhodopsin 2D crystal observed by high-speed atomic force microscopy. *J Struct Biol*. 2009;167:153–8.

---

## Hilbert Phase Contrast (HPC)

- ▶ [Phase Contrast Electron Microscopy](#)

---

## Histone-Lysine *N*-methyltransferase

- ▶ [Protein Lysine Methyltransferase – Computational Studies](#)

---

## History of Infrared Spectroscopy

- ▶ [Vibrational Spectroscopy, A Short History of](#)

---

## History of Raman Spectroscopy

- ▶ [Vibrational Spectroscopy, A Short History of](#)

---

## HIV Protease – Computational Studies

Marc Willem van der Kamp  
Centre for Computational Chemistry, School of  
Chemistry, University of Bristol, Bristol, UK

### Definition

An aspartyl protease that is essential for the life cycle of HIV by cleaving newly synthesized polyproteins into the mature protein components of the infectious HIV virion.

### Basic Characteristics

Three enzymes that are essential for the replication of the human immunodeficiency virus type 1 (HIV-1): a reverse transcriptase, a protease, and an integrase. These enzymes are therefore important targets in the development of anti-AIDS drugs. Common current treatment of HIV/AIDS involves combining drugs

that inhibit the reverse transcriptase and the protease. ▶ [QM/MM methods](#) have been employed to model all three enzymes, to help understand their mechanism, specificity, and effects of genetic variation. HIV-1 protease/inhibitor interactions have been studied extensively with QM/MM methods. For example, the subtractive ▶ [ONIOM QM/MM method](#) was used to investigate the role of a conserved water molecule in six enzyme-inhibitor complexes (Suresh et al. 2008). This water molecule forms four hydrogen bonds to bridge the Ile50/50' NH groups in the active site cleft opposite from the aspartyl dyad (the so-called flaps of the enzyme dimer) and the CO groups at positions P2 and P'1 of the inhibitor. The approved inhibitor tipranavir is designed to replace the water molecule and semiempirical QM/MM (AM1/CHARMM22) ▶ [Molecular Dynamics Simulations](#) indeed indicated that tipranavir interacts with both Asp25/25' and Ile50/50', explaining its potency (Hensen et al. 2004). Garrec et al. (2010) used QM/MM ▶ [Car-Parrinello molecular dynamics](#) to investigate a new class of HIV-1 protease inhibitors, based on the presence of a noncovalent interaction between a tertiary amine and a carbonyl (Gautier et al. 2006). By considering an amino-aldehyde peptide bound to the enzyme in all possible protonation states of the aspartyl dyad, they concluded that the interaction is disrupted through competition with the hydrogen bond network around the conserved water molecule. Inhibitors that combine the noncovalent tertiary amine–carbonyl interaction with replacing the conserved water may therefore be promising. A recent study further demonstrates the complementarity of QM/MM modeling, docking, and experimental methods in the analysis of another new class of HIV-1 protease inhibitors (Makatini et al. 2011).

### Cross-References

- ▶ [Car-Parrinello Molecular Dynamics](#)
- ▶ [Molecular Dynamics Simulations of Lipids](#)
- ▶ [ONIOM](#)
- ▶ [QM/MM Methods](#)
- ▶ [Semiempirical Quantum Mechanical Methods](#)

### References

- Garrec J, Cascella M, et al. Low inhibiting power of N...CO based peptidomimetic compounds against HIV-1

- protease: insights from a QM/MM study. *J Chem Theory Comput.* 2010;6(4):1369–79.
- Gautier A, Pitrat D, et al. An unusual functional group interaction and its potential to reproduce steric and electrostatic features of the transition states of peptidolysis. *Bioorg Med Chem.* 2006;14(11):3835–47.
- Hensen C, Hermann JC, et al. A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J Med Chem.* 2004;47(27):6673–80.
- Makatini MM, Petzold K, et al. Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach. *Eur J Med Chem.* 2011;46(9):3976–85.
- Suresh CH, Vargheese AM, et al. Role of structural water molecule in HIV protease-inhibitor complexes: a QM/MM study. *J Comput Chem.* 2008;29(11):1840–9.

## HIV-1 Reverse Transcriptase – Computational Studies on the Polymerase Active Site

Thanyada Rungrotmongkol<sup>1</sup>, Nadtanet Nunthaboot<sup>2</sup>, Ornjira Aruksakunwong<sup>3</sup> and Supot Hannongbua<sup>4</sup>  
<sup>1</sup>Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand  
<sup>2</sup>Department of Chemistry, Faculty of Science, Mahasarakham University, Mahasarakham, Thailand  
<sup>3</sup>Department of Chemistry, Faculty of Science, Rangsit University, Pathumtani, Thailand  
<sup>4</sup>Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

### Synonyms

Reverse transcriptase

### Definition

Human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT), a multifunctional enzyme, converts the single-stranded viral genomic RNA into the double-stranded DNA, which can then be integrated into the host cell's chromosomes.

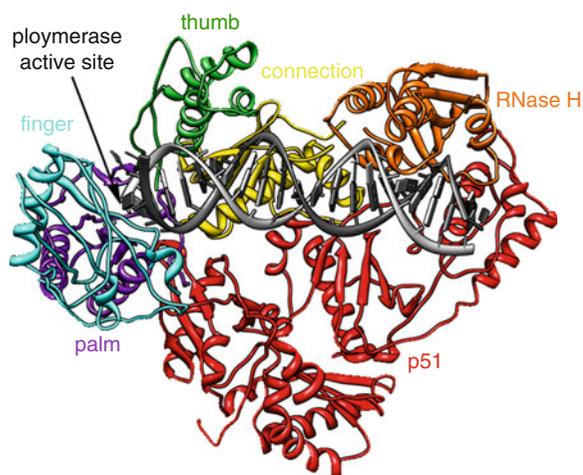
### Introduction

Acquired immunodeficiency syndrome (AIDS), caused by the human immunodeficiency virus (HIV),

has been a worldwide health problem for over 25 years. HIV-1 reverse transcriptase (RT), one of the three essential enzymes in the HIV replication cycle, is an important target for anti-AIDS drug therapy. Two classes of FDA-approved HIV-1 RT inhibitors, non-nucleoside/nucleotide (NNRTI) and nucleoside/nucleotide (NRTI), are important components of the combination therapy using the highly active anti-retroviral therapy (HAART). However, the emergence of drug resistance and their severe side effects limits the efficiency of these antiviral agents. Therefore, new potent and highly selective inhibitors are urgently required.

HIV-1 RT functions by converting the single-stranded viral genomic RNA into double-stranded DNA, which can then be integrated into the genomic DNA of the host cell. The process requires at least three enzymes: RNA-dependent DNA polymerase, DNA-dependent DNA polymerase, and ribonuclease H (RNase H). The enzymatic activity of the RNA/DNA-dependent DNA polymerase is the copying of either an RNA or a DNA template, and that of the RNase H is the RNA degradation of the RNA/DNA heteroduplex (Skowron and Ogden 2006; Sarafianos et al. 2009).

HIV-1 RT, an asymmetric heterodimer protein, is composed of 1,000 amino acids with two related subunits: p66 and p51 (Fig. 1). The larger p66 subunit contains 560 residues and is responsible for the RT catalytic functions. The smaller p51 subunit possesses 440 amino acids and exhibits a structural



**HIV-1 Reverse Transcriptase – Computational Studies on the Polymerase Active Site, Fig. 1** The crystal structure of the HIV-1 RT/dsDNA/dTTP ternary complex (Huang et al. 1998)

role. The p66 subunit comprises the DNA polymerase and RNase H domains. The inhibitor binding sites of NNRTI and NRTI are located in the polymerase domain, which is composed of four distinct subdomains: fingers (residues 1–85 and 118–155), palm (residues 86–117 and 156–236), thumb (residues 237–318), and connection (residues 319–426). The p51 also folds similarly to the four corresponding subdomains (finger, palm, thumb, and connection) but the relative orientations of its subdomains are substantially different from those of p66 (Sarafianos et al. 2009).

### Active Structure of the HIV-1RT/dsDNA/dNTP Ternary Complex

The reaction center of polymerization takes place at the polymerase active site (Fig. 1) of HIV-1 RT in the palm subdomain harboring a catalytic triad of aspartates D110, D185, and D186. Two bound magnesium cations ( $Mg^{2+}$ ) are required to form a complete active site. By molecular dynamics (MD) simulations, both  $Mg^{2+}$  ions were observed to be octahedrally coordinated by six ligands with the Mg-O distance of  $\sim 2 \text{ \AA}$  (Rungrotmongkol et al. 2007), as shown in Fig. 2. The coordination of the  $_A Mg^{2+}$  ion at catalytic site A is comprised of the carboxylate oxygens of D110, D185, and D186, the non-bridging  $O1\alpha$  oxygen of the  $\alpha$ -phosphate of dTTP, the 3'-OH group of the primer terminus, and a water molecule. Note that the last two ligands are not present in the crystal structure (Huang et al. 1998). At the nucleotide-binding site B, the  $_B Mg^{2+}$  ion coordinates with the three non-bridging  $O1\alpha$ ,  $O2\beta$ , and  $O1\gamma$  oxygens of the triphosphate moiety of dTTP, the carboxylate oxygens of D110 and D185, and the backbone carbonyl oxygen of V111. The incoming dNTP in the fully deprotonated form is more tightly bound to the active site and is likely to be more stable than the mono-protonated form due to its greater hydrogen bond formation with the positively charged side chains of K65 and R72 and the peptide backbones of D113 and A114 (Rungrotmongkol et al. 2007; Frank and Carloni 2000).

### Role of the Catalytic Magnesium Ion

DNA polymerases require divalent metal ions for their catalytic function (Bloomfield et al. 2000).

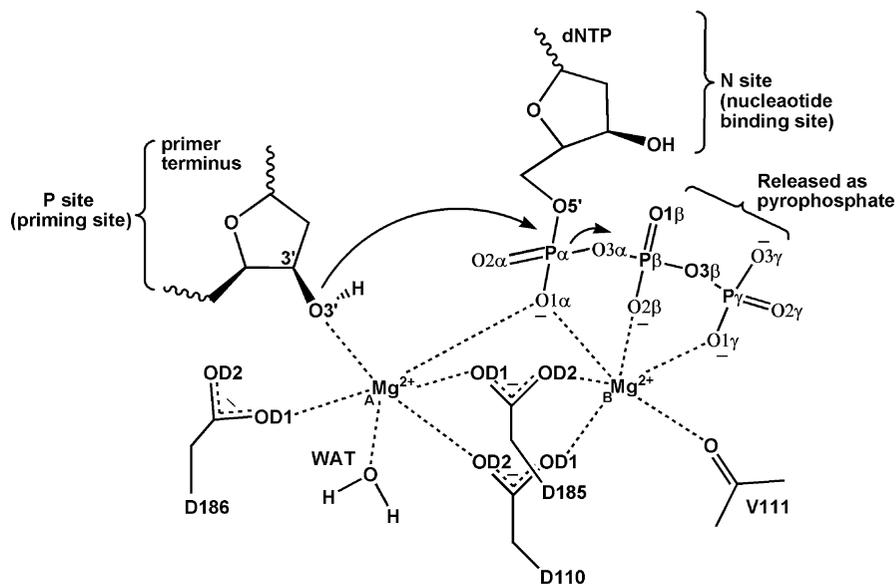
The functions of a DNA polymerase and its fidelity of DNA synthesis in vivo depend on the type of divalent cation (e.g.,  $Mg^{2+}$ ,  $Mn^{2+}$ ,  $Ca^{2+}$ ,  $Zn^{2+}$ ,  $Ni^{2+}$ ,  $Co^{2+}$  or  $Be^{2+}$ ) occupying the polymerase active site. In the HIV-1 RT, the  $_A Mg^{2+}$  ion (at catalytic site A) is thought to decrease the  $pK_a$  of the 3'-OH group of the terminal primer, facilitating the nucleophilic attack on the  $\alpha$ -phosphorus atom of the dNTP substrate. In contrast, the nucleotide-binding  $_B Mg^{2+}$  ion (at site B) is believed to stabilize the triphosphate of the dNTP substrate and the pyrophosphate (PPi) group prior to and after the nucleotidyl transferase reaction, respectively. A  $Be^{2+}$  ion substitution for the  $Mg^{2+}$  ion at either the catalytic or the nucleotide-binding site, acting as a cofactor for the polymerization reaction of HIV-1 RT, has been evaluated theoretically using MD simulations in comparison with experimentally observed pre-steady-state kinetics (Mendieta et al. 2008). The tetrahedral character of the  $Be^{2+}$  coordination was observed at both sites. Three conserved aspartates (D110, D185, and D186) and a water molecule act as the ligands for the  $_A Be^{2+}$  ion complex, while the  $_B Be^{2+}$  coordination environment depends on whether  $Be^{2+}$  or  $Mg^{2+}$  occupies the catalytic site A. Interest is focused on the critical distance for nucleophilic attack from the 3'-OH oxygen of the primer terminus to the  $\alpha$ -phosphorus of the dTTP substrate,  $d[O3'-P\alpha]$ . Having  $Be^{2+}$  or no metal ion at the catalytic site A leads to a significant increase in  $d[O3'-P\alpha]$  by at least  $1.3 \text{ \AA}$  or  $2.3 \text{ \AA}$ , respectively, relative to that of  $3.2 \text{ \AA}$  in the HIV-1 RT ternary complex containing two  $Mg^{2+}$  ions. In contrast, the nucleophilic attack distance is not affected by replacing  $Mg^{2+}$  with  $Be^{2+}$  at the nucleotide-binding site. This finding supports the experimental data, i.e., that DNA polymerase activity is increased by  $Be^{2+}$  ions at a relatively low concentration through binding with the dNTP at site B, while this divalent metal ion at high concentration could compete with the  $Mg^{2+}$  ion in binding at the catalytic site and thus inhibit the DNA polymerization.

### DNA Polymerization

RT is a member of the family of DNA polymerases that assists in DNA replication from the 5'- to 3'-direction (Skalka and Goff 1993). The mechanistic reaction of DNA polymerization contains two

### HIV-1 Reverse Transcriptase – Computational Studies on the Polymerase Active Site,

**Fig. 2** The proposed mechanism of DNA polymerization based on the modeled structure of the HIV-1 RT/dsDNA/dTTP ternary complex (Rungrotmongkol et al. 2007), where the metal chelation in the polymerase active site is also shown



steps: deprotonation of the 3'-primer terminus followed by the nucleotidyl transfer reaction. Although an overall mechanism of polymerization has been suggested, the general base in deprotonation is uncertain, e.g., it has been reported to be the active site aspartate in T7 DNA polymerase (Florián et al. 2003), the  $\alpha$ -phosphate of dNTP in polymerase  $\beta$  (Bojin and Schlick 2007), and ordered water molecule in *sofataricus* DNA polymerase IV (Wang et al. 2007). The only existing data for HIV-1 RT, obtained from the combined QM/MM MD simulations, demonstrates the DNA polymerization reaction pathway with the possibility of the conserved aspartate D185 acting as a proton acceptor to abstract the hydrogen from the 3'-OH group of the primer sugar ring (Rungrotmongkol et al. 2004). Note that D185 is equivalent to D654, which seems to be the general base in T7 DNA polymerase. After deprotonation, the negatively charged hydroxyl moiety of the 3'-primer then attacks the  $\alpha$ -phosphorus of dTTP with formation of the stable pentacovalent intermediate. Finally, cleavage of the P $\alpha$ -O3 $\alpha$  bond occurs to leave pyrophosphate from the newly formed 3'-5' phosphodiester, leading to the elongation of the DNA primer by an added nucleotide. The hydrogen bond interactions with the  $\epsilon$ -amino group K65, the guanidinium group of R74 and the peptide backbone of D113 and A114 are maintained along the reaction pathway.

## References

- Bloomfield VA, Crothers DM, et al. Nucleic acids: structures, properties, and functions. Sausalito: University Science; 2000.
- Bojin MD, Schlick T. A quantum mechanical investigation of possible mechanisms for the nucleotidyl transfer reaction catalyzed by DNA polymerase  $\beta$ . *J Phys Chem B*. 2007;111:11244–52.
- Florián J, Goodman ME, Warshel A. Computer simulation of the chemical catalysis of DNA polymerases: discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. *J Am Chem Soc*. 2003;125:8163–77.
- Frank A, Carloni P. Ab initio molecular dynamics studies on HIV-1 reverse transcriptase triphosphate binding site: implications for nucleoside analog drug resistance. *Protein Sci*. 2000;9:2535–46.
- Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*. 1998;282:1669–75.
- Mendieta J, Cases-González CE, Matamoros T, Ramírez G, Menéndez-Arias L. A Mg<sup>2+</sup>-induced conformational switch rendering a competent DNA polymerase catalytic complex. *Proteins*. 2008;71:565–74.
- Rungrotmongkol T, Hannongbua S, Mulholland AJ. Mechanistic study of HIV-1 reverse transcriptase at the active site based on QM/MM method. *J Theor Comput Chem*. 2004;3:491–500.
- Rungrotmongkol T, Mulholland AJ, Hannongbua S. Active site dynamics and combined quantum mechanics/molecular mechanics (QM/MM) modelling of a HIV-1 reverse transcriptase/DNA/dTTP complex. *J Mol Graph Model*. 2007;26:1–13.
- Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, Arnold E. Structure and function of HIV-1

reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol.* 2009;385:693–713.

Skalka AM, Goff SP. Reverse transcriptase, Cold spring harbor monograph series. Plainview: Cold Spring Harbor Laboratory Press; 1993.

Skowron G, Ogden R. Infectious disease: reverse transcriptase inhibitors in HIV/AIDS therapy; structural studies on HIV reverse transcriptase related to drug discovery. Totowa: Humana Press; 2006. p. 1–47.

Wang L, Yu X, Hu P, Broyde S, Zhang Y. A water-mediated and substrate-assisted catalytic mechanism for *Sulfolobus solfataricus* DNA polymerase IV. *J Am Chem Soc.* 2007; 129:4731–7.

---

## HMM

► [Hidden Markov Modeling in Single-Molecule Biophysics](#)

---

## Homologue Scanning

► [Mutagenic Scanning to Define Binding Hotspots](#)

---

## Homology Modeling of Protein Structures

Torsten Schwede  
Swiss Institute of Bioinformatics, Biozentrum  
University of Basel, Basel, Switzerland

### Synonyms

[Biomolecular modeling](#); [Comparative protein structure modeling](#); [Fold recognition](#); [Protein structure modeling](#); [Template-based modeling](#)

### Definition

The three-dimensional structure of a protein can be modeled starting from its amino acid sequence by using information from homologous proteins with known structure.

## Introduction

Three-dimensional structures of proteins provide invaluable information for our understanding of macro-molecular function at an atomic level of detail. However, despite the tremendous progress in experimental techniques in recent years, structure solution by crystallography or ► [NMR](#) remains a complex process, and as a consequence for the majority of proteins in any given organism no direct experimental structures are available today. Computational approaches for predicting the three-dimensional structure of a protein from its amino acid sequence appear, therefore, as an attractive alternative (Schwede et al. 2008; Venselaar et al. 2009). Various approaches for structure prediction have been developed in recent years, among which homology modeling is currently by far the most accurate and reliable method (Moult et al. 2011). Protein structure homology modeling techniques are used to construct three-dimensional models for proteins from their amino acid sequences using information from homologous proteins with known structure. The expected accuracy of homology models depends on the evolutionary distance between the protein of interest (“target”) and the homologous proteins used as structural “templates.” The relationship between sequence identity and structural similarity, which was first described by Chothia and Lesk for the globin family (Chothia and Lesk 1986), still provides a good first approximation today for the expected accuracy of a homology model.

## Template Detection and Target-Template Alignment

The first step in building a homology model for a target protein consists of searching for homologous proteins with known experimental structures (Table 1, Fig. 1). Pair-wise sequence comparison methods are able to detect and correctly align protein sequences sharing high sequence similarity. For detecting remote homology relationships, more sensitive profile- or Hidden Markov Model (HMM)-based methods are applied (Remmert et al. 2011). Often, several alternative template structures are identified for a given target protein. Besides the evolutionary distance to the target, other factors such as experimental resolution and quality of the template, and presence/absence of

**Homology Modeling of Protein Structures, Table 1** Resources on experimental structure information and structural annotation of genomes

*Resources on experimental structure information*

**PDB:** The worldwide Protein Data Bank (wwPDB; RCSB PDB (USA), PDBe (Europe), and PDBj (Japan)) maintains the single archive of experimental macromolecular structural data

[www.rcsb.org](http://www.rcsb.org)  
[www.PDBe.org](http://www.PDBe.org)  
[www.PDBj.org](http://www.PDBj.org)

**CATH:** CATH is a manually curated hierarchical domain classification of protein structures in the Protein Data Bank

[www.cathdb.info/](http://www.cathdb.info/)

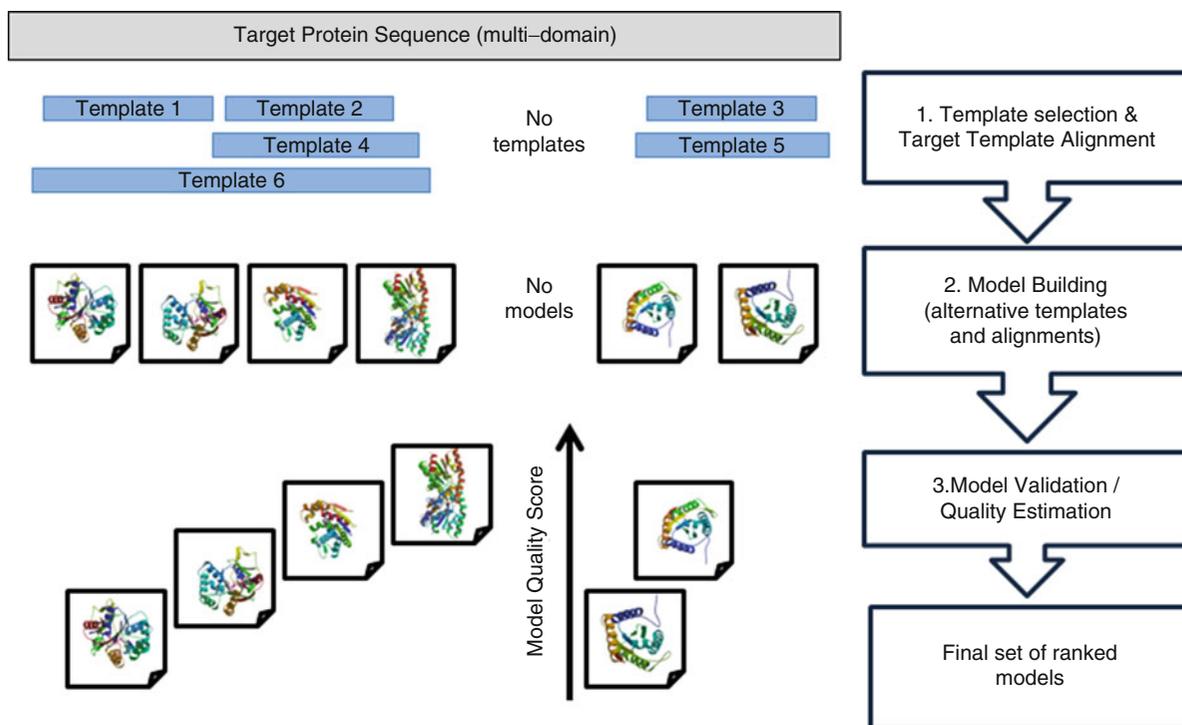
*Resources on structural annotation of genomes*

**SUPERFAMILY** is a database of structural and functional annotation for all proteins and genomes based on a collection of hidden Markov models, which represent structural protein domains at the SCOP superfamily level

<http://supfam.cs.bris.ac.uk/>

**Gene3D** is a database of CATH protein domain assignments for ENSEMBL genomes and Uniprot sequences. Finer grained groupings of superfamilies based on function (FunFams) and structure (SCs) are also provided

<http://gene3d.biochem.ucl.ac.uk/>



**Homology Modeling of Protein Structures, Fig. 1** Typical homology modeling workflow. Starting from the amino acid sequence of the target protein, possible alternative template structures are identified by searching a library of experimental structures (1). Based on the aligned templates, typically one or

more alternative models are constructed (2), and assessed for their expected accuracy using model quality estimation tools (3), resulting in a ranked ensemble of alternative models. Note that homology modeling techniques cannot generate models in the absence of template information

relevant ligands or cofactors should be considered when selecting suitable templates. In many cases, template structures will cover only individual domains rather than the full length of the target sequence.

