TMM 3102: Protein Structure, Function and Disease

- Structural Biology Methods: Molecular Dynamics Simulation
  (October 7th, 2021)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI

(Partially adopted from former lectures by Dr. Maria Musgaard)
Importance: Static v.s. Dynamic

(digits.io/benefits-of-dynamic-digital-signage-over-static-signage/)
Importance: Static v.s. Dynamic

In contrast to the symmetric Sav1866 template, mouse Pgp has evolved with a high degree of asymmetry between its two "halves", which share only 59.4% sequence similarity. More specifically, 77.3% similarity lies between two NBDs and only 43.5% similarity between two halves of TMDs (Figure 2). The differences in the primary sequence result in the asymmetry of chemical properties between pseudo symmetric residues and patches. In general, charged residues (K, R, E, D) are highly hydrophilic with the largest entropy profiles whereas hydrophobic residues prefer to pack together and stabilize each other in the aqueous phase. Therefore, the analysis of different chemical properties based on the residue types of each region can help explain their different dynamics (Figure 2). Specifically, TM1 and TM2 (Figure 2 A) are 5-residues longer and with more charged residues than their pseudo symmetric domains TM6 and TM7 (Figure 2 A). The intracellular end of TM3 (Figure 2B) has more charged residues than TM