The result of the template detection step is an alignment between the target sequence and the sequences of one or more homologous template protein structures.

## Model Building

In the model building step, structural information from the templates is used to construct a three-dimensional model for the target protein using the sequence alignment from the previous step to identify corresponding residues. Two main approaches are used for constructing the model backbone coordinates: Fragment-based methods assemble the core of the model from structural fragments extracted from the templates. Alternatively, restraint optimization methods derive spatial restraints from the template structures and construct a model which optimally satisfies these constraints (Sali and Blundell 1993). While this approach is typically applied to individual domains, the general concept is valid also for modeling multi-domain proteins and oligomeric assemblies – as long as suitable homologous template structures are available (see Table 2 for a list of commonly used modeling tools and resources).

While structurally conserved regions of the target can be modeled using template structure information, variable regions have to be reconstructed “de novo” – specifically structurally variable loop regions and segments corresponding to insertions and deletions in the target-template alignment. Typically sampling techniques are used to enumerate an ensemble of alternative loop conformations which are then scored by empirical scoring functions to select the most likely arrangement (Rohl et al. 2004). In general, reconstructed loop regions are the least reliable parts of a homology model (Fig. 2).

Modeling the correct side chain packing is often (mainly for technical reasons) considered a separate step in homology modeling. Typically, alternative side chain conformations of amino acids are represented in so-called rotamer libraries as a function of backbone geometry (Shapovalov and Dunbrack 2011). The representation of side-chain conformation as discrete states allows for numerically efficient methods for searching optimal conformations such as the dead-end elimination (DEE) algorithm.

One of the remaining challenges in protein structure modeling is the prediction of interactions with ligands, cofactors, and substrates. Most current methods are based on transferring information about ligands observed in homologous protein structures to the target-binding site (Wass et al. 2011). While such comparative approaches work well for

predicting evolutionary conserved interactions in well-characterized protein families, they are limited with respect to predicting changes in specificity and detecting novel, previously not observed, binding sites.

## Model Refinement

One of the last steps in model building often consists in refining the coordinates of the initial model using energy minimization techniques. One aim is to correct unfavorable contacts or suboptimal conformations introduced during the modeling process by identifying energetically more favorable conformations.

A second goal of model refinement is to move the model coordinates closer to the “real structure”: Comparative modeling techniques tend to construct models, which closely resemble the template structure. Refining the models away from the templates closer to the target has been a question of intensive research, and many different solutions have been proposed to address this problem. However, as of today, no solution has emerged which was able to consistently improve the backbone geometry of models closer to the correct reference structure in controlled blind experiments (MacCallum et al. 2011). As a consequence, homology models are often not predictive for the structural changes occurring as a result of sequence variations, such as single point mutations and SNPs.

## Model Quality Assessment, Validation, and Quality Estimation

Depending on the evolutionary distance between the target protein and the template structure, homology models can vary significantly in accuracy. While high-quality models can reach the accuracy of medium-resolution X-ray structures, low-quality models may contain significant errors and inaccuracies. Model validation is, therefore, an essential component of the homology modeling process.

The overall accuracy of different modeling techniques can be assessed retrospectively by comparing blind predictions for a large number of target proteins to their experimental reference structures. Within the biannual CASP experiment – Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (Moult et al. 2011),

**Homology Modeling of Protein Structures, Table 2** *Modeling tools and resources.* The following list provides a selection of commonly used tools and publicly available resources. For a more complete listing, please refer to the annual web server issue of the journal *Nucleic Acids Research*

*Model portals and resource collections*

**Protein Model Portal (PMP):** a portal to protein structure information (both experimental and models), interactive services for model building, and quality assessment [www.proteinmodelportal.org](http://www.proteinmodelportal.org)

**Protein Modeling 101:** a collection of resources and tutorials related to protein structure modeling [www.proteinmodelportal.org/101/](http://www.proteinmodelportal.org/101/)

*Modeling servers and tools*

**HHpred:** homology detection and structure prediction by HMM-HMM comparison [hhpred.tuebingen.mpg.de/hhpred/](http://hhpred.tuebingen.mpg.de/hhpred/)

**I-Tasser:** service for protein structure predictions built based on multiple-threading alignments by LOMETS and iterative TASSER assembly simulations [zhanglab.ccmb.med.umich.edu/I-TASSER/](http://zhanglab.ccmb.med.umich.edu/I-TASSER/)

**Modeller:** software tool for comparative modeling by satisfaction of spatial restraints [salilab.org/modeller/](http://salilab.org/modeller/)

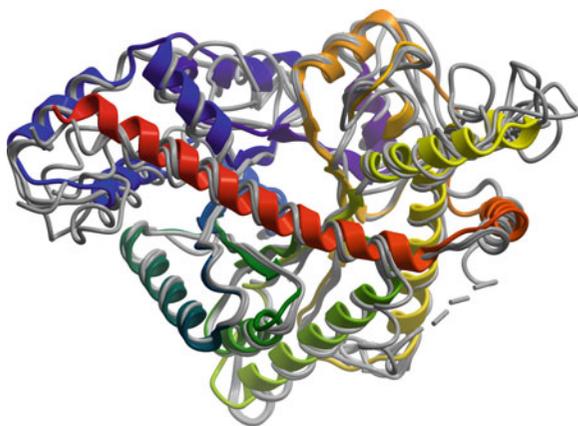
**ModWeb:** comparative protein structure modeling server based on the large-scale protein structure modeling pipeline, ModPipe, and Modeller [salilab.org/modweb](http://salilab.org/modweb)

**Phyre2:** Protein Homology/analogy Recognition Engine: a web server for fold recognition, protein modeling, and binding site prediction [www.sbg.bio.ic.ac.uk/phyre2/](http://www.sbg.bio.ic.ac.uk/phyre2/)

**Robetta:** protein structure prediction server building both ab initio and comparative models of proteins [rosetta.bakerlab.org/](http://rosetta.bakerlab.org/)

**SCWRL4:** efficient software tool for prediction of protein side-chain conformations using a backbone-dependent rotamer library [dunbrack.fccc.edu/scwrl4/](http://dunbrack.fccc.edu/scwrl4/)

**SWISS-MODEL:** fully automated web-based service for protein structure homology modeling [swissmodel.expasy.org](http://swissmodel.expasy.org)



**Homology Modeling of Protein Structures, Fig. 2** Crystal structure of transaldolase from *Corynebacterium glutamicum* in comparison with several homology models. As part of the CAMEO evaluation of structure prediction servers, the amino acid sequence of transaldolase from *Corynebacterium glutamicum* was sent to several modeling pipelines, and the results were compared to the experimental reference structure after the coordinates were released (PDB: 3R5E; Samland et al. (2012)). The experimental reference structure is shown as *ribbon representation*, colored from N- to C-terminus from *blue* to *red*; homology models are shown as *gray tubes*. While the structural core of the protein was modeled consistently and accurately by all modeling servers, significant variations are observed in the loop regions and at the surface of the protein, where also the largest deviations from the experimental reference structure are encountered

homology modeling techniques are evaluated in the category “template-based modeling” (Mariani et al. 2011). The CAMEO project (Continuous Automated Model Evaluation, <http://cameo3d.org/>) continuously evaluates the accuracy and reliability of protein structure prediction services in a fully automated manner. One noteworthy outcome of these efforts is that the difference in accuracy between the best performing modeling methods is small compared to the differences observed between different target proteins. Validation and absolute model quality estimation for individual models is, therefore, essential.

Since at the time of modeling, the “real structure” of the target protein is unknown, the expected accuracy of a model needs to be estimated. Two types of criteria are typically applied in model validation: “Hard criteria” evaluate the physical and stereochemical plausibility of the model’s coordinates – essentially applying the same validation criteria as for experimental structures such as atomic distances, bond lengths, and angles. While these criteria are necessary for correct model geometry, they are not sufficient to ensure that the model correctly represents the structure of the target protein. “Soft criteria” are used to estimate the expected quality of a model, or to rank models within an ensemble of alternatives. Widely used for this

**Homology Modeling of Protein Structures, Table 3** Tools for model validation and quality assessment*Protein model validation tools (stereochemical quality)*

**PROCHECK**: checks the stereochemical quality of a protein structure by analyzing its overall and residue-by-residue geometry [www.ebi.ac.uk/thornton-srv/software/PROCHECK/](http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/)

**WHAT\_CHECK**: protein verification tools from the WHAT IF package; detailed analysis of the stereochemical quality of a protein structure or model [swift.cmbi.ru.nl/gv/whatcheck/](http://swift.cmbi.ru.nl/gv/whatcheck/)

**MOLPROBITY**: all-atom validation for macromolecular crystallography and NMR structures [molprobity.biochem.duke.edu/](http://molprobity.biochem.duke.edu/)

*Model quality estimation tools*

**QMEAN**: quality estimation server for protein structure models using a composite scoring function, including distance-dependent interaction potentials of mean force, torsion angle potentials, and solvation potentials [swissmodel.expasy.org/qmean/](http://swissmodel.expasy.org/qmean/)

**ModEval**: evaluates various aspects of model quality using the atomic distance-dependent statistical potential DOPE and statistical potential GA341 for assessing the reliability of a model [salilab.org/modeval/](http://salilab.org/modeval/)

**ModFold**: a quality assessment program used for global and local assessment of models, using the comparison of multiple models without structural alignments [www.reading.ac.uk/bioinf/ModFOLD/](http://www.reading.ac.uk/bioinf/ModFOLD/)

**ProQres**: a neural network predictor that predicts the quality of different parts of protein model based on a number of structural features [www.sbc.su.se/~bjornw/ProQ/](http://www.sbc.su.se/~bjornw/ProQ/)

*Model quality assessment*

**CASP**: the community wide experiment on the critical assessment of techniques for protein structure prediction aims at an in-depth and objective assessment of our current abilities and inabilities in the area of protein structure prediction [predictioncenter.org](http://predictioncenter.org)

**CAMEO**: the continuous automated model evaluation project continuously evaluates the accuracy and reliability of various protein structure prediction services in a fully automated manner [cameo3d.org](http://cameo3d.org)

approach are statistical potentials of mean force for atomic interactions, backbone geometry, or solvent exposure, which are derived from high-resolution crystal structures (Table 3). In case a large number of models based on independent modeling methods are available, consensus approaches have been shown to be successful in selecting the most accurate models within these ensembles (Kryshtafovych et al. 2011). In the context of practical applications of models, error estimates for individual residues in single models would be required on an absolute scale, which is, however, a much more difficult task.

## Applications of Models

With the development of stable, reliable, and fully automated techniques, protein structure homology models have found widespread use in biomedical research as source for structure information in cases

where no suitable experimental structure of the target protein is available. Since different applications have various requirements on the quality of the structural information, ultimately the accuracy of individual models determines their usability for specific applications (Petrey and Honig 2005; Schwede et al. 2009).

Typical applications of homology models include experimental design for protein engineering and functional characterization studies by site-directed mutagenesis; the interpretation of disease-related mutations and SNPs to understand the molecular mechanisms of a disease; phasing of X-ray diffraction data in protein crystallography by ► [molecular replacement](#) (Schwarzenbacher et al. 2008); or virtual ligand screening in structure-based drug discovery projects. Often, computational structure models are combined with other sources of low-resolution experimental constraints, such as cryo-EM maps or proteomics cross-linking data, in an integrative approach to protein structure solution (Robinson et al. 2007).

## Modeling Tools and Resources

Protein structure homology modeling techniques have matured in recent years from research projects into stable and reliable automated pipelines. SWISS-MODEL was the first web server to offer homology modeling as a fully automated service to the research community (Guex et al. 2009). Today, a broad variety of software tools and web servers for protein structure homology modeling are available, which are easy to use also for the non-expert. Table 2 provides a selection of commonly used modeling tools and servers.

## Summary

The amount of new protein sequences is growing at an astounding rate thanks to ongoing genome sequencing and meta-genomics efforts. As a consequence, the gap between the number of known protein sequences and experimental structures is widening continuously – despite the tremendous progress made in optimizing the experimental structure determination process, for example, as part of worldwide structural genomics efforts. Computational techniques for protein structure modeling allow extrapolating the available structure information to homologous protein sequences (Godzik 2011). Today, structural information – either experimental or homology models – is available for the majority of amino acids in well-characterized model organism proteomes like *E. coli* or *T. maritima*. Fully automated, reliable, and accurate homology modeling pipelines make structure information readily available for a wide spectrum of structure-guided applications in biomedical research. However, homology models still fall short of accuracy in comparison to experimental structures, and the development of reliable techniques for model refinement (Zhang et al. 2011) closer to the native structure remains a challenging research goal for the coming years.

## Cross-References

- ▶ [Alignment of Protein Sequences](#)
- ▶ [Macromolecular Crystallography: Overview](#)
- ▶ [Modeling of Antibody Structures](#)
- ▶ [Molecular Dynamics Simulations of Lipids](#)
- ▶ [Molecular Replacement](#)

- ▶ [Protein Oligomers](#)
- ▶ [Protein Structural Models – Evaluating Quality](#)
- ▶ [Structural Genomics](#)
- ▶ [Structural Impact of SNPs](#)

## References

- Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 1986; 5:823–6.
- Godzik A. Metagenomics and the protein universe. *Curr Opin Struct Biol.* 2011;21:398–403.
- Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis.* 2009;30:S162–73.
- Kryshtafovych A, Fidelis K, Tramontano A. Evaluation of model quality predictions in CASP9. *Proteins.* 2011;79: 91–106.
- MacCallum JL, Pérez A, Schnieders MJ, Hua L, Jacobson MP, Dill KA. Assessment of protein structure refinement in CASP9. *Proteins.* 2011;79:74–90.
- Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based protein structure predictions in CASP9. *Proteins.* 2011;79:37–58.
- Moult J, Fidelis K, Kryshtafovych A, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)–round IX. *Proteins.* 2011;79:1–5.
- Petrey D, Honig B. Protein structure prediction: inroads to biology. *Mol Cell.* 2005;20:811–9.
- Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods.* 2011;9:173–5.
- Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. *Nature.* 2007;450:973–82.
- Rohl CA, Strauss CE, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins.* 2004;55:656–77.
- Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234:779–815.
- Samland AK, et al. Conservation of structure and mechanism within the transaldolase enzyme family. *FEBS J.* 2012;279:766–778.
- Schwarzenbacher R, Godzik A, Jaroszewski L. The JCSG MR pipeline: optimized alignments, multiple models and parallel searches. *Acta Crystallogr D Biol Crystallogr.* 2008;64:133–40.
- Schwede T, Sali A, Eswar N, Peitsch MC. Protein structure modeling. In: Schwede T, Peitsch MC, editors. *Computational structural biology: methods and applications.* Singapore: World Scientific; 2008. p. 3–35.
- Schwede T, Sali A, Honig B, Levitt M, Berman HM, Jones D, Brenner SE, Burley SK, Das R, Dokholyan NV, Dunbrack RLJ, Fidelis K, Fiser A, Godzik A, Huang YJ, Humblet C, Jacobson MP, Joachimiak A, Krystek SRJ, Kortemme T, Kryshtafovych A, Montelione GT, Moult J, Murray D, Sanchez R, Sosnick TR, Standley DM, Stouch T, Vajda S, Vasquez M, Westbrook JD, Wilson IA. Outcome of

- a workshop on applications of protein models in biomedical research. *Structure*. 2009;17:151–9.
- Shapovalov MV, Dunbrack RLJ. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*. 2011;19:844–58.
- Venselaar H, Krieger E, Vriend G. Homology modeling. In: Gu J, Bourne P, editors. *Structural bioinformatics*. 2nd ed. Hoboken: Wiley; 2009.
- Wass MN, David A, Sternberg MJ. Challenges for the prediction of macromolecular interactions. *Curr Opin Struct Biol*. 2011;21:382–90.
- Zhang J, Liang Y, Zhang Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure*. 2011;19:1784–95.

---

## Hybrid Circuits

- [Dynamic Clamp: Synthetic Conductances and Their Influence on Membrane Potential](#)

---

## Hybrid Quantum Mechanical/Molecular Mechanical Methods

- [QM/MM Methods](#)

---

## HYDFIT and Related Packages for Linear Molecules

Álvaro Ortega<sup>1</sup> and José García de la Torre<sup>2</sup>

<sup>1</sup>Departamento de Química Física, Facultad de Química, Universidad de Murcia, Murcia, Spain

<sup>2</sup>Faculty of Chemistry, Department of Physical Chemistry, University of Murcia, Spain

### Synonyms

[Hydrodynamics of macromolecules](#)

### Definition

HYDFIT is the global term for the different software packages for the fitting of multi-property and multi-sample sets of experimental data to simple models like rigid rods, disks or ellipsoids, or semiflexible wormlike chains.

## Introduction

The HYDFIT programs are intended for the determination of shape and size, or flexible parameters of macromolecules and nanoparticles, which is done by comparison of computational predictions with experimental data. Single-HYDFIT is intended for the analysis of several solution properties of a single macromolecule or particle, Multi-HYDFIT works with values of properties for a series of samples with varying molecular weight or length. These programs are straightforward for the determination of the optimum values of the model parameters that fit best a series of experimental values. The optimization is done by minimizing a target (error) function. The comparison between experimental and calculated quantities is made in terms of equivalent radii or ratios of radii which are most adequate for expressing the structure-properties relationship in a manner that is consistent and equilibrated, for the various properties and different macromolecular conformations. The equivalent radii depend on both size and shape, while the ratios of radii are only function of shape (dimensionless) indicators. Single-HYDFIT is oriented to the determination of structural parameters of ellipsoidal, cylindrical particles and the flexibility and structure of linear semiflexible chains, while Multi-HYDFIT focuses on macromolecules which can make a series of sizes or molecular weights, maintaining the same overall structural parameters, such as cylinders and wormlike chains.

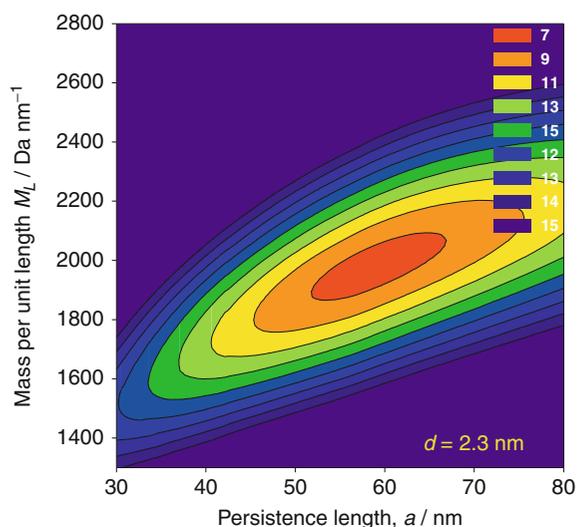
## Application to Linear Molecules. Calculation of Properties

The focus of this entry is centered in the application of HYDFIT programs to fully or semiflexible linear macromolecules. There is a variety of synthetic polymers and biological macromolecules which possesses a semiflexible skeleton, and can be properly represented by the ► [wormlike chain model](#). The model describes the whole spectrum of chains with different degrees of chain stiffness from rigid rods to random coils, and is particularly useful for representing stiff chains. In the literature this chain is sometimes referred to as a Porod–Kratky chain (McNaught and Wilkinson 1997).

The basics of HYDFIT falls on the comparison of the hydrodynamic properties calculated for a given structure or set of structural parameters and the experimental properties measured in the laboratory. About 40 years ago, Yamakawa and Fujii presented a theoretical treatment of the solution properties of the wormlike model (Yamakawa and Fujii 1973, 1974). Zimm and others (Zimm 1980; Hagerman and Zimm 1982), proposed to calculate hydrodynamic properties of flexible macromolecules by the Monte Carlo rigid-body method. In the rigid-body treatment, conformations of the chain are generated by a Monte Carlo procedure. For each conformation, the properties are individually computed using hydrodynamically rigorous procedures, and the final values are the conformational averages. This treatment has been shown to predict correctly experimental results for fully flexible random-coil polymers in both ideal and good solvent conditions (Garcia de la Torre et al. 1984). The latter is the methodology applied by HYDFIT, the procedure requires rigid-body hydrodynamic calculation which will be performed on bead models of the ► [wormlike chain](#). For such calculation advanced methodology embodied in bead-model hydrodynamics (Carrasco and García de la Torre 1999) is employed.

### Minimization Function. Equivalent Radii and Ratios of Radii

For the purpose of handling the results of the Monte Carlo simulation – in regard to both their dependence on structural parameters and their use in the determination of such parameters from experimental data – it is very useful the employment of the concepts of equivalent radii and ratios of radii (Ortega and García de la Torre 2007). The equivalent radii are the values of the radius of a spherical particle that would have a given value of some property, they do depend only on the size and conformation of the solute, which in a ► [wormlike chain](#) is defined by the contour length,  $L$ , the persistence length,  $a$  and the diameter,  $d$ . For the solution hydrodynamic or ► [light scattering properties](#), namely, the radius of gyration,  $R_G$ , the ► [translational diffusion](#),  $D_t$ , or equivalently, the ► [sedimentation coefficient](#),  $s$ , etc. the equivalent radii can be defined with straightforward equations. Some examples are shown below:

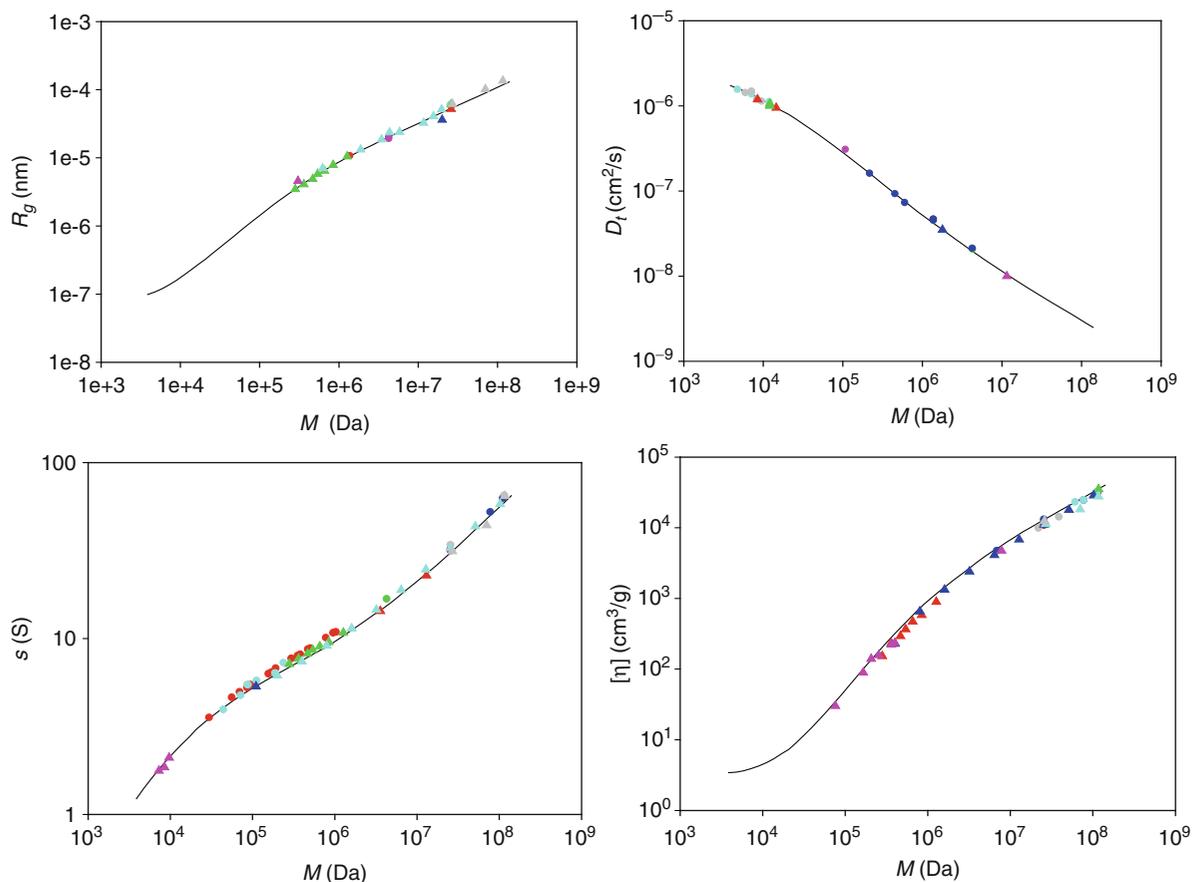


**HYDFIT and Related Packages for Linear Molecules, Fig. 1** HYDFIT contour plot of the percent typical deviation,  $100\Delta$ , with fixed  $d = 23 \text{ \AA}$ , allows to predict correct persistence length and  $M_L$  of the double-stranded DNA molecule

$$a_G = \sqrt{\frac{5}{3}}R_G; \quad a_T = \frac{k_B T}{6\pi\eta_0 D_t}; \quad a_I = \left(\frac{3M[\eta]}{10\pi N_A}\right)^{1/3}$$

The equivalent radii can be then combined into ratios of radii which do not depend on the absolute size of the particle but only on the shape or conformation, therefore being perfect candidates for the goal of finding a minimization function which accounts for all the properties in a weighted manner and sensitive enough to structural parameters, such as shape and flexibility of the molecule. In the case of wormlike chains,  $GT$ ,  $GI$ , and  $IT$  ratios are functions of the quantities that determine their conformation, which are the ratios of parameters,  $L/d$  and  $L/a$ , but not on the individual values of  $L$ ,  $a$ , and  $d$ . The notation  $XY$  defines the ratio of equivalent radii  $a_X/a_Y$ . With the results from the computer simulation for the Monte Carlo procedure, the ratios  $GT$ ,  $GI$ , and  $IT$  are obtained for any wormlike molecule with given values of  $L$ ,  $a$ , and  $d$ , then the equivalent radii  $a_T$ ,  $a_I$ , and  $a_G$  are extracted and finally  $D_t$  and  $[\eta]$  are obtained, the latter requiring the value of the molecular weight, or that of the relative parameter  $M_L$ .

The ultimate goal of HYDFIT for studying the ► [wormlike model](#) should be to provide a protocol to extract the structurally relevant information,



**HYDFIT and Related Packages for Linear Molecules, Fig. 2** Plots of the experimental values for some hydrodynamic properties of double-stranded DNA (filled symbols) and

calculated values from optimum parameters obtained by HYDFIT (line),  $a = 560 \text{ \AA}$ ,  $M_L = 195.0 \text{ Da/\AA}$ , and  $d = 23 \text{ \AA}$

contained in parameters like  $a$ ,  $M_L$ , and  $d$ , from experimental data. Usually, experimental values of the solution properties are available for a series of samples with varying molecular weight. In order to analyze simultaneously several properties of various samples, in a search for structural parameters, HYDFIT follows a strategy (Ortega and García de la Torre 2007), implemented in both the Single-HYDFIT and Multi-HYDFIT computer programs, which is based on equivalent radii.

$Y$  denotes any of the observable properties,  $R_G$ ,  $D_t$ ,  $s$ ,  $[\eta]$ , etc. As it has been just noted above, previous results allow the prediction of these properties as functions of molecular weight from given values of the parameters. In the HYDFIT strategy, the properties are transformed into equivalent radii,  $a_Y$ , which are considered functions of molecular weight, with

a dependence in turn on the model parameters. In the case of the ► wormlike chain, this can be expressed as  $a_Y(\text{cal}) (M; a, M_L, d)$ . Then, the HYDFIT procedure optimizes the parameters seeking the minimization of a target function such as:

$$\Delta^2(a, M_L, d) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[ \left( \sum_Y w_Y \right)^{-1} \sum_Y w_Y \left( \frac{a_Y(\text{cal}) - a_Y(\text{exp})}{a_Y(\text{exp})} \right)^2 \right]$$

where the outermost sum runs in the Multi-HYDFIT program over the  $N_s$  samples with different molecular weight and the innermost sum runs over the available properties for each sample. In the case of

Single-HYDFIT, the target function is the same containing only the innermost sums, as only one set of samples is analyzed. The advantage of using ratios of radii is that the  $a_Y$  are different but of the same order of magnitude, and in this way the properties are given a similar importance. Furthermore, the target function uses relative deviations,  $(a_Y(\text{cal}) - a_Y(\text{exp}))/a_Y(\text{exp})$ . The user of HYDFIT could optionally use different weights  $w_Y$  for the various properties for specific purposes – for instance, to weight the properties according to their experimental precision. For more details and examples of this strategy, see the original reference (Ortega and García de la Torre 2007).

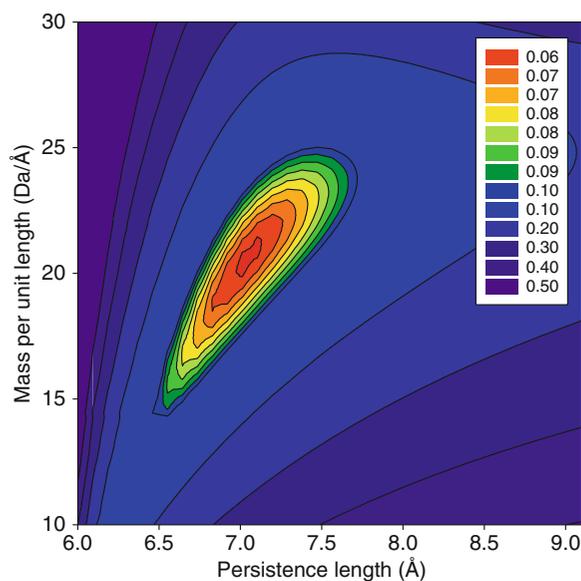
If the interest is the analysis of data for a single sample, there is also an alternative route based on the ratios of radii, which determines primarily the particle's shape, described by the parameter  $p$ , instead of the two sizes  $L$  and  $d$ ; this is particularly useful, for instance, in the cylinder-shaped particle's case. For this purpose another target function based on ratios is defined, and included in the program Single-HYDFIT:

$$\nabla^2 = \left( \sum_{XY} w_{XY} \right)^{-1} \sum_{XY} w_{XY} [XY(\text{cal}) - XY(\text{exp})]^2$$

## Examples of HYDFIT Calculations

To express the accuracy of HYDFIT software, some examples of the calculation of the parameters  $a$ ,  $M_L$ , and  $d$  of some synthetic and biological polymers are shown here. Results for the double-stranded DNA molecule, which is very well determined and whose experimental properties are available in large amount, are presented and, as an example of synthetic polymers, the calculations made for poly isobutylene (PIB), an extreme case of a very flexible chain are shown (Fig. 1).

DNA is the paradigmatic example of the ► [wormlike chain](#), its solution properties have been determined by many authors. To determine the mass per unit length ( $M_L$ ) and the persistence length,  $a$  (which is an estimation of its rigidity), HYDFIT is run with the input of 147 experimental property–molecular weight values. With one single execution of Multi-HYDFIT, results are obtained, in perfect agreement with the established values for this macromolecule. Figure 1 shows contour plots of the  $\Delta$  function, indicating the minimum of this function at  $M_L = 195.0 \pm 4.0 \text{ Da}/\text{\AA}$  and  $a = 560 \pm 20 \text{ \AA}$  (Fig. 2).



**HYDFIT and Related Packages for Linear Molecules, Fig. 3** HYDFIT contour plot of the percent typical deviation,  $100\Delta$ , of the poly isobutylene molecule with fixed  $d = 6.9 \text{ \AA}$ , showing the minimum of the deviation at values of persistence length,  $a = 7.1 \text{ \AA}$ , and  $M_L = 20.5 \text{ Da}/\text{\AA}$

To show the validity of the methodology implemented in HYDFIT for the whole range of rigidities, the calculation of the wormlike parameters of the PIB, a very flexible wormlike chain is also included, in contrast to the rigidity of the DNA molecule. As can be seen in Fig. 3, the parameters found for this molecule are  $M_L = 20.5 \text{ Da}/\text{\AA}$  and  $a = 7.1 \text{ \AA}$ , being  $d = 7.1 \text{ \AA}$ .

## Cross-References

- [Dynamic Light Scattering](#)
- [HYDRO Suite of Computer Programs for Solution Properties of Rigid Macromolecules](#)
- [Sedimentation Velocity Analytical Ultracentrifugation](#)
- [Worm-Like Chain \(WLC\) Model](#)
- [X-Ray Scattering of Lipid Membranes](#)

## References

- Carrasco B, García de la Torre J. Hydrodynamic properties of rigid particles. Comparison of different modelling and computational procedures. *Biophys J.* 1999;76:3044–3047.

- García de la Torre J, López Martínez MC, Tirado MM, Freire JJ. Monte Carlo study of hydrodynamic properties of flexible linear chains. Analysis of several approximate methods. *Macromolecules*. 1984;17:2715–2722.
- Hagerman P, Zimm BH. Monte Carlo approach to the analysis of the rotational diffusion of wormlike chains. *Macromolecules*. 1982;15:148–154.
- McNaught AD, Wilkinson A. *Compendium of chemical terminology*. The gold book. 2nd ed. London: Blackwell Science; 1997.
- Ortega A, García de la Torre J. Equivalent radii and ratios of radii from solution properties as indicators of macromolecular conformation, shape, and flexibility. *Biomacromolecules*. 2007;8:2464–2475.
- Yamakawa H, Fujii M. Translational friction coefficient of wormlike chains. *J Chem Phys*. 1973;6:408–414.
- Yamakawa H, Fujii M. Intrinsic viscosity of wormlike chains. Determination of the shift factor. *Macromolecules*. 1974;7:128–135.
- Zimm BH. Chain molecule hydrodynamics by the Monte-Carlo method and the validity of the Kirkwood-Riseman approximation. *Macromolecules*. 1980;13:592–602.

---

## Hydration

- ▶ [Terahertz Spectroscopy: Solvation of Biomolecules](#)
- ▶ [Vibrational Spectroscopy with Neutrons](#)
- ▶ [X-Ray Diffraction and Crystallography of Oligosaccharides and Polysaccharides](#)

---

## HYDRO Suite of Computer Programs for Solution Properties of Rigid Macromolecules

José García de la Torre<sup>1</sup> and Álvaro Ortega<sup>2</sup>

<sup>1</sup>Faculty of Chemistry, Department of Physical Chemistry, University of Murcia, Spain

<sup>2</sup>Departamento de Química Física, Facultad de Química, Universidad de Murcia, Murcia, Spain

### Synonyms

[Hydrodynamics of macromolecules](#)

### Definition

The HYDRO suite is a collection of computer tools for predicting hydrodynamic and other solution properties

from macromolecular structures, and to gather structural information from such properties.

A collection of free-domain, user-friendly, and well-documented computer programs to predict properties from the three-dimensional shape of rigid particles, with various level of resolution, from very coarse-grained models to atomic level.

## Introduction

Hydrodynamic properties, like translational and rotational diffusion, intrinsic viscosities, relaxation times, and other properties like those related to scattering of radiation, for example, the radius of gyration or the distribution of distances, are, in the case of rigid particles, directly determined from their size and shape, and are, therefore, potential sources of information about their overall structures. Thus, the validation of structures obtained by, say crystallography, microscopy, etc., against experimental measurements of such solution properties, or the determination of structural information by global fitting or optimization, requires the employment of such predictive computational tools.

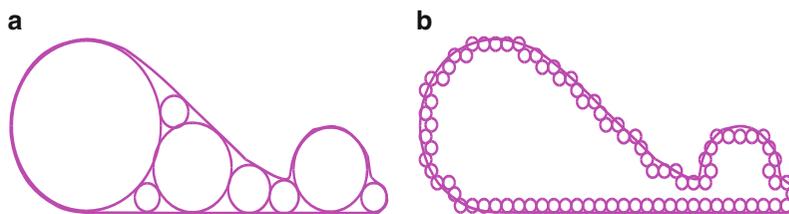
## Bead Models and Computer Programs

All the computational procedures in the HYDRO suite represent the structure of the rigid particle as a bead model, which is an array of  $N$  spheres such that its overall size and shape reproduces adequately those of the particle under consideration (See [Fig. 1a](#)). The reason for using beads as the building blocks is that the hydrodynamics of a spherical particle is simpler than that of any other shape. In an array of spherical elements, an important aspect is the hydrodynamic interaction (HI), which makes the friction experienced by any element dependent on those of the rest of them. The theoretical and computational treatment of HI is also simpler with spherical elements. The HI effect is the central aspect of the hydrodynamics, and it is remarkable that it determines the computing time, which increases remarkably with the number of beads, being proportional to  $N^3$ .

The idea of using bead models to represent macromolecules comes from classical works in polymer physics, which proposed to model chain macromolecules, either coiled because of flexibility, or those appearing as rigid straight rods, as strings of beads.

### HYDRO Suite of Computer Programs for Solution Properties of Rigid Macromolecules, Fig. 1

(a) Scheme of a bead model, in strict sense. (b) Scheme of a bead/shell model



The concept was adopted for modeling the complex shapes of rigid biological macromolecules by Bloomfield et al. (1967). (For a review see García de la Torre and Bloomfield 1981). In the early applications, models were constructed as agglomerates of a moderate number (such that the computing requirements were accessible by them) of non-overlapping spheres. The more advanced procedures are based on the bead/shell model, in which a large number of small “minibeads” are placed on the surface of the particle (see Fig. 1b).

### HYDRO++

This computer program implements the simplest bead model, such as that in Fig. 1a. A real example is shown in Fig. 2a. The simplicity of this kind of model allows for an extremely fast calculation of properties, but the initial versions of the computational procedures (program HYDRO) had some deficiencies regarding the prediction of rotational diffusion intrinsic viscosity. In the latest version, HYDRO++ (García de la Torre et al. 2007), these problems have been corrected. The user has to supply a main input file plus another file containing a list of the three coordinates and radii of the beads. This method is usually applied to low-resolution models, but the number of beads in the model is not limited, and it can be applied to any other instance like, for instance, elongated filamentous structures modeled as strings of beads.

### HYDROPRO and HYDRONMR

These programs (García de la Torre et al. 2000a, b) were intended for the calculation of solution properties of models with atomic resolution, starting from atomic coordinates contained in PDB-formatted files. In the primary hydrodynamic model, whose properties are to be calculated, each atom is replaced by a sphere. The radii of the elements are larger than the van der Waals

atomic radii, including a contribution due to hydration. From a global fit for a large set of proteins this radius has been estimated to be  $\cong 3.0 \text{ \AA}$ , which corresponds well to a typical van der Waals radius of  $\cong 1.8 \text{ \AA}$  plus a monolayer of hydration of about  $\cong 1.2 \text{ \AA}$ . Properties predicted by HYDROPRO with this setting agree well with experimental data, within statistical error. HYDRONMR predicts the correlation time  $\tau_c$  of macromolecular structures and, for globular proteins, the sequence of  $T1$  and  $T2$  relaxation times for each amino acid residue. An example showing the quality of the prediction is shown in Fig. 3.

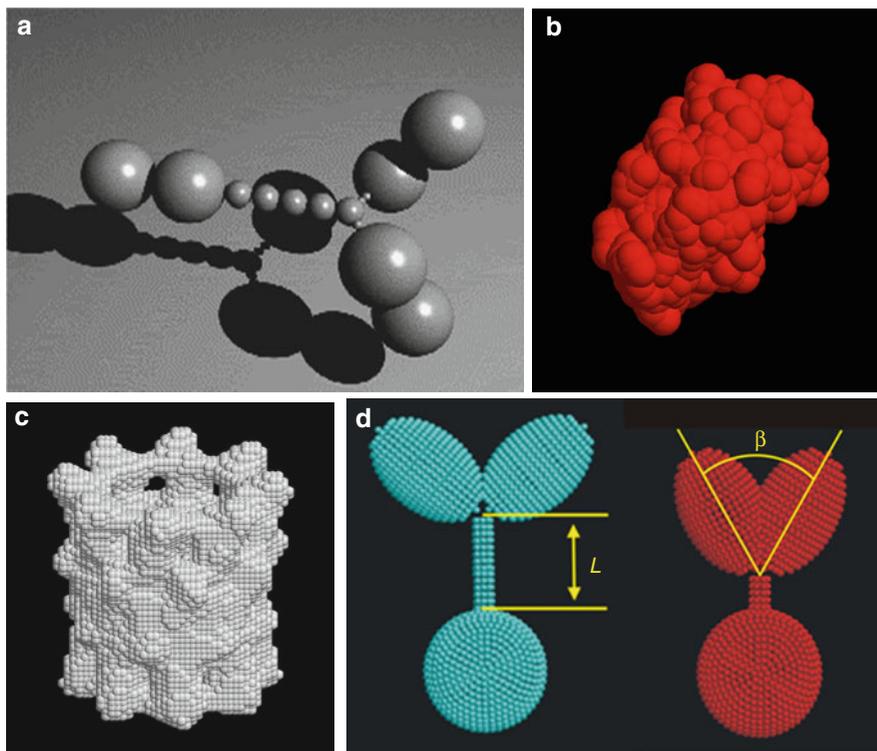
Although HYDROPRO was designed for atomic-resolution elements, it has recently been noticed (Frembgen and Elcock 2009; Ortega et al. 2011) that it works fairly well with primary elements based on more coarse-grained models in which each bead represents not an atom, but a full amino acid residue. In this case, the structural file would be a PDB-formatted list of the  $C\alpha$  coordinates, where the beads will be centered. The radius of the primary element should be in this case about  $\cong 5.1 \text{ \AA}$ . The quality of the prediction is practically as good as that of fully atomic model, which indicates that such level of detail, although available and of course useful for the HYDROPRO calculation, is not fully necessary. Yet another case of pushing the HYDROPRO calculation for an even lower resolution model is the possibility of applying this program for models derived from ab initio structural determination using ► [small-angle scattering](#), which provides a model of large “dummy pseudo-atoms” (Ackerman et al. 2003). For each individual structure, the radius of the elements in the primary mode has to be adjusted as to reproduce the hydrated volume estimated for the protein.

### HYDROMIC

While atomic- and residue-level structures can be routinely determined for moderately small proteins and oligonucleotides, much larger macromolecules

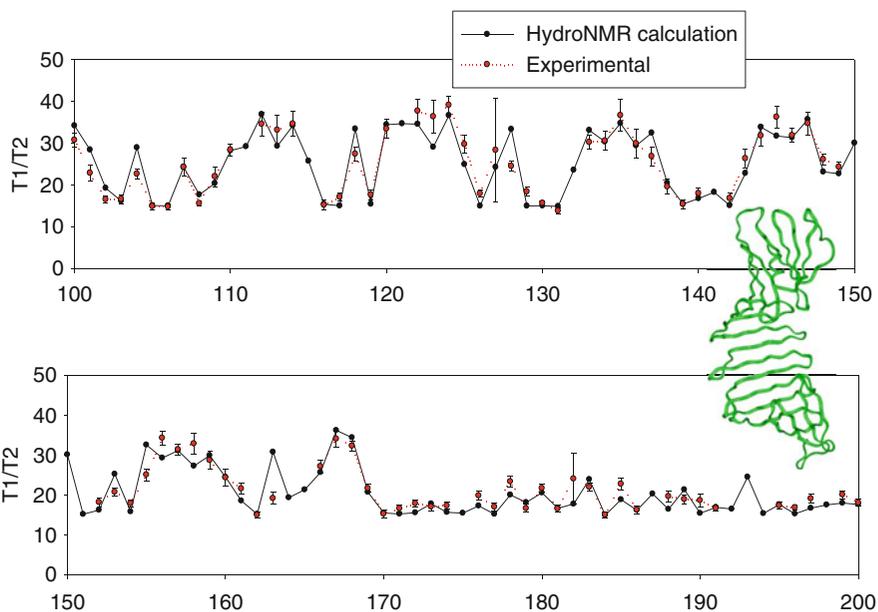
### HYDRO Suite of Computer Programs for Solution Properties of Rigid Macromolecules, Fig. 2

(a) HYDRO++ bead model for an IgG3 antibody molecule.  
 (b) HYDROPRO and HYDRONMR primary models for lysozyme  
 (c) HYDROMIC model for CCT chaperonin  
 (d) HYDROSUB model for wild type and mutant IgG3 antibodies



### HYDRO Suite of Computer Programs for Solution Properties of Rigid Macromolecules, Fig. 3

Sequence of the ratio of NMR relaxation times  $T_1/T_2$ , calculated by HYDRONMR for the outer surface protein OSP, for amino acids 100–200, showing the agreement between predictions and experimental results



and macromolecular complexes are investigated by ► [cryo-electron microscopy](#) and related techniques, which supply three-dimensional reconstructions of the particle in the form of an electron-density map (see [Fig. 2c](#)). In the same way as HYDROPRO works with

the atomic or residue coordinates in a PDB-formatted file, there is another program, HYDROMIC, which accepts the outcome of those microscopy methods for the prediction of solution properties. (For an example, see [García de la Torre et al. 2001](#))

## HYDROSUB

Large macromolecules and macromolecular complexes are often constituted by well-defined domains or subunits. Even if high-resolution information of those structures were available (which is a rare case, as these structures are not amenable to X-ray or ► [NMR](#) determination), the main structural aspect is the overall arrangement of these subunits. Modeling those subunits with simple shapes like ellipsoids or cylinders, allows the reduction of the structural description to such overall aspects, which may be those determining both the solution properties and the main features of their physiological function. This is the purpose of program HYDROSUB (García de la Torre and Carrasco [2002](#)), whose structural input file is a list of the sizes, positions, and orientations of the subunits. The size and shape of the individual can be obtained from the atomic-level structure, when it is available from crystallography (this is the origin of the “crystallohydrodynamics” term) or other higher-resolution model. Alternatively, the information on subunits can be gathered from if their experimental solution properties are available, and analyzed in terms of the classical theories for ellipsoids and rods. Then the HYDROSUB user would specify the Cartesian coordinates of the center of each subunit, and the two polar angles (only two are needed thanks to the axial symmetry of the subunits) that determine the orientation of the subunit in space, all with reference to any conveniently chosen system of Cartesian axes.

## Other Programs

This entry has concentrated on the HYDROxxx programs that perform the basic task of predicting properties from structure. The usual purpose of the users may go in the inverse direction: determination of structural information from solution properties. For such “inverse problem” we have devised strategies based on optimization protocols – the global fitting approach, described in a separate article in this Encyclopedia. We just mention the names of the tools for such purpose: ► [Single-HYDFIT](#), ► [Multi-HYDFIT](#), and ► [HYDROFIT](#). The latter is closely related to the HYDROxxx programs and other tools, generically named Multi-HYDROxxx, which facilitate the HYDROxxx calculations for a number of plausible

structures (eventually taking advantage of multi-core computing platforms), from which the optimum, best fitting structure can be selected by HYDROFIT.

This entry treats exclusively the tools for rigid particles. Flexible entities require different procedures that, in addition to hydrodynamics, consider their conformational variability. This is particularly important when considering not just overall coefficients like those of sedimentation or viscosity, but problems related to the internal dynamics of the flexible particle. Physical methods adequate for these problems are Monte Carlo and Brownian dynamics simulation. Computer programs for carrying out easily those simulations, for models quite similar to those of the rigid particles, are also available (García de la Torre et al. [2010](#)).

## Cross-References

- [Crystallohydrodynamics of IgG](#)
- [Dynamic Light Scattering](#)
- [Electron Microscopy](#)
- [HYDFIT and Related Packages for Linear Molecules](#)
- [NMR](#)
- [X-Ray Scattering of Lipid Membranes](#)

## References

- Ackerman CJ, Harnett MM, Harnett W, Kelly SM, Svergun DI, Byron O. 19 Å solution structure of the filarial nematode immunomodulatory protein, ES-62. *Biophys J.* 2003;84:489–500.
- Bloomfield VA, Dalton WO, van Holde KE. Frictional coefficients of multisubunit structures. *Biopolymers.* 1967;5:135–48.
- Frembgen-Kesner T, Elcock A. Striking effects of hydrodynamic interaction on the simulated diffusion and folding of proteins. *J Chem Theory Comput.* 2009;5:242–56.
- García de la Torre J, Bloomfield VA. Hydrodynamic properties of complex, rigid, biological macromolecules. Theory and applications. *Q Rev Biophys.* 1981;14:81–139.
- García de la Torre J, Carrasco B. Hydrodynamic properties of rigid macromolecules composed of ellipsoidal and cylindrical subunits. *Biopolymers.* 2002;63:163–7.
- García de la Torre J, del Rio Echenique G, Ortega A. Calculation of rotational diffusion and intrinsic viscosity of bead models for macromolecules and nanoparticles. *J Phys Chem B.* 2007;111:955–61.
- García de la Torre J, Huertas ML, Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J.* 2000a;78:719–30.

- García de la Torre J, Huertas ML, Carrasco B. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson.* 2000b;147:138–46.
- García de la Torre J, Llorca O, Carrascosa JL, Valpuesta JM. HYDROMIC: prediction of hydrodynamic properties of rigid macromolecular structures obtained from electron microscopy. *Eur Biophys J.* 2001;30:457–62.
- García de la Torre J, Ortega A, Amorós D, Rodríguez Schmidt R, Hernández Cifre JG. Methods and tools for the prediction of hydrodynamic coefficients and other solution properties of flexible macromolecules in solution. A tutorial minireview. *Macromol Biosci.* 2010;10:721–30.
- Ortega A, Amorós D, García de la Torre J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J.* 2011;101:892–98.

## Hydrodynamic Modeling of Carbohydrate Polymers

Gordon A. Morris<sup>1</sup> and Stephen E. Harding<sup>2</sup>

<sup>1</sup>Chemical and Biological Sciences, School of Applied Sciences, University of Huddersfield, Huddersfield, West Yorkshire, UK

<sup>2</sup>School of Biosciences, NCMH Laboratory, University of Nottingham, Sutton Bonington, Leicestershire, UK

## Synonyms

Hydrodynamics of macromolecules

## Introduction

Polysaccharides, mucins, and other large molecular weight glycoconjugates tend to exhibit quite different hydrodynamic properties compared with proteins. This is not only because of their general larger size (reaching to molar masses  $>50 \times 10^6$  g/mol) for some polysaccharides like amylopectin, their greater non-ideality (through molecular co-exclusion and charge effects – as represented by virial coefficients), and greater diversity of shapes and flexibilities. The primary structure of carbohydrate polymers is not coded by a genetic template; so they are also polydisperse (as represented by their molecular weight distribution or composition distribution). Nucleic acids are also

a class of glycoconjugate (poly-deoxyribose or poly-ribose backbones linked by phosphor-diester instead of glycosidic bonds).

This means that compared with proteins, the grammar of enquiry is somewhat different: We can use hydrodynamic methods to ascertain the molecular weight distribution (or the ratio of molar mass averages, such as  $M_z/M_w$ ), overall conformation type (between the extremes of globular/spherical, rod and random coil), flexibility (as manifested by the persistence length  $L_p$ ), and quaternary interactions (self-association or interactions with other macromolecules or small ligands).

## Representation of Solution Conformation

### The Haug Triangle

In the simplest of terms, the solution conformation of polysaccharides can be described as spheres, coils, or rods (Smidsrød and Andresen 1979). These three “extremes” of conformation are often represented as the points of the Haug triangle (Fig. 1).

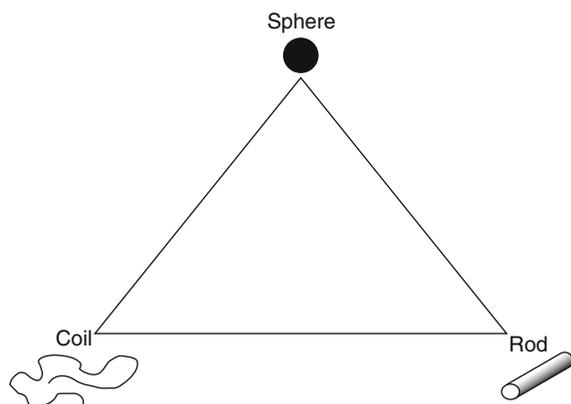
### Conformation Zoning (Normalized Scaling Relations)

An improvement on the Haug triangle is conformation zoning (or normalized scaling relations) which qualitatively estimates the conformation of polysaccharide based on a series of hydrodynamic measurements (Fig. 2a). The sedimentation conformation zoning (Pavlov et al. 1999) plot  $k_s M_L$  versus  $[s]/M_L$  enables an estimate of the “overall” solution conformation of a macromolecule in solution ranging from Zone A (rigid rod) to Zone E (globular or branched). The parameter  $[s]$  is related to the sedimentation coefficient by the relation

$$[s] = \frac{s^0_{20,w} \eta_{20,w}}{(1 - \bar{v} \rho_{20,w})} \quad (1)$$

$\rho_{20,w}$  and  $\eta_{20,w}$  are, respectively, the density and viscosity of water at 20.0°C and  $M_L$  is the mass per unit length

$$M_L = \frac{m}{l} \quad (2)$$



### Hydrodynamic Modeling of Carbohydrate Polymers,

**Fig. 1** Schematic representation of the “extremes” of polysaccharide conformation (Haug triangle), where most polysaccharide conformations will be described by a position on either the rod-coil or coil-sphere edges of the triangle

where  $m$  is the mass of an average monomer and  $l$  is the length of an average monomer ( $\sim 0.5$  nm).

Pavlov et al. (1999) described a further procedure to represent the conformation of polymers in solution based on the relationship between their molar mass, intrinsic viscosity, and mass per unit length,  $M_L$ , and has been recently applied to sugar beet pectin and its homogalacturonan (smooth) and rhamnogalacturonan (hairy) regions (Fig. 2b), and it has been demonstrated that the smooth regions tend to behave more like rods and the hairy regions more like flexible coils.

For a homologous series of polysaccharides of different molecular weights, the conformation can be estimated from the molecular weight dependency of a number of hydrodynamic parameters, for example, intrinsic viscosity ( $[\eta]$ ), sedimentation coefficient ( $s_{20,w}^0$ ), root-mean-square “radius of gyration” ( $r_g$ ), and the translational diffusion coefficient ( $D_{20,w}^0$ ) (Fig. 3a–d).

$$[\eta] = k' M^a \quad (3a)$$

$$s_{20,w}^0 = k'' M^b \quad (3b)$$

$$r_g = k''' M^c \quad (3c)$$

$$D_{20,w}^0 = k'''' M^{-\varepsilon} \quad (3d)$$

where the power law coefficients  $a$ ,  $b$ ,  $c$ , and  $\varepsilon$  depend on the polymer conformation (Table 1).

Fujita (1962) examined the possibility of transforming a distribution of sedimentation coefficients into a distribution of molecular weights for linear polymers and concluded that the sedimentation velocity determination of  $g(s)$  allows the evaluation of the molecular weight distribution of a given polymer if information about the relation between  $s$  and  $M$  is available from other sources (e.g., Eq. 3b). In Fujita (1962), the sedimentation coefficient  $s$  was given for the case of random coils  $s = k'' M^{0.5}$ . More generally  $s = k'' M^b$  where  $b = 0.4$ – $0.5$  for a coil,  $\sim 0.15$ – $0.2$  for a rod, and  $\sim 0.67$  for a sphere.

$$f(M) = g(s) \cdot b \cdot k''^{1/b} \cdot s^{(b-1)/b} \quad (4)$$

To do the transformation, the conformation type or  $b$  needs to be known and at least one pair of  $s$ - $M$  values is needed to define the  $k''$ . This method *only* applies to the infinite dilution or non-ideality free sedimentation coefficient distribution and therefore valid for values of  $s$  (or a distribution of  $s$  values) extrapolated to zero concentration or  $s$  values measured at low enough concentrations where non-ideality effects are small (Harding et al. 2011).

### Tsvetkov, Eskin, and Frenkel Relations

The validity of the MHKS parameters can be further explored by the calculation of their corresponding Tsvetkov, Eskin, and Frenkel (TEF) relations (Tsvetkov et al. 1970).

$$a = 2 - 3b \quad (5a)$$

$$b = 1 - c \quad (5b)$$

$$c = \frac{a + 1}{3} \quad (5c)$$

### The $\rho$ Parameter

A further estimate of molecular conformation can be obtained by the  $\rho$  parameter which has theoretical limits of 0.78, 1.7, and 2 for hard spheres, random coils ( $\theta$ -conditions), and rigid rods, respectively (Burchard 1992).

$$\rho = \frac{r_g}{r_H} \quad (6)$$

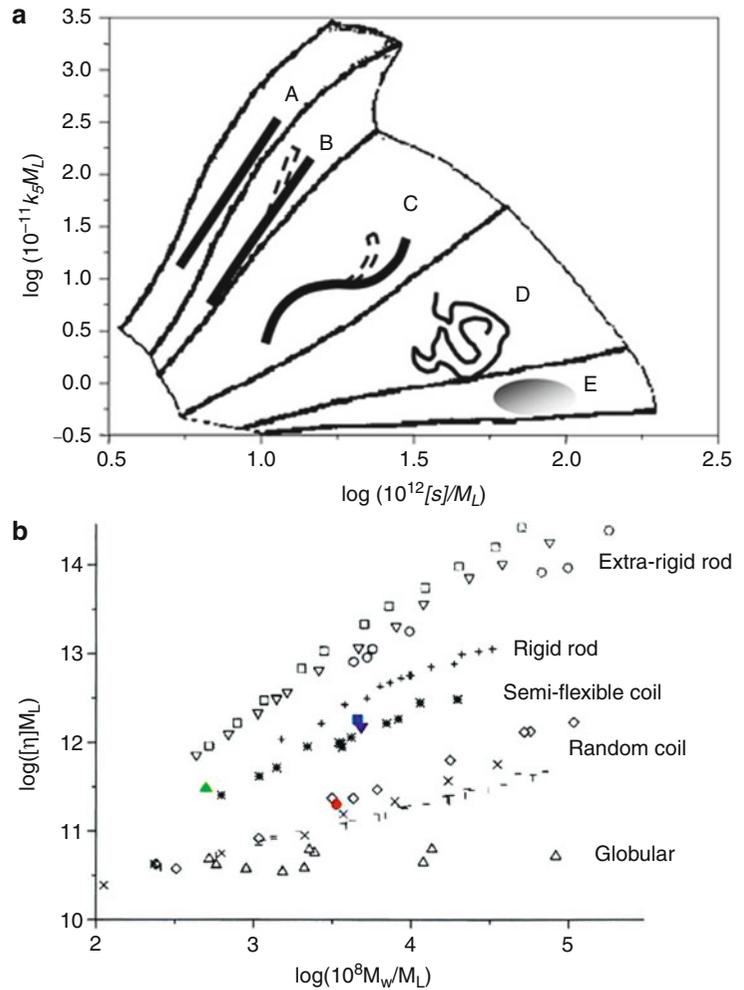
### Hydrodynamic Modeling of Carbohydrate Polymers, Fig. 2

(a) The sedimentation conformation zoning plot.

Zone A: rigid rod; Zone B: rod; Zone C: semiflexible coil; Zone D: random coil; Zone E: globular sphere (Adapted from Pavlov et al. 1999).

(b) Normalized scaling plot of  $[\eta]M_L$  versus  $M_w/M_L$  (Adapted from Pavlov et al. 1999; Morris et al. 2010)

where the solution conformations for pectin A (■) and its RG-I (●) and HG (▲) fractions are indicated, a typical citrus pectin is shown for comparison (▼). All other symbols are as defined previously (Pavlov et al. 1999): schizophyllan (□), DNA (○), globular proteins (Δ), xanthan, poly(1-vinyl-2-pyrrolidone) (◇), cellulose nitrate (+), pullulan (×), methyl cellulose (\*), poly- $\alpha$ -methylstyrene (-) and polystyrene (l)



### Translational Frictional Ratio and Perrin Function

The translational frictional ratio,  $f/f_0$ , is a parameter which depends on conformation and molecular expansion through hydration effects (Tanford 1961). It can be measured experimentally from the sedimentation coefficient, hydrodynamic radius, or translational diffusion coefficient and molecular weight:

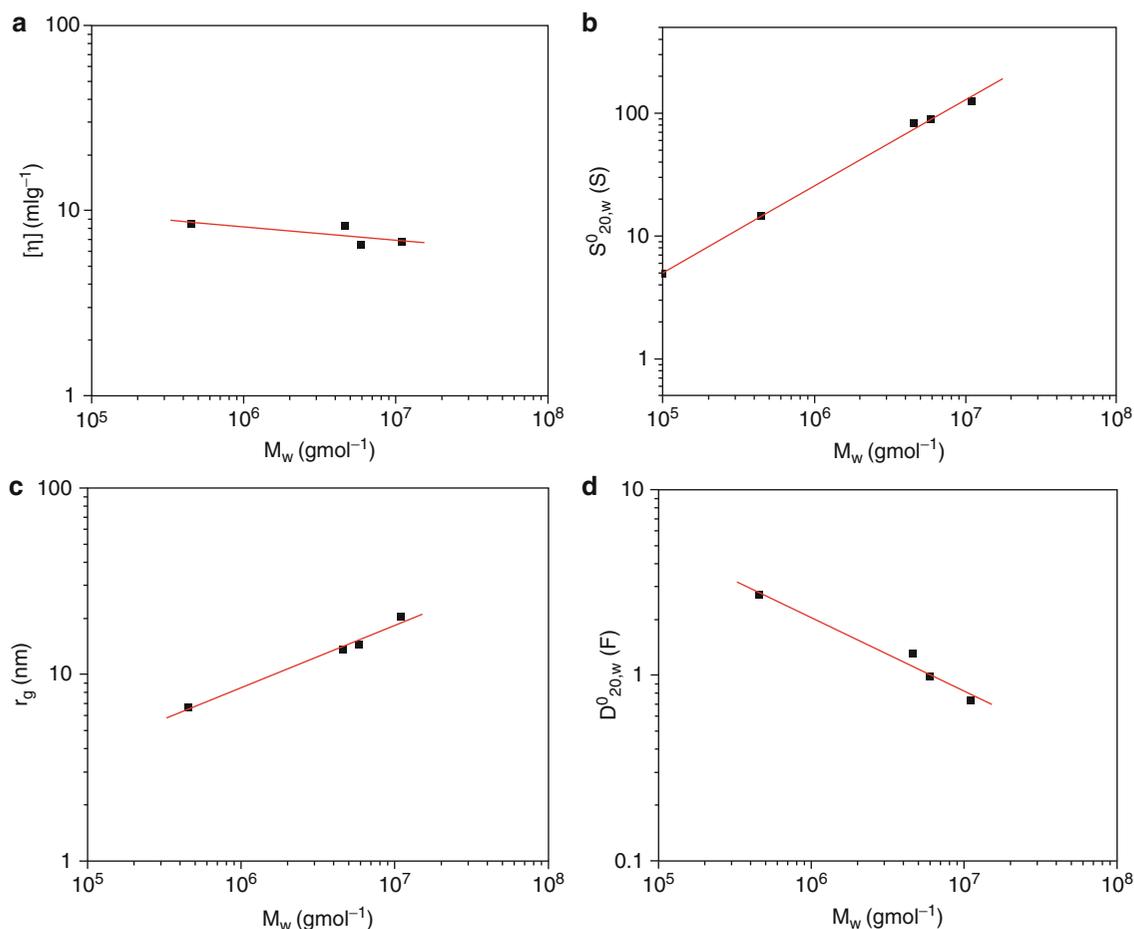
$$\frac{f}{f_0} = \frac{M_w(1 - \bar{v}\rho_{20,w})}{(N_A 6\pi\eta_{20,w} s_{0,20,w}^0)} \left( \frac{4\pi N_A}{3\bar{v}M_w} \right)^{\frac{1}{3}} \quad (7a)$$

$$\frac{f}{f_0} = r_H \left( \frac{4\pi N_A}{3\bar{v}M_w} \right)^{\frac{1}{3}} \quad (7b)$$

$$\frac{f}{f_0} = \frac{k_B T}{(6\pi\eta_{20,w} D_{20,w}^0)} \left( \frac{4\pi N_A}{3\bar{v}M_w} \right)^{\frac{1}{3}} \quad (7c)$$

where  $N_A$  is Avogadro's number and  $k_B$  is the Boltzmann constant.  $f$  is the friction coefficient of the molecule and  $f_0$  the corresponding value for a spherical particle of the same mass and (anhydrous) volume (Tanford 1961). It should therefore be noted that the translational frictional ratio has the same molar mass dependency as the hydrodynamic radius.

$$\frac{f}{f_0} = k'''' M_w^c \quad (7d)$$



**Hydrodynamic Modeling of Carbohydrate Polymers, Fig. 3** The Mark–Houwink–Kuhn–Sakurada (MHKS) plots for glycogen. The slopes of all four plots are consistent with a spherical (globular) conformation (Zone E). (a) the MHKS

viscosity plot ( $a = -0.07$ ); (b) the MHKS sedimentation plot ( $b = 0.71$ ); (c) the MHKS  $r_g$  plot ( $c = 0.33$ ); and (d) the MHKS diffusion plot ( $\epsilon = 0.40$ )

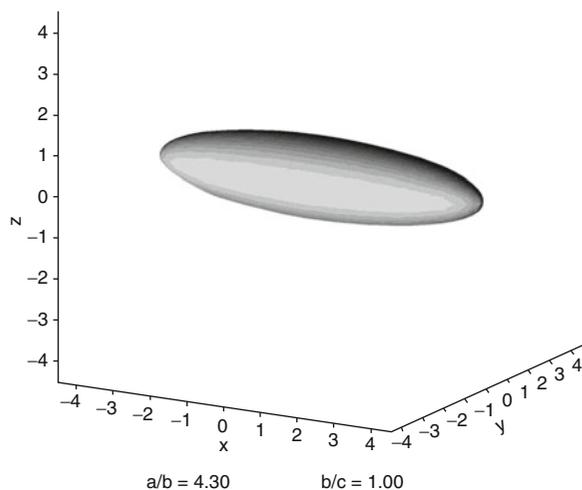
**Hydrodynamic Modeling of Carbohydrate Polymers, Table 1** The Mark–Houwink–Kuhn–Sakurada (MHKS) power law exponents ( $a$ ,  $b$ ,  $c$ , and  $\epsilon$ ), Wales–van Holde ( $k_s/[\eta]$ ) and  $\rho$  parameters for the five different conformations described by sedimentation conformation zoning (see Fig. 2a)

	Zone A	Zone B	Zone C	Zone D	Zone E
$a$	>1.4	0.8–1.4	0.5–0.8	0.2–0.5	0.0
$b$	<0.2	0.2–0.4	0.4–0.5	0.5–0.6	0.67
$c$	>0.8	0.6–0.8	0.5–0.6	0.4–0.5	0.33
$\epsilon$	>0.8	0.6–0.8	0.5–0.6	0.4–0.5	0.33
$k_s/[\eta]$	<0.2	0.2–0.4	0.4–1.0	1.0–1.4	1.6
$\rho = r_g/r_H$	>2.0	1.0–2.0			0.78–1.0

Knowledge of the hydration,  $\delta$ , (g or solvent per g of macromolecule) allows the estimation of the Perrin (frictional ratio due to shape) parameter,  $P$ .

$$P = \left( \frac{f}{f_0} \right) \left[ \frac{\bar{v}}{(\bar{v} + \delta)} \right]^{\frac{1}{3}} \quad (8)$$

The axial ratio ( $a/b$ ) can be calculated from the Perrin parameter (Fig. 4) using, for example, the ELLIPS1 routine (Harding 1997).



**Hydrodynamic Modeling of Carbohydrate Polymers, Fig. 4** Schematic representation of the solution conformation of glycogen (Zone E) in terms of prolate ellipsoids ( $x$ ,  $y$ , and  $z$  represent the orthogonal axes in which the ellipsoid lies, and  $a$ ,  $b$ , and  $c$  are ellipsoid semi-axes ( $a \geq b \geq c$ ) in the  $x$ ,  $y$ , and  $z$  directions with  $c = a$  for an oblate ellipsoid and  $c = b$  for a prolate ellipsoid). The axial ratio of  $\sim 4.3$  is calculated from  $P = 1.2$

### Wales–van Holde Ratio

The Wales–van Holde ratio,  $R$ , is a hydration-independent estimation of conformation (Wales and van Holde 1954).

$$R = \frac{k_s}{[\eta]} \quad (9)$$

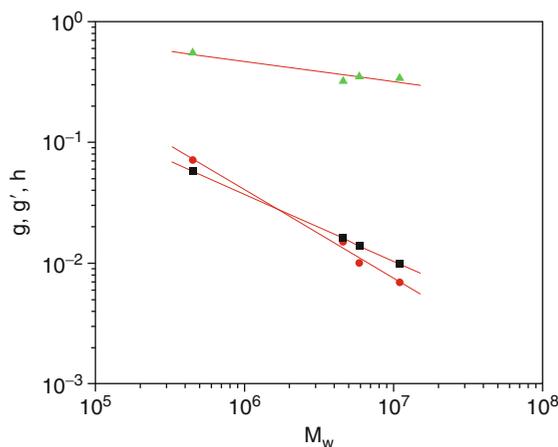
As with the Perrin function molecules, the axial ratio ( $a/b$ ) can be calculated from the Wales–van Holde ratio using, for example, the ELLIPS1 routine (Harding 1997).

### Shrinking Factors

The degree of branching of a hyper-branched macromolecule can be estimated from the shrinking factors (Burchard et al. 1980; Zimm and Stockmayer 1949),  $g$ ,  $g'$ , and  $h$  (Fig. 5).

$$g = \frac{r_{g,z,branched}^2}{r_{g,z,linear}^2} \quad (10a)$$

$$g' = \frac{[\eta]_{branched}}{[\eta]_{linear}} \quad (10b)$$



### Hydrodynamic Modeling of Carbohydrate Polymers, Fig. 5

The molar mass dependency of the shrinking factors  $g$  (■),  $g'$  (●) and  $h$  (▲) for glycogen. The slopes for  $g$ ,  $g'$ , and  $h$  are  $-0.55$ ,  $-0.73$ , and  $-0.17$ , respectively, which correspond to estimated degrees of branching of 40%, 9%, and 8.5%. **N.B.** the values for  $g$ ,  $g'$ , and  $h$  are very sensitive to the choice of model for the linear polymer of the same molar mass (in this case pullulan) and this may go some way to explaining the different estimations for the number of branches

$$h = \frac{f_{branched}}{f_{linear}} = \frac{(f/f_0)_{branched}}{(f/f_0)_{linear}} \quad (10c)$$

The molar mass dependency of  $g$ ,  $g'$ , and  $h$  allows an estimation of the number of branch points per molecule,  $n$ , and  $M_{BU}$  ( $= M_w/n$ ), the molar mass of the branching unit (Eq. 11a, b, and c),  $M_{BU}$  from which we calculate the number of monosaccharide residues per branch, for example,  $M_{BU}/162$  for glycogen.

$$g = \left[ \left(1 + \frac{n}{7}\right)^{0.5} + \frac{4n}{9\pi} \right]^b \quad (11a)$$

$$g' = \left[ \left(1 + \frac{n}{7}\right)^{0.5} + \frac{4n}{9\pi} \right]^{b'} \quad (11b)$$

$$h = \left[ \left(1 + \frac{n}{7}\right)^{0.5} + \frac{4n}{9\pi} \right]^{b''} \quad (11c)$$

where  $b$ ,  $b'$ , and  $b''$  are the slopes of the molar mass dependencies of  $g$ ,  $g'$ , and  $h$ , respectively.

### Draining Effects

The researcher needs to be aware of the possible contribution from intrachain solvent draining effects

where some solvent is conjectured to freely flow through the interstices between the atoms or segments of a macromolecule, although these effects are usually small compared to the dominant hydrodynamic interactions between such segments (Tanford 1961).

### Estimation of Persistence Length

The linear flexibility of polymer chains can also be represented quantitatively in terms of the persistence length,  $L_p$ , of equivalent *worm-like chains* where the persistence length is defined as the average projection length along the initial direction of the polymer chain. In the case of a theoretical perfect random coil,  $L_p = 0$ , and for the equivalent rigid rod (see Harding 1997)  $L_p = \infty$ , although in practice, limits of  $\sim 1$  nm for random coils (e.g., pullulan) and 200 nm for a rod (e.g., schizophyllan) are more appropriate (Tombs and Harding 1998).

### Burchard–Stockmayer–Fixman (BSF) Plot

This is perhaps the simplest way of estimating the persistence length. It involves plotting  $[\eta]/M_w^{1/2}$  versus  $M_w^{1/2}$  and the persistence length is calculated from the intercept,  $K_\theta$  (Stockmayer and Fixman 1963) (Fig. 6a).

$$K_\theta = \Phi \left( \frac{2L_p}{M_L} \right)^{3/2} \quad (12)$$

where  $\Phi$  is the Flory coefficient  $\sim 2.86 \times 10^{26} \text{ mol}^{-1}$ .

### Hearst Approach

The Hearst approach is similar to BSF and it involves the calculation of  $1/K_\theta$  from the slope of  $[\eta]/M_w$  versus  $M_w^{1/2}$  (Fig. 6b) Hearst (1963).

### Bushin–Bohdanecky Method

This is a popular method devised by Bushin et al. (1981) and Bohdanecky (1983) for estimating chain persistence lengths particularly for semiflexible polymers, and has been applied to a range of polysaccharides. In its simplest form, the Bushin–Bohdanecky

method involves plotting  $\left( \frac{M_w^2}{[\eta]} \right)^{1/3}$  versus  $M_w^{1/2}$  and from the slope  $L_p$  can be calculated using the following relation and tabulated values (Bohdanecky 1983) of the coefficient  $B_0$  (Fig. 6c):

$$\begin{aligned} \left( \frac{M_w^2}{[\eta]} \right)^{1/3} &= A_0 M_L \Phi^{-1/3} \\ &+ B_0 \Phi^{-1/3} \left( \frac{2L_p}{M_L} \right)^{-1/2} M_w^{1/2} \quad (13) \end{aligned}$$

### Yamakawa–Fujii Equation

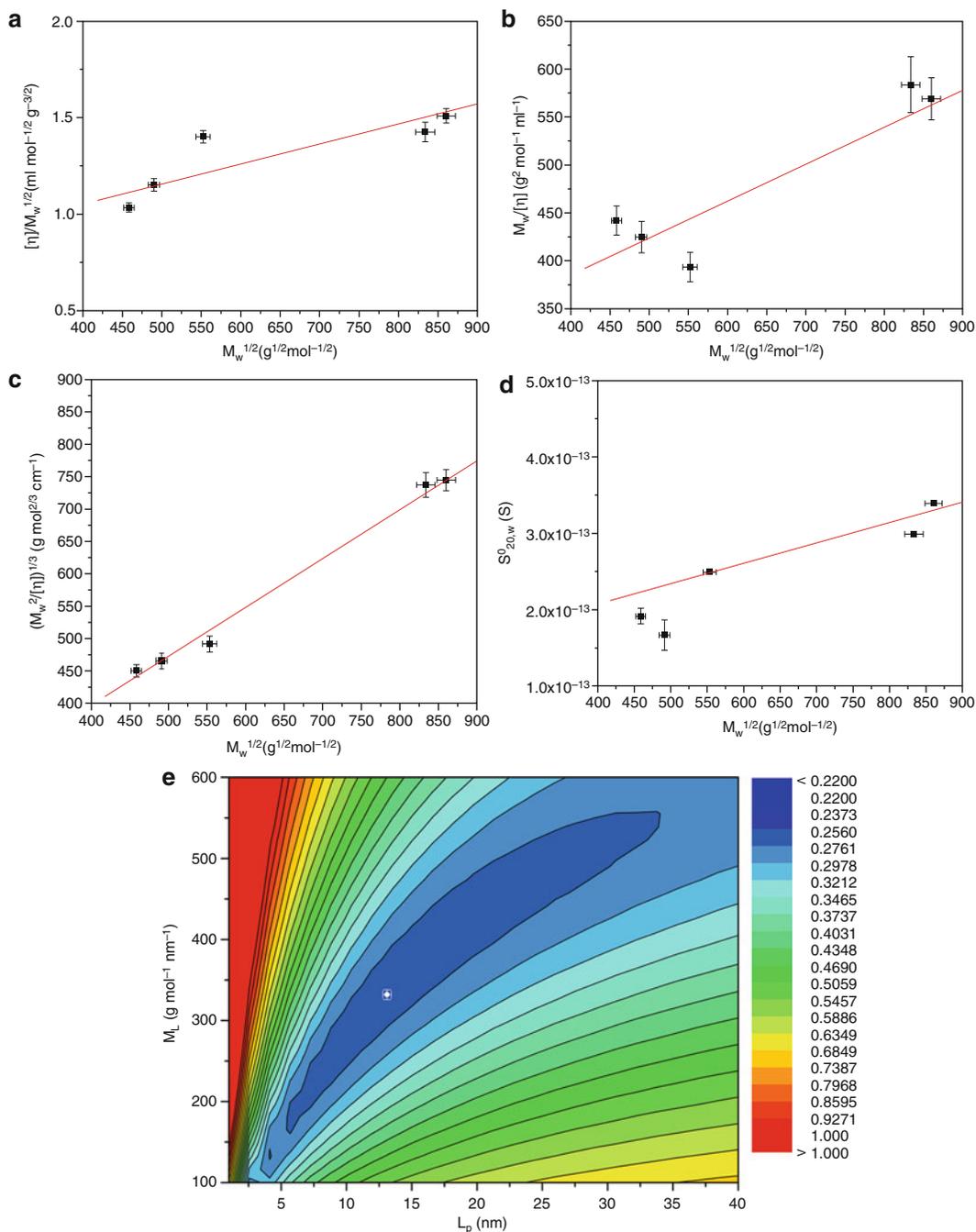
Hearst and Stockmayer (1962) first reported the sedimentation coefficient in relation to worm-like chain parameters, later refined by Yamakawa and Fujii (1973). The original relation given by Yamakawa and Fujii relating the sedimentation coefficient:

$$\begin{aligned} s^0 &= \frac{M_L(1 - \bar{v}\rho_0)}{3\pi\eta_0 N_A} \\ &\times \left[ 1.843 \left( \frac{M_w}{2M_L L_p} \right)^{1/2} + A_2 + A_3 \left( \frac{M_w}{2M_L L_p} \right)^{-1/2} + \dots \right] \quad (14) \end{aligned}$$

Yamakawa and Fujii (1973) showed that  $A_2 = \ln(d/2L_p)$  and  $A_3 = 0.1382$  if the  $L_p$  is much larger than the chain diameter,  $d$ . The persistence length can then be estimated from the slope of  $s_{20,w}^0$  versus  $M_w^{1/2}$  (Fig. 6d).

### The HYDFIT Algorithm

The way these approaches are implemented can lead to a significant variability in the results, that is, contrary to expectation,  $L_p$  is model dependent (Ortega and Torre 2007). This is ably demonstrated by the different persistence lengths calculated by the Burchard–Stockmayer–Fixman, Hearst, Bushin–Bohdanecky, and Yamakawa–Fujii approaches (Fig. 6a–d). The persistence length and mass per unit length can be estimated using Multi-HYDFIT program which considers data sets of hydrodynamic parameters for different molecular weights. It then performs a minimization procedure, finding the best values of  $M_L$  and  $L_p$  and chain diameter  $d$  satisfying the Bushin–Bohdanecky (Bushin et al. 1981; Bohdanecky 1983) and Yamakawa and Fujii (1973) equations (Eqs. 13 and 14). The Multi-HYDFIT program then “floats” the variable parameters in order to find a minimum of the multi-sample target (error) function,  $\Delta$  (Ortega and Torre 2007) (Fig. 6e). In this procedure,  $\Delta$  is calculated using equivalent radii, where the equivalent radius ( $a_x$ ) is



### Hydrodynamic Modeling of Carbohydrate Polymers,

**Fig. 6** The estimation of the persistence length,  $L_p$ , for konjac glucomannan (Zone C) using different approaches. (a) BSF plot where  $L_p = 3 \text{ nm}$  from the intercept. (b) Hearst plot where  $L_p = 1 \text{ nm}$  from the slope. (c) Bushin-Bohdanecky plot where  $L_p = 8 \text{ nm}$  from the slope. (d) Yamakawa-Fujii plot where  $L_p = 34 \text{ nm}$  from the slope. (e) Solutions to the Bushin-Bohdanecky and Yamakawa-Fujii equations using equivalent

radii approach. The x-axis and y-axis represent  $L_p$  (nm) and  $M_L$  ( $\text{g mol}^{-1} \text{ nm}^{-1}$ ), respectively. The target function,  $\Delta$ , is calculated over a range of values for  $M_L$  and  $L_p$ . In these representations, the values of  $\Delta$  function are represented by the full color spectrum, from the minimum in the target function in blue ( $\Delta = 0.220$ ) to red ( $\Delta \geq 1$ ). The calculated minimum ( $L_p = 13 \text{ nm}$  and  $M_L = 330 \text{ g mol}^{-1} \text{ nm}^{-1}$ ) is indicated

**Hydrodynamic Modeling of Carbohydrate Polymers, Table 2** Representative dilute solution conformation parameters of selected glycopolymers

	<i>a</i>	<i>b</i>	<i>c</i>	$\epsilon$	$k_p/[\eta]$	$\rho$	$ff_0$	$L_p$ (nm)	Zone
Alginate	0.73–1.31	–	0.52–0.54	–	0.6	–	9	12–15	B/C
Capsular polysaccharides	0.4–0.9	–	–	–	0.8–1.6	–	3–9	2–20	C/D
Chitosan	0.77–1.10	0.24–0.25	0.55–0.56	–	0.16–0.73	–	11–16	4–35	B/C
Galactomannan	0.70–0.77	0.12–0.65	0.54–0.57	–	0.15–0.41	–	8–17	2–12	C
Glycogen	–0.07–0	0.71	0.31–0.33	0.38–0.40	–	0.7–1.0	1.7–2.8	–	E
Heparin	0.90	0.38	0.38	0.62	1.04–2.98	1.34–1.52	1–3	4–6	C
$\kappa$ -carrageenan	0.67–0.90	–	0.68	–	0.39–0.9	–	7–9	2–3	B/C
$\iota$ -carrageenan	0.77	–	0.68	–	0.16	–	5	4	B/C
Konjac glucomannan	0.74–0.78	0.32	–	–	0.4	–	9–14	1–34	C
Methyl cellulose	0.83	0.39	–	–	0.30–0.75	–	10–12	10–17	C
Mucin	0.56–1.1	0.4	0.31–0.69	–	1.5	1.7–2.4	–	–	A/B/C
Pectin	0.62–0.94	0.17	0.57	–	0.10–0.85	0.6–1.0	7–10	10–15	A/B/C
Pullulan	0.66–0.67	0.45	0.58	0.51	1.27–1.49	1.40–1.66	2–5	1–2	D
Xanthan	1.23	0.26	1.00	–	0.28	–	14–19	100–150	A/B
Xyloglucan	0.55–0.67	0.42	0.51	–	0.12–1.44	–	2–6	4–15	C/D

defined as the radius of an equivalent sphere having the same value as the determined property, for example, intrinsic viscosity ( $a_1$ ).  $\Delta$  is thus a dimensionless estimate of the agreement between the theoretical calculated values for the selected hydrodynamic parameter of a particular molar mass, persistence length, and mass per unit length and the experimentally measured parameter.

## Summary

The solution conformation of polysaccharides, mucins, and glycopolymers can be estimated in a variety of ways, and estimates for a number of commercially important biomacromolecules are summarized in Table 2. We find that glycopolymers cross the whole range of conformations in solution (Zones A–E), but that Zone C (semiflexible coil) appears the most common. As we can see from Fig. 6, the estimation of conformation (especially in terms of persistence length) is very sensitive to the choice of model and it is therefore our belief that when trying to estimate solution conformation of polysaccharides (or any other flexible macromolecule), the quality of the estimate is determined by the amount of experimental data available – a combination of hydrodynamic data is always desirable.

## Cross-References

- ▶ [Dynamic Light Scattering](#)
- ▶ [HYDFIT and Related Packages for Linear Molecules](#)
- ▶ [Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules](#)
- ▶ [Multiangle Light Scattering from Separated Samples \(MALS with SEC or FFF\)](#)
- ▶ [Sedimentation Equilibrium Analytical Ultracentrifugation](#)
- ▶ [Sedimentation Velocity Analytical Ultracentrifugation](#)

## References

- Bohdanecky M. New method for estimating the parameters of the wormlike chain model from the intrinsic viscosity of stiff-chain polymers. *Macromolecules*. 1983;16:1483–93.
- Burchard W. Static and dynamic light scattering approaches to structure determination of biopolymers. In: Harding SE, Sattelle DB, Bloomfield VA, editors. *Laser light scattering in biochemistry*. Cambridge: Royal Society of Chemistry; 1992. p. 3–22.
- Burchard W, Schmitt M, Stockmayer WH. Influence of hydrodynamic pre-averaging on quasi-elastic scattering from flexible linear and star-branched macromolecules. *Macromolecules*. 1980;13:1265–72.
- Bushin SV, Tsvetkov VN, Lysenko YB, Emel'yanov VN. Conformational properties and rigidity of molecules of

- ladder polyphenylsiloxane in solutions according to the data of sedimentation-diffusion analysis and viscometry. *Vysokomol Soedin A*. 1981;23:2494–503.
- Fujita H. *Mathematical theory of sedimentation analysis*. New York: Academic; 1962. p. 182–92.
- Harding SE. The intrinsic viscosity of biological macromolecules. Progress in measurement, interpretation and application to structure in dilute solution. *Prog Biophys Mol Biol*. 1997;68:207–62.
- Harding SE, Schuck P, Abdelhameed AS, Adams G, Kok MS, Morris GA. Extended Fujita approach to the molecular weight distribution of polysaccharides and other polymeric systems. *Methods*. 2011;54:136–44.
- Hearst JE. Rotatory diffusion constants of stiff-chain macromolecules. *J Chem Phys*. 1963;38:1062–5.
- Hearst JE, Stockmayer WH. Sedimentation constants of broken chains and wormlike coils. *J Chem Phys*. 1962;37:1425–33.
- Morris GA, Ralet M-C, Bonnin E, Thibault J-F, Harding SE. Physical characterisation of the rhamnogalacturonan and homogalacturonan fractions of sugar beet (*Beta vulgaris*) pectin. *Carbohydr Polym*. 2010;82:1161–7.
- Ortega A, de la Torre JG. Equivalent radii and ratios of radii from solution properties as indicators of macromolecular conformation, shape, and flexibility. *Biomacromolecules*. 2007;8:2464–75.
- Pavlov GM, Harding SE, Rowe AJ. Normalized scaling relations as a natural classification of linear macromolecules according to size. *Prog Colloid Polym Sci*. 1999;113:76–80.
- Smidsrød O, Andresen I-L. *Biopolymerkjemi*. Trondheim: Tapir Press; 1979.
- Stockmayer WH, Fixman MJ. On the estimation of unperturbed dimensions from intrinsic viscosities. *J Polym Sci C*. 1963;1:137–41.
- Tanford C. *Physical chemistry of macromolecules*. New York: Wiley; 1961.
- Tombs MP, Harding SE. *An introduction to polysaccharide biotechnology*. London: Taylor and Francis; 1998. p. 14–20.
- Tsvetkov VN, Eskin V, Frenkel S. *Structure of macromolecules in solution*. London: Butterworths; 1970.
- Wales M, van Holde KE. The concentration dependence of the sedimentation constants of flexible macromolecules. *J Polym Sci*. 1954;14:81–6.
- Yamakawa H, Fujii M. Translational friction coefficient of wormlike chains. *Macromolecules*. 1973;6:407–15.
- Zimm BH, Stockmayer WH. The dimensions of chain molecules containing branches and rings. *J Chem Phys*. 1949;17:1301–14.

- ▶ [HYDRO Suite of Computer Programs for Solution Properties of Rigid Macromolecules](#)
- ▶ [Hydrodynamic Modeling of Carbohydrate Polymers](#)
- ▶ [US-SOMO: Methods for Construction and Hydration of Macromolecular Hydrodynamic Models](#)

---

## Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules

Georges M. Pavlov

Institute of Macromolecular Compounds, Russian Academy of Sciences, St. Petersburg, Russia  
Institute of Physics, St. Petersburg State University, St. Petersburg, Russia

### Introduction

The hydrodynamic assessment of the overall conformation of macromolecules in solution using normalized scaling relations will be reviewed in this entry. Macromolecules in infinitely dilute solution occupy a volume in which the distribution of the polymer matter (substance) is determined by the chemical structure of the repeat unit, by the architecture/topology of macromolecule, and by the interaction between monomers and solvent molecules. An individual macromolecule is regarded as a fractal object (cluster of connected repeat units, connected monomer clusters), which may be represented by the actual fractal dimension (de Gennes 1979; Mandelbrot 1982). This dimension  $d_f$  may be determined from the dependence of the number (N) of single particle clusters on the distance  $r$  for which the number calculated for identical particles (the repeat unit) N is directly proportional to the mass M:

$$N \sim M \sim r^{df} \quad (1)$$

---

## Hydrodynamics of Macromolecules

- ▶ [Crystallohydrodynamics of IgG](#)
- ▶ [Electro-optics and Macromolecular Hydrodynamics](#)
- ▶ [Ellips and Covol](#)
- ▶ [Intrinsic Viscosity](#)
- ▶ [HYDFIT and Related Packages for Linear Molecules](#)

More general characteristic dimensions of a linear polymer chain may also be represented in terms of the radius of gyration or the end-to-end distance. The *fractal dimension* characterizes the distribution of polymer matter in the volume occupied by a macromolecule. The value of this volume depends on the

ability of a linear macromolecule to coil, on the cross-sectional diameter of the chain, and on the nature of the interactions between polymer matter with the solvent molecules.

Information about macromolecular size may be obtained from hydrodynamic studies and/or by the study of scattered electromagnetic radiation or neutron scattering (Tanford 1961; Tsvetkov et al. 1970; Cantor and Schimmel 1980; Benoit and Higgins 1992) – see also ► [HYDFIT and Related Packages for Linear Molecules](#); ► [X-Ray Scattering of Lipid Membranes](#); ► [Neutron Scattering of Membranes](#) – but this paragraph is focused on *interpretation* of the hydrodynamic data.

Highly useful hydrodynamic values are the translational diffusion coefficient ( $D_o$ ), the velocity sedimentation coefficient ( $s_o$ ) – where the subscript “o” indicates the value at infinite dilution – and the intrinsic viscosity ( $[\eta]$ ); they are related to the size or dimensions ( $D_o$ ,  $s_o$ , and  $[\eta]$ ) and the molecular weight or molar mass ( $s_o$  and  $[\eta]$ ) of a macromolecule (Tanford 1961; Tsvetkov et al. 1970; Yamakawa 1971; Cantor and Schimmel 1980; Fujita 1990):

$$D_o = kT/P\eta_0 \langle h^2 \rangle^{1/2} \quad (2)$$

$$s_o = M(1 - \rho_0 v)/N_A P\eta_0 \langle h^2 \rangle^{1/2} \quad (3)$$

$$[\eta] = \Phi \langle h^2 \rangle^{3/2}/M \quad (4)$$

In (2) (3), (4),  $\langle h^2 \rangle$  is the average square of the end-to-end distance,  $N_A$  is Avogadro’s number,  $\Phi(L/A, d/A, \epsilon)$  and  $P(L/A, d/A, \epsilon)$  are the Flory hydrodynamic parameters,  $L$  is the chain contour length,  $d$  is the diameter of the chain,  $A$  is the statistical segment length ( $= 2L_p$  where  $L_p$  is the “persistence length”), and  $\epsilon$  is a parameter characterizing the thermodynamic “quality” of a polymer-solvent system.

To characterize a particular macromolecular component, the experimental values must be extrapolated to zero polymer concentration  $c$  (g/ml), using, for example, the following relations:

$$D = D_o(1 + k_{Dc} + k_{D'}c^2 + \dots), \quad (5)$$

$$s = s_o(1 + k_{sc} + k_{s'}c^2 + \dots), \text{ and} \quad (6)$$

$$\eta = \eta_o(1 + k_{\eta c} + k_{\eta'}c^2 + \dots). \quad (7)$$

Equation (7) easily transforms into the classical Huggins equation  $(\eta_r - 1)/c = k_{\eta} + k_{\eta'}c + \dots$  and by definition  $k_{\eta} \equiv [\eta]$  when  $c \rightarrow 0$ , where  $\eta_r = \eta/\eta_o$  is the relative viscosity of solution (see ► [Intrinsic viscosity](#)). The values of  $k_{D}$  and  $k_s$  have the same dimensions or units (ml/g) as the  $k_{\eta} \equiv [\eta]$  value. Often, the concentration dependence parameter of the sedimentation coefficient  $k_s$  – also called the Gralen coefficient – is compared with  $[\eta]$  giving the dimensionless ratio  $k_s/[\eta]$ . Finally for  $k_s$ , a relation similar to (4) may be considered (Pavlov and Frenkel 1995):

$$k_s = B \langle h^2 \rangle^{3/2}/M, \quad (8)$$

where  $B$  is a parameter with dimension or units  $\text{mol}^{-1}$ .

The coefficient  $k_{D}$  depends in general both on  $k_s$  and on the second virial coefficient  $A_2$  (Tsvetkov et al. 1970):

$$k_{D} = k_s + v - 2A_2M, \quad (9)$$

where  $\bar{v}$  is the partial specific volume of polymer. To eliminate the common solvent properties, the intrinsic values of velocity sedimentation coefficient  $[s]$  and translational diffusion coefficient  $[D]$  prove practical to use:  $[s] \equiv s_o\eta_o/(1 - \bar{v}\rho_o)$  and  $[D] \equiv D_o\eta_o/T$ . These “intrinsic” parameters, like the intrinsic viscosity, depend only on the main characteristics of a macromolecular component.

In polymer science, Mark-Kuhn-Houwink-Sakurada (MHKS) type or hydrodynamic scaling relationships equations are widely used (Fujita 1990; Brandrup et al. 1999). The comparison of hydrodynamic characteristics with molar mass allows obtaining canonical MHKS relationships:

$$P_i = K_i M^{b_i}, \quad (10)$$

where  $P_i$  represents one of the hydrodynamic characteristics  $[\eta]$ ,  $D_o$ ,  $s_o$ , and  $k_s$ . The following correlations exist in the polymer homologous series among scaling indices  $b_i$  (Tsvetkov et al. 1970; Pavlov et al. 1990):

$$|b_D| = (1 + b_{\eta})/3, |b_D| + b_s = 1 \text{ and} \quad (11)$$

$$b_{ks} = (2 - 3b_s)/b_s \quad (12)$$

For instance, between  $s_o$  and  $M$  and also between  $k_s$  and  $s_o$ , there are the relations  $s_o = K_s M^{b_s}$  and  $k_s = K_{k_s} s_o^{b_{k_s}}$ . The indices  $b_s$  and  $b_{k_s}$  are linked by (12). These relations follow from the comparison of the corresponding fundamental (2, 3, 4, 8) with (10) and the supposition that the sizes of macromolecules are the same value under the different experimental conditions (velocity sedimentation, translational diffusion, and viscous flow). It should be noted however that these relations are meaningful only for a homologous series of polymers with different conformational status. For linear macromolecules, “homology” may be defined as a series of macromolecules which have the same mass per unit length  $M_L$  for different chain contour lengths  $L$ . The mass per unit length  $M_L$  is also called a *shift factor* (Yamakawa 1971) or *linear density of the polymer chain* as defined by  $M_L = M/L = M_o/\lambda$  with  $M$  the molar mass,  $M_o$  the molar mass of the repeat unit, and  $\lambda$  a projection of repeat unit to the main direction of macromolecular chain. This value may be determined from X-ray scattering data, or be calculated in the simple case of aliphatic chains. For the very rigid polymer chains, the estimation of this value from the hydrodynamic data is possible using the model of a weakly bending rod or a cylinder (Tsvetkov 1989). This value may change in the following limits:  $3 \leq 10^{-9} \times M_L \text{ g.mol}^{-1} \text{ cm}^{-1} \leq 22$ . A low value corresponds, for example, to flexible chains (polystyrene, pullulan), and a high value corresponds, for example, to the triple helical polysaccharide from the *Schizophyllum commune*.

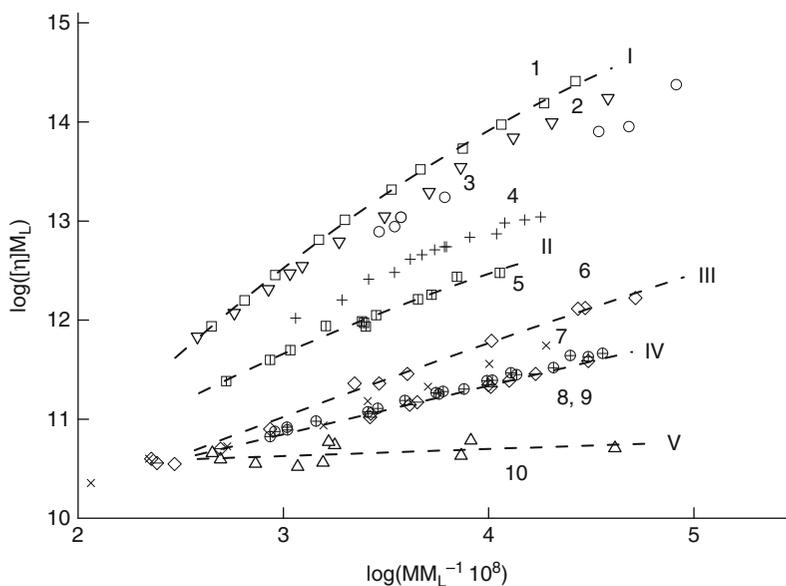
### First Normalization of the MKHS Equation for Linear Density of the Polymer Chain $M_L$

Mark-Kuhn-Houwink-Sakurada equation can be normalized to take into consideration the linear density or mass per unit length of the polymer chain  $M_L$ . In fact, a library of hydrodynamic data values for different types of linear polymer systems (including also their molar masses determined by one of various absolute methods – light scattering, sedimentation-diffusion analysis, or sedimentation equilibrium) has been collected over the last 50–60 years: from all this data, it is possible to make some representative generalizations. For instance, the data of  $[\eta]$  values cover the range of contour lengths of linear chains varying by three orders

of magnitude and the range of statistical segment lengths varying approximately 500 times. There are also plentiful data for the translational frictional properties of polymer systems in dilute solution with regards the velocity sedimentation and translational diffusion coefficients (Brandrup et al. 1999; Pavlov 2007; Pavlov et al. 1997, 1999) and covering both natural and synthetic polymers.

Polymer systems may be tentatively divided into several classes: globular (such as globular proteins and glycogen), flexible chain under  $\theta$  conditions (polystyrene (Pst), poly- $\alpha$ -methylstyrene (P $\alpha$ mst), and polymethylmetacrylate (PMMA)), flexible chain in thermodynamically good solvents (Pst, P $\alpha$ mst, PMMA, polyisobutylene, pullulan, and polyvinylpyrrolidone), rigid chains (cellulose derivatives), and extra-rigid chain polymers (xanthan, schizophyllan polysaccharides, and DNA). Data also used in this analysis were obtained for comb-like polymers and copolymers (molecular brushes) showing a large chain diameter.

Data analysis can be carried out using plots of the MHKS type normalized with the  $M_L$  value:  $\log([\eta]M_L) = f(\log(M/M_L)) = f(\log L)$  (Fig. 1). It follows from Equation (2) that  $([\eta]M_L) \sim \langle h^2 \rangle^{3/2}/L \sim V/L$  where  $L = M/M_L$  is the contour length of the macromolecule and  $V$  is the volume occupied by it in solution. Consequently, the value of  $[\eta]M_L$  characterizes the volume occupied by the chain section corresponding to unit contour length of linear macromolecule. This value will be larger, for greater equilibrium rigidity of macromolecule and better thermodynamic quality of the solvent. Hence, linear macromolecules in Fig. 1 are expanded into a spectrum by sizes which are mainly determined by the statistical segment length of the macromolecular chain. Zones that form linear macromolecules are designed in Fig. 1 as follows: Zones I–II are the zones of rigid-chain macromolecules, III–IV are the zones of flexible-chain polymers, and V is the zone of the globular state. Notice that it is clear from this plot that the excluded volume effects (III–IV zone) have a smaller influence on macromolecular size or dimensions in comparison with the contribution of the equilibrium rigidity of a chain. In the case of absence of intrachain volume effects, this dependence may be predicted using the Yamakawa-Fujii intrinsic viscosity theory for wormlike cylinders (Yamakawa 1971; Tsvetkov 1989).



**Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules, Fig. 1** Normalized plot of  $[\eta]M_L$  versus  $L$  for the following polymer systems: 1 – schizophyllan, 2 – xanthane, 3 – DNA, 4 – cellulose nitrate, 5 – methylcellulose, 6 – polyvinylpyrrolidone, 7 – pullulan, 8, 9 – polystyrene and poly- $\alpha$ -methylstyrene in  $\theta$  conditions, and 10 – globular proteins. Approximate curves have shown the

following zones: I–II – extra rigid – rigid chains ( $400 \text{ nm} \geq A \geq 15 \text{ nm}$ ), III–IV – flexible chain in thermodynamically good solvents and in  $\theta$  conditions ( $A \approx 2 \text{ nm}$ ), and V – globular system ( $A < 1 \text{ nm}$ ). The references to the original publication concerning the compared polymer systems may be found in the following publications: Pavlov et al. 1997, 1999, 2008. The same concerns to Figs. 2, 3, 4, 5, and 6

In a similar way, the dependence of velocity sedimentation coefficients normalized for  $M_L$ ,  $[s]/M_L$  on  $L$ , will display the opposite trend (Fig. 2). According to Svedberg's (3), the ratio  $[s]/M_L \sim L/\langle h^2 \rangle^{1/2}$  can be obtained and this ratio  $[s]/M_L$  characterizes the degree of coiling of the macromolecule. The ratio  $[s]/M_L$  will be greater for “compact” macromolecules occupying smaller volume and smaller for rigid macromolecules. Such arguments are also applicable to plotting the characteristic translational diffusion coefficient  $[D]$  depending on  $L$ .

What is very informative is a direct comparison of the concentration dependence coefficient  $k_s$  (6 and 8) with the sedimentation coefficient  $s_0$ , taking into account the mass per unit length  $M_L$  (Pavlov and Frenkel 1995). In fact,  $(k_s M_L) \sim \langle h^2 \rangle^{3/2}/L$  and the slope of the dependence  $(k_s M_L)$  versus  $f([s]/M_L)$  will be proportional to  $A^2$ , whereas the slope of a plot of  $[\eta]M_L$  versus  $f(L)$  will be proportional to  $A^{3/2}$ . This means that a plot of  $\log(k_s M_L)$  versus  $f(\log([s]/M_L))$  will be more sensitive to the equilibrium rigidity than the plot of  $\log([\eta]M_L)$  versus  $f(\log L)$ . The conformational zoning of macromolecules using a plot of  $(k_s M_L)$  versus

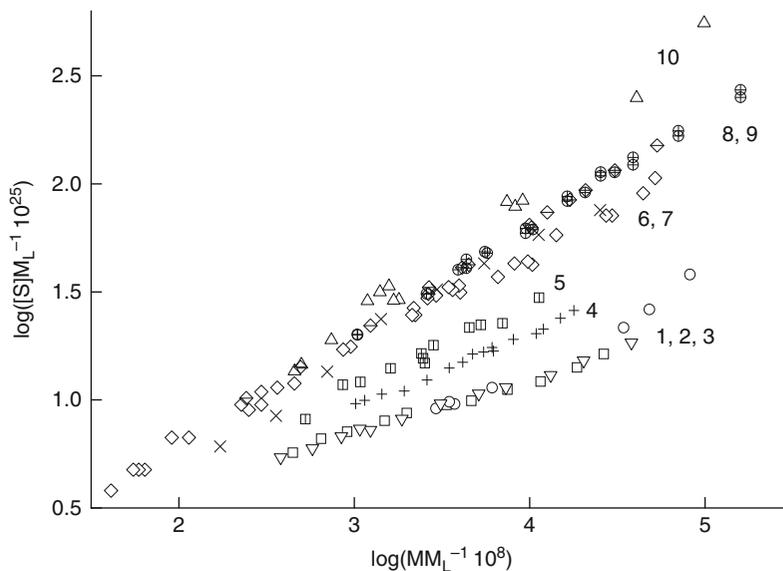
$f([s]/M_L)$  was proposed in Pavlov et al. (1997, 1999). This once again underlines the fact that the study of velocity sedimentation of a homologous series of macromolecules is an informative way for investigating conformation and, in some cases, can be used as a self-contained method (Pavlov and Frenkel 1995; Pavlov 1997).

## Second Normalization of the MHKS Relationships for Statistical Segment Length $A$

The next step is to take into account a parameter known as the statistical segment length in a further consideration of the hydrodynamic data obtained for homologous series.

Quantitative evaluation of the statistical segment length  $A$  (also known as the “Kuhn segment length”) can be estimated by Gray-Bloomfield-Hearst theory which considers the dependence of the translational friction coefficient of a wormlike chain on molar mass taking into account the influence of

**Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules, Fig. 2** Plot of the dependence of  $[\eta]M_L$  on  $L$  on a double logarithmic scale for the same systems as in Fig. 1



intramolecular draining and excluded volume effects on chain size (Gray et al. 1967):

$$[s]P_0N_A = [3/(1-\varepsilon)(3-\varepsilon)]M_L^{(1+\varepsilon)/2}A^{-(1-\varepsilon)/2}M^{(1-\varepsilon)/2} + [M_L P_0/3\pi] \left[ \ln(A/d) - (1/3)(A/d)^{-1} - \phi(\varepsilon) \right] \quad (13)$$

where  $P_0 = 5.11$  and  $\varepsilon = 2b_D - 1 = (2b_\eta - 1)/3$  is the parameter characterizing the thermodynamic quality of the polymer-solvent system, which may be calculated from the indexes  $b_D$  and  $b_\eta$ ,  $d$  is the hydrodynamic diameter of the chain (cross-sectional diameter), and  $\phi(\varepsilon) = 1.431 + 2.64\varepsilon + 4.71\varepsilon^2$ .

The viscosity data analysis may be carried out for all ranges of  $M$ , supposing that the molecular size is the same in translational and rotational friction phenomena, and results in the following equation (Pavlov et al. 1990):

$$(M^2\Phi_0/[\eta])^{1/3} = [s]P_0N_A \quad (14)$$

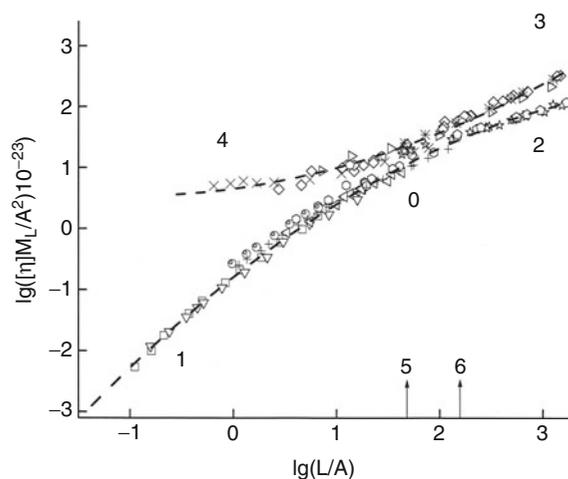
The dependence of  $[\eta]M_L$  on  $L$  can thus be additionally normalized for the segment length  $A$ . In this case, it follows from (2) and (4) that the generalized coordinates are  $[\eta]M_L/A^2$  which is a function of the relative contour length  $L/A \equiv M/(M_L A)$ :

$$[\eta]M_L/A^2 = \Phi(\varepsilon) \left[ 1 + (5/6\varepsilon + (1/6)\varepsilon^2) \right]^{-1} (L/A)^{(1-3\varepsilon)/2} \quad (15)$$

In Fig. 3 (based on data from Fig. 1), the data have been collected together and delineated by lines 1-2-3-4. Line 1-2 defines the region corresponding to polymers in the absence of intracoil excluded volume interactions and its contour represented by the following empirical expression:  $y_{\text{exp}} = -0.795 + 1.34x - 0.141x^2$ . The slope at each point is given by  $b_\eta = dy/dx = 1.34 - 0.282x$ .  $b_\eta$  varies between 1.62 and 0.5, while the number of Kuhn segments in the chain correspondingly varies between 0.1 and 1,000. This chain behavior without excluded volume effects is also predictable on the basis of the Yamakawa-Fujii theories of wormlike cylinders which gives comparable agreement through the following relationship:  $y_{\text{theor}} = -0.865 + 1.24x - 0.094x^2$ .

It is clear from Fig. 3 that the branches 1-2 and 3-4 approach each other in a region marked as “0.” The 0 or “zero” region is not an actual single point but is instead represents a locality.

Fujita (1990) generalized the behavior of flexible chains in a thermodynamically good solvent and found that, when  $L/A < 50$  ( $\log(L/A) \approx 1.7$ ), the excluded volume effects are no longer pronounced. The value of  $\log(L/A) \approx 1.7$  corresponds to this zero region. Branch 3-0 limits the region of flexible-chain polymers with maximum volume effects (thermodynamically good solvent). Branch 4-0 limits the region corresponding to a polymer with maximally large chain diameter. Lines 3-0 and 4-0 may also be approximated by the same contour defined by  $y = 0.650 + 0.230x + 0.116x^2$ .



### Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules, Fig. 3

Double normalized dependence of  $[\eta]M_L/A^2$  on  $(L/A)$  for the same polymer systems as in Fig. 1. In this plot, the data of polystyrene and poly- $\alpha$ -methylstyrene in thermodynamically good solvents and data of comb-like copolymer have been added. Branch 1-0-2 corresponds to chains without excluded volume effects, 3-0 corresponds to chains with excluded volume effects, 4-0 corresponds to chains with diameter which comparable with statistical segment length, 5 –  $\lg(L/A) = 1.7$ , and 6 –  $\lg(L/A) = 2.1$ . See text for details

This corresponds to the following expression for the slope:  $b_\eta = 0.230 + 0.232x$ .  $b_\eta$  changes in this case from 0 to 0.9 as the number of Kuhn segments in the chain increases from 0.1 to 1,000, respectively. This indicates the similarity of contributions to the size of macromolecules – one part, the excluded volume effects in the range of high  $M$ , and by another part, the diameter of chain in the range of low  $M$ . All these data can be divided into two subspaces 1-0-4 and 2-0-3. The subspace 1-0-4 corresponds to polymers of low molar mass when intracoil volume interactions become negligible. This region corresponds to the case of draining chains when the contribution of diameter to the total chain size can be considerable. The region 1-0-4 is characterized by the value of parameter  $\varepsilon = 0$  and a wide variation in the  $d/A$  value ( $0.005 < d/A < 0.5$ ). Subspace 2-0-3 corresponds to high molar mass polymers, i.e., corresponding to the case of nondraining coils when hydrodynamic draining is not manifest and the contribution of the chain diameter to total coil size is negligible. Region 2-0-3 is characterized by a change in the thermodynamic quality of the

solvent and corresponds to changes in the parameter  $\varepsilon$  in the range:  $0 \leq \varepsilon \leq 0.25$ .

The zero region corresponds to the parameter range  $1.7 \leq \lg(L/A) \leq 2.1$  (Fujita 1990; Pavlov 2007). In the region of relative contour length less than  $L/A < 50$ , intracoil volume effects virtually do not appear in thermodynamically good solvents. The plot in Fig. 3 is the locus of values of the intrinsic viscosity Flory parameter  $\Phi$  corresponding to the set (or multitude) of linear macromolecules.

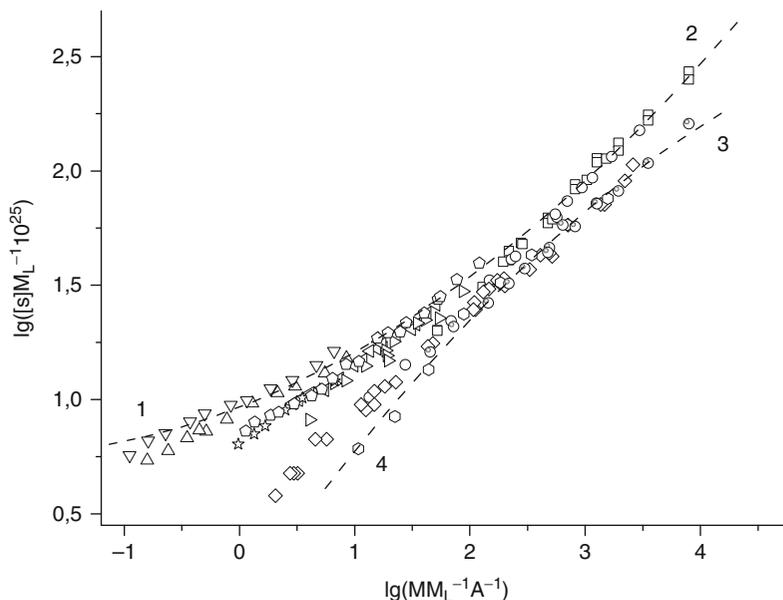
In the case of translational friction, the double logarithmic normalized plots will be in terms of the following:  $[s]/M_L$  or  $[D]M/RM_L$  versus  $(L/A)$  for the sedimentation coefficient and translational diffusion coefficient, respectively, where  $R$  is the gas constant. Figure 4 demonstrates the double normalized plot of the sedimentation coefficient, which is the locus of the translational friction Flory parameter  $P$  corresponding to the set of linear polymers. The parameter  $P$  is analogous to the Stokes coefficient  $6\pi$  in the relation of the translational friction coefficient  $f$  with the end-to-end distance of linear chain  $\langle h^2 \rangle$ :  $f = P\eta_0 \langle h^2 \rangle^{1/2}$ . Lines 1-2 and 3-4 in Fig. 4 have the same meaning as in Fig. 3, and virtually, all experimental data are situated inside these two curved lines. Hence, plots 1-4 show clearly that the main parameters determining the hydrodynamic characteristics of linear chains in solution are the statistical segment length (or persistence length) and the linear density of the chain. The values contributing less are the hydrodynamic diameter – the effect of which is manifested in the range of low molar mass – and the thermodynamic quality of the solvent – manifested at high molar mass. It should be emphasized that, to evaluate adequately the equilibrium rigidity of the chain from hydrodynamic characteristics, it is necessary to precisely take into account the effects of intramolecular draining and volume effects (Pavlov 2002).

### General Comparison of Hydrodynamic Properties of Macromolecules of Different Forms

In Figs. 1 and 2, data have been presented for linear polymers for which the linear chain density  $M_L$  can be determined. For nonlinear molecules, this approach is not suitable, because in this case, it is necessary to consider only the volume density of the polymer

### Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules,

**Fig. 4** Double normalized dependence of  $[\eta]/M_L$  on  $(L/A)$  for the same polymer systems as in Fig. 3. See text for details



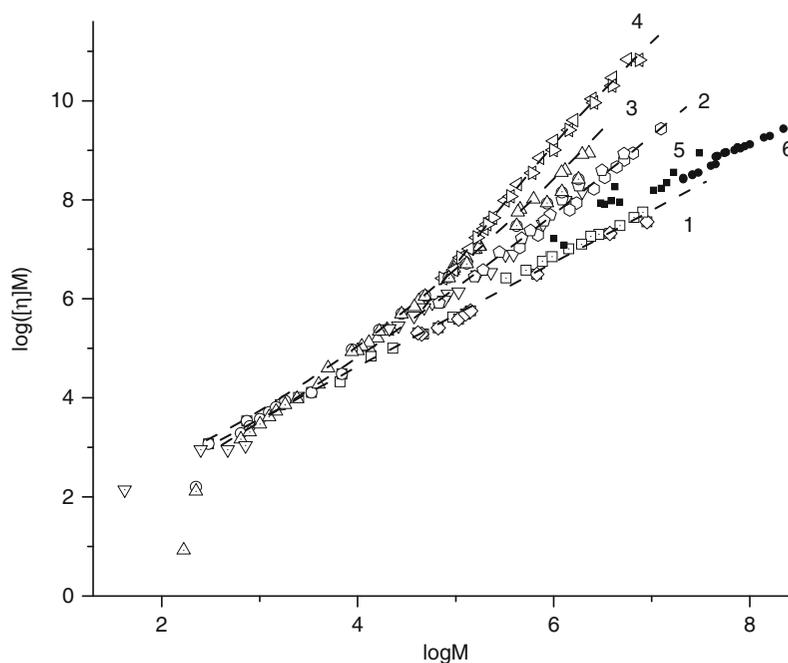
substance. Just this consideration makes it possible to compare both linear and general macromolecules in the system of coordinates  $[\eta]M$  versus  $M$  (Fig. 5). The product  $([\eta]M)$  is claimed to be proportional to the volume  $V$  occupied by the (macro)molecule in solution. Note that  $([\eta]M)$  is the key value in the Benoit universal calibration interpreting the results from exclusion gel permeation chromatography of polymers. In the first approximation, the slope of these dependences will be inversely proportional to the average intracoil density ( $\sim \log(1/\rho)$ ). The upper curves are located along the ordinate, and the lower along the density of polymer substance in the volume limited by the polymer molecule.

Star-like polystyrenes and short-branched molecules such as glycogen, lignins, and dendrimers can be added to those to linear systems as shown in Fig. 1. Data concerning supramolecular structures such as micelles formed by polystyrenes-polymethacrylic acid block copolymers and dendrigrafts can also be taken into account in this case. Finally, data on the intrinsic viscosity of low molecular weight chelate compounds (alkaline FeTaNa solution, cadmium-ethylene-diamine complex) and those of mineral salts (such as LiCl in dimethylacetamide, and LiCl in  $H_2O$ ) can also be taken into account. The limiting slopes of these dependences at high  $M$  range from 1.0 to 2.4 on passing from dendrimers, glycogen, and globular proteins to comb-like chain, then flexible linear chain,

and, finally, to extra-rigid macromolecules (Fig. 5) in accordance with the fact that  $([\eta]M) \sim M^{(1+bm)}$ . When the molar mass decreases, the differences between different types of macromolecules also decreases.

In the region of low molecular mass ( $M < 1,000$ ), when the dimension of the moving macromolecules becomes comparable with those of the solvent, the  $[\eta]$  value often loses its simple physical sense. In this region of  $M$ , a negative value of  $[\eta]$  is possible, which obviously cannot be taken into consideration in this kind of plot. This means that, in this case, dissolved molecules disturb considerably the organization and properties of solvent continuum, especially its dynamic viscosity. A large scattering of data is observed in this  $M$  region.

Let us now compare the most compact molecular systems: globular proteins, dendrimers, short-branched molecules such as glycogen and lignins, molecular brushes (comb-like polymers and copolymers), star-like polystyrenes, and supramolecular structures such as micelles. The dependence for globular proteins, dendrimers, and glycogen is a straight line (Fig. 5, line 1) over the entire molecular mass range (four orders of magnitude) with a slope of  $1.05 \pm 0.01$  and with a linear correlation coefficient  $r = 0.9984$ . This slope value means that, to a good approximation,  $V \sim M$  over the entire molar mass range. In other words, the average density of these molecular systems is virtually constant for the entire molecular



**Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules, Fig. 5** Dependence of  $[\eta]M$  versus  $M$  on a double logarithmic scale for different polymer systems: 1 – globular proteins, dendrimers, and glycogen; 2 – comb-like polymers; 3 – polystyrene, poly- $\alpha$ -methylstyrene, and PVP in thermodynamically good solvents; 4 – polysaccharides schizophyllan and xanthan; and 5 and 6 – supramolecular micelles formed by block copolymers of polystyrene-polymethacrylic acid and polystyrene – block – poly

(ethylene/propylene) copolymers. Approximate curves according to data shown for the following: 1 – globular and short-branched polymer systems (1.0,  $d_f = 3.0$ ); 2 – molecular brushes (1.4,  $d_f = 2.14$ ); 3 – flexible chain in thermodynamically good solvents (1.75,  $d_f = 1.71$ ); and 4 – extra-rigid polymers (2.4,  $d_f = 1.25$ ). Values corresponding to limiting slopes ( $1 + b_\eta$ ) in the region of high molecular mass and corresponding fractal dimension are shown in parentheses

range. These molecular systems (globular proteins, dendrimers, and glycogen) possess higher average densities in solution than all other macromolecules. Other macromolecular systems of this series (comb-like and star-like polymers and copolymers) also give a single line (Fig. 5, line 2), the slope of which increases from 1.3 (for  $\log M = 4$ ) to 1.5 (for  $\log M = 6$ ) moving to the region of high  $M$ . The average density of these macromolecular systems in solution is lower than those of globular proteins. Interestingly, supramolecular systems such as micelles formed by polystyrene-polymethacrylic acid block copolymers are situated between the line (1) and (2) in the system of coordinates  $[\eta]M$  versus  $M$ , and so have the intermediary polymer density in solution between the globular proteins/dendrimers and polymer brushes.

The translational friction data of linear and branched macromolecules may be considered in the coordinates  $M/[s] = R/[D] = [f]N_A$  as a function of  $M$ ,

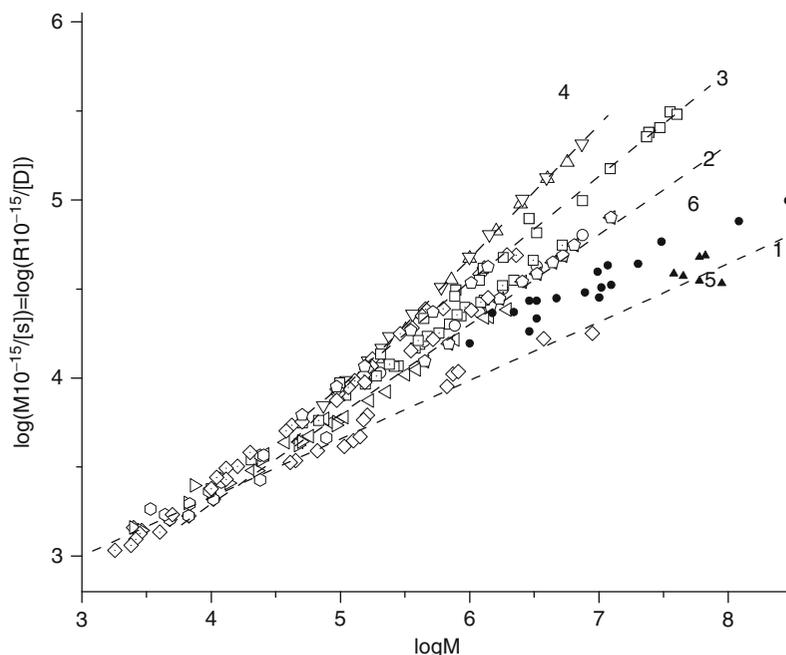
where  $[f] \equiv f_0/\eta_0$ . This plot is presented in Fig. 6 for the same classes of polymers as in Fig. 5 and has the same features as Fig. 5. Figure 6 also provides the possibility of distinguishing macromolecules of different types following their translational friction properties in the region of high molecular mass.

### Examples of Normalized Hydrodynamic Plots in Interpreting Hydrodynamic Data

Applicability of normalized hydrodynamic plots can be illustrated with the data obtained for the sodium poly(4-styrenesulfonate) (NaPSS) in aqueous solutions as studied by Takahashi et al. (1967), Hirose et al. (1999), and Pavlov et al. (2008) at various ionic forces and in a wide interval of molar mass. In salt-free solutions, NaPSS shows polyelectrolyte behavior, at addition NaCl polyelectrolyte effects suppress to be

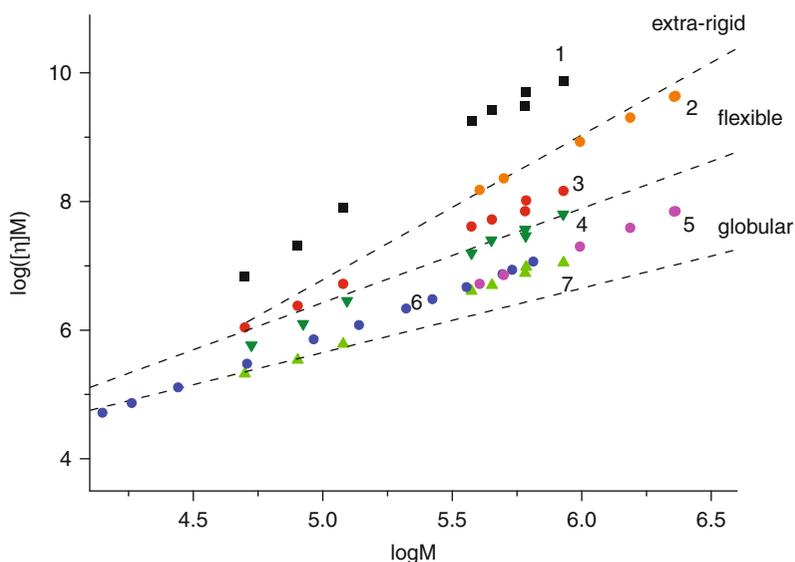
### Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules, Fig. 6

Dependence of  $M/[\eta] = R/[D]$  versus  $M$  on a double logarithmic scale for the following: 1 – globular and short-branched polymer systems, 2 – molecular brushes, 3 – flexible chain in thermodynamically good solvents, 4 – extra-rigid polymers, and 5 and 6 – supramolecular micelles formed by block copolymers of polystyrene – polymethacrylic acid and supramolecular polystyrene – polyoxyethylene core-shell amphiphilic dendrigrafts



### Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules, Fig. 7

Dependence of  $[\eta]M$  versus  $M$  on a double logarithmic scale for sodium poly(4-styrenesulfonate) at different ionic strength: 1 – in water salt-free solution, 2 – 0.005 M NaCl, 3 – 0.2 M NaCl, 4 – 1 M NaCl, and 5–7 4.17 M NaCl. 1, 3, 4, and 7 are the data from Pavlov et al. (2008), 2 and 5 are the data from Takahashi et al. (1967), and 6 is the data from Hirose et al. (1999)

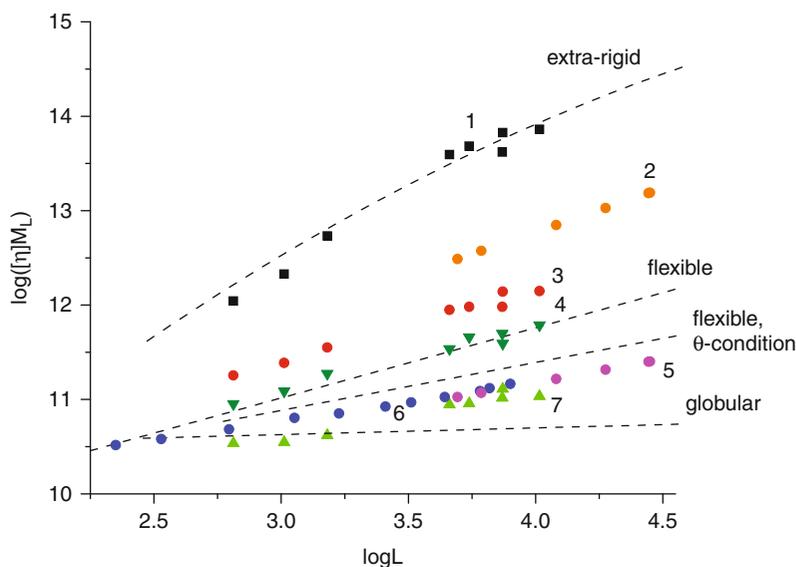


also attended by the decrease of the coil sizes. The data concerning values of intrinsic viscosity is considered. The molar masses have been determined by one of the absolute methods, for example, by sedimentation-diffusion analysis or by light scattering.

The intrinsic viscosity in salt-free solutions is defined as the initial slope of the following dependences:  $\ln \eta_r$  versus  $f(c)$  at  $c \rightarrow 0$ , where  $\eta_r$  is the relative

viscosity of a solution at concentration  $c$  (Pavlov et al. 2006). All over raw experimental data is analyzed in coordinates  $([\eta]M)$  on  $M$  (Fig. 7). It is obvious that at transition from salt-free solutions to solutions with the moderate ionic strength (0.005 M, 0.2 M) and then to solutions with the high ionic strength (4.17 M), the conformation of macromolecules of NaPSS molecules undergoes changes from rodlike conformation up to

**Hydrodynamics of Macromolecules: Conformation Zoning for General Macromolecules, Fig. 8** Normalized plot of  $[\eta]M_L$  versus  $L = M/M_L$  for sodium poly (4-styrenesulfonate) at different ionic strength which are the same as in Fig. 7



conformation of strongly dense coil. More quantitative comparison can be lead comparing data in coordinates  $[\eta]M_L$  from  $L$  (Fig. 8) which shows that in salt free solvent NaPSS molecules belong to extra-rigid zone, at 0.005 M solvent NaPSS molecules be found in rigid macromolecules zone, at 1 M belong to the zone of flexible macromolecules, and at 4.17 M macromolecules of NaPSS be found between the zone of flexible macromolecules in  $\theta$  conditions and the zone of globular conformation. These conclusions are supported by estimation of the statistical segment lengths (Hirose et al. 1999; Pavlov et al. 2008).

## Conclusions

The relations (2, 3, 4) and (1 and 10) allow to the possibility of relating the fractal dimension to the scaling indices of the MHKS relationships:

$$d_f = 3/(1 + b_\eta) = (1 - b_s)^{-1} = |b_D|^{-1}. \quad (16)$$

Globular proteins, glycogen, and dendrimers are characterized by high fractal dimension  $d_f \approx 3$ . These kinds of macromolecules resemble macroscopic particles. The macromolecular brushes and other supramolecular systems like micelles and dendrigrafts have fractal dimension between  $2 < d_f < 3$ , and linear macromolecules have  $d_f$  between  $1.1 \leq d_f \leq 2$  following the equilibrium rigidity of chains (“soft matter”).

It is clear that the macromolecular systems are ranged in Figs. 3, 4, 7, and 8, following their fractal dimensions (gradually decreasing or increasing).

Consequently, the vast quantity of experimental data gives the possibility of arranging macromolecules of different topology accordingly to their hydrodynamic volume or size in the system of coordinates  $[\eta]M$  versus  $M$  or  $M/[s]$  versus  $M$ , respectively. The consideration of the mass per unit length leads to the size spectrum of linear macromolecules in the system of coordinates  $[\eta]M_L$  versus  $L$ . Further consideration of the statistical Kuhn segment length (or persistence length) shows that the principal parameters governing the sizes of macromolecules in solution are the statistical segment length (equilibrium rigidity of chain) and the mass per unit length. The second order parameters are thermodynamic quality of solvent and hydrodynamic diameter of chain which determines the draining effects inside moving macromolecules in solution.

Generalization of the available library of experimental hydrodynamic data suggests that, at a considerable degree of probability, the characteristics of a new linear polymer system, the fractal dimension of which lies between  $1 < d_f \leq 3$ , will be located inside the region determined by the lines I and V in a system of coordinates  $[\eta]M_L - L$  as shown in Fig. 1. The characteristics of any new soluble polymer system will be located inside the region determined by lines 1 and 4 in a system of coordinates  $[\eta]M - M$  (Fig. 5).

## Summary

There are fundamental relationships that link hydrodynamic parameters to the molecular weight or molar mass and size of polymer chains. These include scaling or power-law relations known in polymer science as the Mark-Kuhn-Houwink-Sakurada equations. These have been further developed using the mass per unit length of the polymer chain and a parameter known as the statistical segment length, and relations have been developed for representing the conformations of a variety of macromolecular conformation types, whether they be synthetically made or naturally occurring.

## Cross-References

- ▶ [Cellulose](#)
- ▶ [Dynamic Light Scattering](#)
- ▶ [Lipid Lateral Diffusion](#)
- ▶ [Multiangle Light Scattering from Separated Samples \(MALS with SEC or FFF\)](#)
- ▶ [Polysaccharides: Biophysical Properties](#)
- ▶ [Sedimentation Velocity Analytical Ultracentrifugation](#)
- ▶ [X-Ray Scattering of Lipid Membranes](#)

## References

- Benoit H, Higgins J. Neutron scattering by polymers. Oxford: Oxford University Press; 1992.
- Brandrup J, Immergut EH, Grulke EA, editors. Polymer handbook. New York: Wiley; 1999.
- Cantor CR, Schimmel PR. Biophysical chemistry. San Francisco: Freeman; 1980.
- de Gennes P. Scaling concepts in polymer physics. Ithaca: Cornell University Press; 1979.
- Fujita H. Polymer solutions. Amsterdam: Elsevier; 1990.
- Gray HB, Bloomfield VA, Hearst JE. Sedimentation coefficient of linear and cyclic wormlike coils with excluded-volume effects. *J Chem Phys.* 1967;46:1493–507.
- Hirose E, Iwamoto Y, Norisuye T. Chain stiffness and excluded-volume effects in sodium poly(styrenesulfonate) solutions at high ionic strength. *Macromolecules.* 1999;32:8629–34.
- Mandelbrot B. The fractal geometry of nature. New York: Freeman; 1982.
- Pavlov GM. The concentration dependence of sedimentation for polysaccharides. *Eur Biophys J.* 1997;25:385–97.
- Pavlov GM. Evaluation of draining and volume effects in the interpretation of hydrodynamic data for linear macromolecules. *Progr Coll Polym Sci.* 2002;119:149–58.
- Pavlov GM. Size and average density spectrums of macromolecules obtained from hydrodynamic data. *Eur Phys J E.* 2007;22:171–80.
- Pavlov GM, Frenkel SYA. Sedimentation parameters of linear polymers. *Progr Coll Polym Sci.* 1995;99:101–8.
- Pavlov GM, Panarin EF, Korneeva EV, Kurochkin K, Baikov V, Uschakova V. Hydrodynamic properties of polyvinylpyrrolidone molecules in dilute solutions. *Makromolek Chem.* 1990;191:2889–99.
- Pavlov GM, Rowe AJ, Harding SE. Conformation zoning of large molecules using the analytical ultracentrifuge. *Trends Anal Chem.* 1997;16:401–5.
- Pavlov GM, Harding SE, Rowe AJ. Normalized scaling relations as a natural classification of linear macromolecules according to size. *Progr Coll Polym Sci.* 1999;113:76–80.
- Pavlov GM, Gubarev AS, Zaitseva II, Sibileva MA. On the intrinsic viscosity determination of polyelectrolyte chains in aqueous salt-free solution. *J Appl Chem.* 2006;79:1423–8.
- Pavlov GM, Zaitseva II, Gubarev AS, Korneeva EV, Gavrilova II, Panarin EF. Conformation of sodium poly(4-styrenesulfonate) macromolecules in aqueous solutions. *DAN RAN (Doklady Chemistry).* 2008;419:770–1.
- Takahashi A, Kato T, Nagasawa M. The second virial coefficient of polyelectrolytes. *J Phys Chem.* 1967;71:2001–10.
- Tanford C. Physical chemistry of macromolecules. New York: London: Wiley; 1961.
- Tsvetkov VN. Rigid-chain polymers. New York: Consultants Bureau; 1989.
- Tsvetkov VN, Eskin VE, Frenkel SYA. Structure of macromolecules in solutions. London: Butterworths; 1970.
- Yamakawa H. Modern theory of polymer solutions. New York: Harper and Row; 1971.

---

## Hydrodynamics of Nucleic Acids: Modeling Overall Conformation and Dynamics

José García de la Torre  
Faculty of Chemistry, Department of Physical Chemistry, University of Murcia, Spain

## Definition

The shape of rigid structures can be represented by various models with the desired detail. The wormlike chain is a versatile model for the conformation of long nucleic acids. The overall hydrodynamic properties are the sedimentation and diffusion coefficients, intrinsic viscosity, and rotational and relaxation times.

## Introduction

The essential roles played by nucleic acids are, in many aspects, related to the overall conformation (i.e., size, shape, flexibility, etc.) and dynamics (intramolecular, diffusive, etc.) of the macromolecular chain. Although such roles will take place in the complex environment of the cell, for the understanding or characterization of such aspects, the proper way is the study of properties in dilute solution, where the features of the individual macromolecules are manifested without influence from their surroundings. Thus, solution properties like the hydrodynamic coefficients (diffusion, ► [sedimentation](#), viscosity), dynamic spectroscopies (fluorescence, NMR relaxation), flow behavior, and scattering (light, x-rays, neutrons) (Van Holde et al. 1998; Serdyuk et al. 2007) are primary sources of structural information. The relationship between such properties (and, ultimately, their biological functions), on the one hand, and the conformational and dynamical aspects, on the other, requires theoretical treatments and computational tools that would allow the calculation of properties from structure or for us to extract conformational information from properties.

Although there exist other important topologies among DNA and nucleic acids, like circular or superhelical forms, hairpin loops, single-stranded chains, etc., the relevant structure, which is the subject of this article, is that of linear chains of double-stranded double-helical, linear DNA (dsDNA). The DNA double helix is a quite stiff structure, but still allows for some flexibility, whose influence on conformation and dynamics will depend on the size of the DNA. Thus, DNA can adopt a range of conformations, from that of small rigid oligonucleotides to that of very long, flexible native macromolecules. Also, the range of level of detail in modeling can range from fully atomic to more coarse-grained descriptions and even to idealized polymeric chains.

For the purpose of computing or simulating overall conformation and solution properties, hydrodynamic calculation procedures are available for rigid bead models (Carrasco and García de la Torre 1999). Flexible molecules can be treated by Monte Carlo approaches when one considers only conformation and overall properties, and Brownian Dynamics with the inclusion of hydrodynamic interactions is the methods of choice for long-scale internal dynamics and single-molecule behavior (Schlick 2002; García de la Torre et al. 2005, 2010).

## The Wormlike Chain Model

The wormlike chain (WMC) model is the most widely used paradigm for describing the overall conformation and overall hydrodynamic properties of dsDNA. In simple words, the WMC is a filament of contour length  $L$ , diameter  $d$ , with a partial flexibility gauged by the so-called persistence length,  $a$ , which is interpreted as the part of the contour over which the filament can be considered as essentially rigid. The model includes the limit of the rigid rod ( $L/a \ll 1$ ) and the fully flexible random coil ( $L/a \gg 1$ ). The relationship between  $L$  and molecular weight  $M$  is expressed by a constant, the mass per unit length  $M_L$ . As described in classical monographs (see Bloomfield et al. 1999), this model is adequate to describe the conformation and overall properties of dsDNA over the whole range of sizes, from small oligonucleotides to native DNAs. The usually accepted values of the WLC parameters are  $a = (50 \pm 5)$  nm,  $d = (2.3 \pm 0.3)$  nm, and  $M_L = (1,900 \pm 100)$  Da/nm (the indicated ranges are guessed from the author's experience).

The theoretical description of properties of WLCs has been the subject of intensive work for many years. Theoretical approximations, and separate treatments of the WLC for different  $L/a$  ranges, had been common until a recent, remarkable achievement: the development of a computational scheme based on Monte Carlo calculation and rigid-body hydrodynamics (► [HYDRO](#)) of properties from the WLC parameters (Amoros et al. 2011). A global fit analysis (► [HYDFIT](#)) has been able to describe the various overall solution properties of dsDNA from 8 to 200,000 base pairs, in an unified manner, with results for the WLC that match very well the commonly accepted values for the double helix.

## A Simple Model for Very Short dsDNA: The Rigid Rod

For very short dsDNA oligonucleotides ( $L/a \ll 1$ , up to, say, 20–30 base pairs), the WLC reduces to a rigid rod. In this range,  $L/d$  is indeed moderate ( $< 10$ ). The most widely employed computational approach for the solution properties for such short, thick rods is implemented in a scheme, which has been recently updated (Ortega and García de la Torre 2003). Diffusion coefficients determined by ► [dynamic light scattering](#) and NMR relaxation have been successfully

interpreted in terms of this approach (Eimer and Pecora 1991; Boisbouvier et al. 2003).

### A Simple Model for Very Long dsDNA: The Bead-and-Spring Chain Model

For very long dsDNAs ( $L/a \gg 1$ , with a random-coil conformation), which are still amenable to conventional determination of overall solution properties (covering up to  $M \cong 10^8$ , like that of DNA from T2 bacteriophage), the classical Rouse–Zimm bead-and-spring chain (BSC) model, in which each element in the chain represents a long piece (subchain) of the DNA, has been the usual representation for theoretical and computational purposes (Rubinstein and Colby 2003). The WMC model still describes the overall properties in this range, but it is inadequate to describe the internal dynamics of the DNA chain, which must obviously play a role in the function of such long, native DNA molecules. The internal (intra-molecular) dynamics of DNA had been classically characterized by bulk properties of DNA solutions (Bloomfield et al. 1999), until it was observed directly in the behavior of single molecules during the pioneering study by Chu and coworkers (Perkins et al. 1997). These authors observed the stretching of single DNA molecules when subjected to an extensional flow and the relaxation to the coil conformation upon cessation of the flow. Using a modified BSC, the finitely extensible (FENE) chain model, and a rigorous description of hydrodynamic interactions in Brownian dynamics simulations, this behavior can be quantitatively predicted using Brownian dynamics simulation (Schroeder et al. 2004; García de la Torre et al. 2009). In addition to predicting the essential quantity for internal dynamics, namely, the longest relaxation time, the so-called molecular individualism, i.e., the various paths and diverse instantaneous conformations of individual molecules during unfolding and refolding, can be simulated by this technique.

### More Detailed dsDNA Models

Modeling a dsDNA with  $M \cong 10^8$  Da as a BSC with up to, say, 100 beads, means that each element represents a subchain of about 1 million Da. While this works for some aspects of overall dynamics, more detail may be

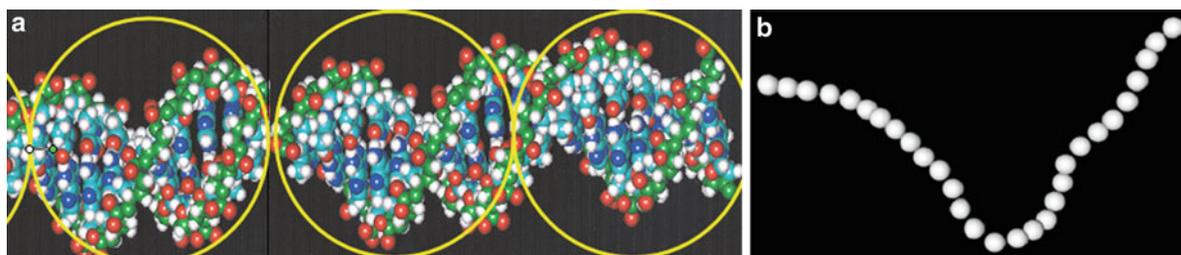
needed in some cases. The level of detail may range from a description of the thickness of the macromolecular chain – which is ignored in the BSM – to an atomic description. Certainly, a higher level of detail requires, for a DNA piece of some given length, a higher number of elements in the computational model, and this brings the obvious consequence of a higher computational cost. Thus, high-detail models are evidently restricted to DNAs of moderate size.

### Touching Beads Model

Modeling a polymer chain as a string of touching beads (looking like a pearl necklace) has also been common practice in polymer physics. The touching bead model (TBM) consists of a string of  $N$  touching beads whose diameter,  $b$ , is the same as the distance between the centers of neighbor beads. The fact that the TBM does not have a uniform cross section can have a remedy making some equivalence between this model and a cylindrical one. A string of touching beads has the same volume as a cylinder of the same total length and diameter  $d$  if one takes  $b = (3/2)^{1/2} d = 1.225 d$ . If one accepts  $d = 2.3$  nm for dsDNA, the beads would have a diameter and bead-to-bead distance  $b \cong 2.8$  nm. Thus, the ratio of number of beads to number of base pairs (nbp) is  $N/\text{nbp} \cong 0.12$ ; one bead would then represent about eight base pairs or about  $M_1 = M/N \cong 5,000$  Da in molecular weight.

Hagerman and Zimm (1981) first used this model in Monte Carlo (MC) simulations of conformational and overall hydrodynamic coefficients, and Allison (1986) showed how it can be implemented for Brownian dynamics (BD) simulation of internal dynamics. For more recent applications, see García de la Torre (2007), García de la Torre et al. (2009). The maximum number of beads typically accessible in present implementations is, approximately, 2,000 in MC and 200 in BD; therefore, the TBM can be employed to represent dsDNA of up to 15,000 base pairs in MC or 1,500 in BD simulations.

An essential aspect when devising the TBM for dsDNA is the representation of the stiffness of the double-helical filament. The chain whose segments are the bonds between beads can be assimilated to a discrete wormlike chain if some bending potential is assigned to the angle,  $\alpha$ , formed by two consecutive bonds. It has been shown (Allison 1986) that with



**Hydrodynamics of Nucleic Acids: Modeling Overall Conformation and Dynamics, Fig. 1** (a) Scheme of the touching bead model (TMB) model of dsDNA. (b) Snapshot of a TBM with 29 beads, representing a model of dsDNA with 209 base pairs

a quadratic potential  $V(\alpha) = \frac{1}{2}Q\alpha^2$ , the chain of segments represents a WCM with persistence length  $a$  if one takes  $V = (b/kTa)$ . This bending potential and its associated forces are employed, respectively, in MC and BD simulations (García de la Torre 2007; Amoros et al. 2011) (Fig. 1).

### Residue-Level Models

A further level of detail in modeling the conformation (shape and flexibility) and dynamics of DNA can be achieved with mesoscale models in which one element in the model represents one repeating unit, i.e., nucleotide in the case of nucleic acids. These are sometimes called coarse-grained models, in the sense that the detail is coarser than in a fully atomistic description (see below). For double helical B-DNA, the equilibrium conformation of the model will reproduce the shape and primary dimensions (rise per base pair, diameter, and major-minor-groove asymmetry) of the double helix. For the calculation of hydrodynamic properties, the elements are regarded as spherical beads with a frictional radius adequately chosen.

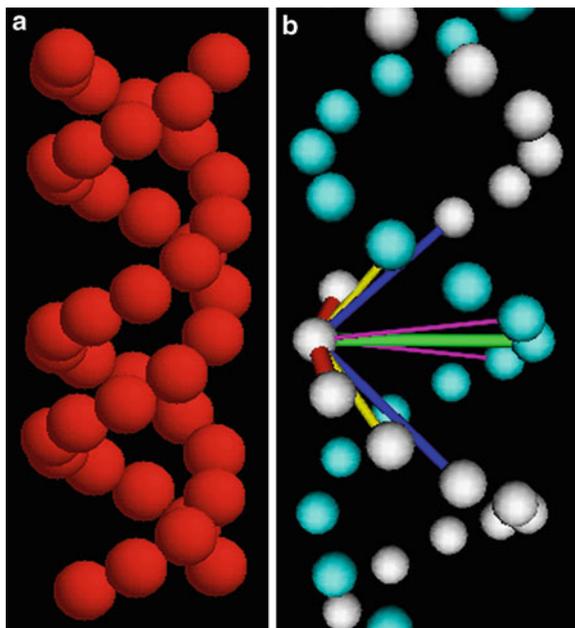
A rigid model suffices for the description of the translational and rotational (overall tumbling) of very short oligonucleotides (below, say, 25 base pairs). However, for longer DNA pieces – or even if one wishes to describe the internal, torsional dynamics of the oligonucleotides – flexibility must be taken into account. This can be easily done by means of stiff but still elastic connectors (quadratic-potential springs) between the beads. In the spirit of the elastic network models, a minimal model can be constructed joining every bead with their closest neighbors: for instance, six on its own strand (three next neighbors on each side) and three in the opposite one (its pair and

the pair's next neighbors). With a proper choice of the stiffness of those springs, the well-known bending and torsional flexibility of the double helix can be modulated. This modeling scheme was early confirmed by predictions of both overall (translational and end-over-end tumbling) and internal (NMR relaxation) dynamics of short oligonucleotides (Huertas et al. 1997).

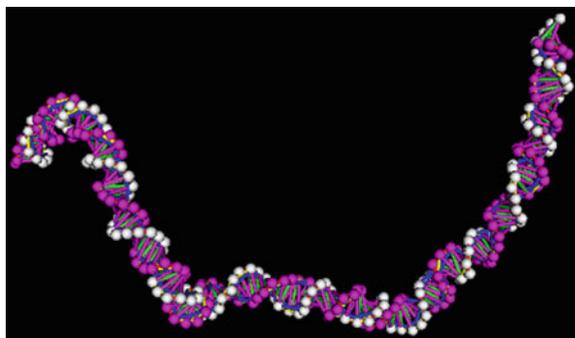
For a Brownian dynamics simulation, the presently accessible number of beads in the model make it possible to carry out simulations of these mesoscale models, with one bead per nucleotide, for DNA pieces having up to, approximately, 200–300 nucleotides (longer DNAs would require touching beads models). Nonetheless, this range can cover cases of present interest involving moderately long nucleic acids (Figs. 2 and 3).

### Models from Atomic-Level Structures

With the present proliferation of atomic-level structures of biomacromolecules, the employment of such detailed structures in dynamic modeling is desirable. Even if such high-level details are not significant for some overall and large-scale internal dynamics, certainly atomic structures should be valid starting points for computational predictions. In the field of rigid proteins, this scheme has been implemented in the program HYDROPRO and other programs of the ► [HYDRO suite](#). Each (non-hydrogen) atom in the structure is represented by one spherical element in the model, having some fixed hydrodynamic radius, which is practically the same, about 0.28 nm for both protein and nucleic acid atoms (Fernandes et al. 2002; Ortega et al. 2011). Hydrodynamic coefficients of rigid nucleic acids and nucleoprotein complexes can be



**Hydrodynamics of Nucleic Acids: Modeling Overall Conformation and Dynamics, Fig. 2** (a) Scheme of model for dsDNA with one bead per nucleotide, shown in the straight (equilibrium) conformation. Such model, considered to be rigid, is valid for predicting overall solution properties of small dsDNA. (b) Scheme showing the stiff connections of each element in the elastic network model (Huertas et al. 1997)



**Hydrodynamics of Nucleic Acids: Modeling Overall Conformation and Dynamics, Fig. 3** Snapshot of a dsDNA fragment at an instantaneous conformation during the simulation of the elastic network model (García de la Torre et al. 2009) of a fragment with 135 base pairs

predicted with uncertainties not larger than experimental errors.

Even if atomic-resolution structures are available, one may wish to devise the hydrodynamic model with a lower resolution, with one bead per residue. This has the advantage of decreasing the number of elements in

the model, with the subsequent reduction in computing requirements. This approach performs practically as well as that with atomic models (Ortega et al. 2011).

## Computer Programs

Computer programs for simulating the conformation, dynamics, and solution properties of nucleic acids are freely available from <http://leonardo.inf.um.es/macromol/>. For rigid models, one can employ programs in the ► [HYDRO suite](#):

- HYDRO++ calculates solution properties rigid bead models in general.
- HYDROPRO and HYDRONMR compute solution properties of rigid models with either atomic.

For flexible models, ranging from the BSM to the more detailed TSM and the BPN, one can employ:

- MONTEHYDRO, for Monte Carlo simulation of conformational and overall dynamic quantities
- SIMUFLEX, for Brownian dynamics simulation (including hydrodynamic interaction) of any conformational or dynamic aspect, including single-molecule event

Furthermore:

- WORMCYL is available for the calculation of solution properties of rodlike and ► [wormlike](#) models.
- ► [HYDFIT](#) provides tools to fit the parameters of rodlike and WLC models to experimental data.

## Summary

Modeling, computing, and simulation procedures are available to describe the conformation and dynamics of nucleic acids, and particularly double-stranded DNA, from small oligonucleotides to large, native macromolecules. Various levels of detail can be adopted in the models, which can be adapted to the size of the case under study so that the computing or simulation becomes feasible with available computational resources. The problems that are covered range from bulk solution properties to single-molecule events.

## Cross-References

- [Dynamic Light Scattering](#)
- [HYDFIT and Related Packages for Linear Molecules](#)

- ▶ [HYDRO Suite of Computer Programs for Solution Properties of Rigid Macromolecules](#)
- ▶ [Neutron Scattering of Membranes](#)
- ▶ [Nucleic Acid NMR – Introduction](#)
- ▶ [Protein Conformational Dynamics by Relaxation Dispersion](#)
- ▶ [Sedimentation Velocity Analytical Ultracentrifugation](#)
- ▶ [Single-Particle Tracking](#)
- ▶ [Wide and Small Angle X-Ray Scattering](#)
- ▶ [Worm-Like Chain \(WLC\) Model](#)
- ▶ [X-Ray Scattering of Lipid Membranes](#)

## References

- Allison SA. Brownian dynamics simulation of wormlike chains. *Macromolecules*. 1986;19:118–24.
- Amoros D, Ortega A, García de la Torre J. Hydrodynamic properties of wormlike macromolecules: Monte Carlo simulation and global analysis of experimental data. *Macromolecules*. 2011;44:5788–97.
- Bloomfield VA, Crothers DM, Tinoco Jr I. *Nucleic acids: structure, properties and function*. Sausalito: University Science; 1999.
- Boisbouvier J, Wu Z, Ono A, Kainosho M, et al. Rotational diffusion tensor of nucleic acids from  $^{13}\text{C}$  NMR relaxation. *J Biomol NMR*. 2003;27:133–42.
- Carrasco B, García de la Torre J. Hydrodynamic properties of rigid particles. Comparison of different modelling and computational procedures. *Biophys J*. 1999;76:3044–57.
- Eimer W, Pecora R. Rotational and translational diffusion of short rodlike molecules in solution: oligonucleotides. *J Chem Phys*. 1991;94:2324–9.
- Fernandes M, Ortega A, García de la Torre J. Calculation of hydrodynamic properties of small nucleic acids from their atomic structures. *Nucleic Acids Res*. 2002;30:1782–8.
- García de la Torre J. Dynamic electro-optic properties of macromolecules and nanoparticles in solution. A review of computational and simulation methodologies. *Colloids Surf B*. 2007;56:4–15.
- García de la Torre J, Ortega A, Pérez Sanchez HE, Hernández Cifre JG. MULTIHIDRO and MONTEHYDRO: conformational search and Monte Carlo calculation of solution properties of rigid and flexible macromolecular models. *Biophys Chem*. 2005;116:12–128.
- García de la Torre J, Hernández Cifre JG, Ortega A, Schmidt RR, Fernandes MX, Pérez HE, Pamies R. SIMUFLEX: algorithms and tools for simulation of the conformation and dynamics of flexible molecules and nanoparticles in dilute solution. *J Chem Theor Comput*. 2009;5:2606–18.
- García de la Torre J, Ortega A, Amorós D, Rodríguez SR, Hernández Cifre JG. Methods and tools for the prediction of hydrodynamic coefficients and other solution properties of flexible macromolecules in solution A tutorial minireview. *Macromol Biosci*. 2010;10:721–30.
- Hagerman P, Zimm BH. Monte Carlo approach to the analysis of the rotational diffusion of wormlike chains. *Biopolymers*. 1981;20:1481–502.
- Huertas ML, Navarro S, López Martínez MC, García de la Torre J. Simulation of the conformation and dynamics of a double-helical model for DNA. *Biophys J*. 1997;73:3142–53.
- Ortega A, García de la Torre J. Hydrodynamic properties of rod-like and disk-like particles in dilute solution. *J Chem Phys*. 2003;114:9914–9.
- Ortega A, Amorós D, García de la Torre J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J*. 2011;101:892–8.
- Perkins TT, Smith DE, Chu S. Single polymer dynamics in an elongational flow. *Science*. 1997;276:2016–21.
- Rubinstein M, Colby RH. *Polymer physics*. New York: Oxford University Press; 2003.
- Schlick T. *Molecular modelling and simulation: an interdisciplinary guide*. New York: Springer; 2002.
- Schroeder CM, Shaqfeh ESG, Chu S. Effect of hydrodynamic interactions on DNA dynamics in extensional flow: simulation and single Molecule experiment. *Macromolecules*. 2004;37:9242–56.
- Serdyuk IN, Zaccai NR, Zaccai J. *Methods in molecular biophysics: structure, dynamics, function*. Cambridge: Cambridge University Press; 2007.
- Van Holde KE, Johnson WC, Ho PS. *Physical biochemistry*. Upper Saddle River: Prentice-Hall; 1998.

---

## Hydrogen Atom Transfer (HAT)

- ▶ [Proton-Coupled Electron Transfer](#)

---

## Hydrogen Bonding

- ▶ [Rhodopsins – Intramembrane Signaling by Hydrogen Bonding](#)
- ▶ [Ultraviolet Resonance Raman \(UVRM\) Spectroscopy Studies of Structure and Dynamics of Proteins](#)
- ▶ [X-Ray Diffraction and Crystallography of Oligosaccharides and Polysaccharides](#)

---

## Hydrogen Exchange

- ▶ [H/D Exchange](#)

## Hydrogen–Deuterium Exchange

► [Infrared Spectroscopy of Protein Dynamics: H/D Exchange](#)

## Hyperfine Shifts

Ivano Bertini and Giacomo Parigi  
Magnetic Resonance Center (CERM) and Department  
of Chemistry, University of Florence, Sesto Fiorentino  
(FI), Italy

### Synonyms

[Pseudocontact and contact shifts](#)

### Definition

Paramagnetic metal ions affect the NMR spectra of molecules. The hyperfine shift is the difference between the chemical shift of a nucleus in a paramagnetic molecule and that in the diamagnetic analog obtained after removal of the paramagnetic metal ion, its substitution with a diamagnetic ion, or reduction/oxidation of the paramagnetic metal to a diamagnetic state. The hyperfine shift comprises the contact and the pseudocontact shifts. The contact shift is the result of the presence of unpaired electron spin density at the observed nucleus, which may occur through direct spin delocalization or spin polarization. Contact shifts are experienced by nuclei that are positioned few chemical bonds away from the paramagnetic metal ion or even further away if there are  $\pi$  bonds, and they contain structural information. In few cases ( $\text{Fe}_4\text{S}_4^{2+}$  clusters, Fe(III) heme proteins), the contact shifts of some nuclei are related to angles between the metal ion and the ligand nuclei through heuristically obtained Karplus relationships. They may be as large as hundreds of ppm and of course may tend to zero.

Pseudocontact shifts arise from the nonzero averaging upon rotation of the dipolar coupling between the nuclear magnetic moment and the average electron magnetic moment induced by the external magnetic

field. They are related to geometrical parameters through the equation:

$$\delta^{\text{pcs}} = \frac{1}{12\pi r^3} \left[ \Delta\chi_{ax}(3\cos^2\vartheta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\vartheta\cos 2\varphi \right] \quad (1)$$

where  $\Delta\chi_{ax}$  and  $\Delta\chi_{rh}$  are the axial and rhombic anisotropy parameters of the paramagnetic susceptibility tensor, given by

$$\Delta\chi_{ax} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2}$$

$$\Delta\chi_{rh} = \chi_{xx} - \chi_{yy}$$

and  $r$ ,  $\vartheta$ , and  $\varphi$  are the spherical coordinates of the nuclei in the reference frame in which the paramagnetic susceptibility tensor  $\chi$  is diagonal. Pcs values depend only on the anisotropy of  $\chi$  and not on its magnitude; if the magnetic susceptibility is isotropic, the rotational average of the dipolar energy is zero, and no pcs occurs. They depend neither on the applied magnetic field, nor on the observed nucleus. They depend on the distance of the nucleus with respect to the metal ion and on its angular position. Hyperfine shifts of nuclei, which are expected not to experience contact shifts because of the distance from the metal ion, are due to pcs. They have been extensively used for the calculations of molecular structures. Pcs can be univocally calculated once the molecular structure and the magnetic susceptibility anisotropy tensor are known. The anisotropy tensor can be calculated from the experimental pcs (provided that at least five values are available, if the metal position is known) and the molecular structure. The molecular structure cannot be determined from pcs and the anisotropy tensor alone due to the large degeneracy in the nuclear positions resulting from [Eq. 1](#). However, the availability of pcs arising from multiple metal ions with different tensors and of other structural restraints (like NOEs) can solve this degeneracy. Pseudocontact shifts can be detected up to several tens of Å, depending on the magnitude of the magnetic susceptibility anisotropies.

A further contribution to the hyperfine shifts of nuclei close to the paramagnetic metal ion, besides the contact shift and the pseudocontact shift described by [Eq. 1](#), can arise from the ligand-centered pseudocontact shifts resulting from the presence of unpaired spin density within non-s ligand orbitals. Nuclei far from the metal ion experience only the

pseudocontact contributions, which allow to determine the parameters of Eq. 1 – if a structural model becomes available – and then to calculate pcs of nuclei close to the paramagnetic center. The latter nuclei experience also contact and ligand-centered pseudocontact shifts. Both of them can be estimated by subtracting the pcs from the hyperfine shifts. The separation between contact and ligand-centered pseudocontact shifts is very difficult: often the latter are neglected or estimated through density functional calculations.

### Cross-References

- ▶ [Cytochromes](#)
- ▶ [Hemes](#)

- ▶ [Iron–Sulfur Clusters](#)
- ▶ [Paramagnetic Metalloproteins](#)

---

### Hyperpolarization-Activated Channels

- ▶ [HCN Channels: Biophysics and Functional Relevance](#)

---

### HyuP (“Permease” in the *hyu* Operons from *Pseudomonas* sp. NS671)

- ▶ [Mhp1, the Na<sup>+</sup>-Hydantoin Membrane Transport Protein](#)

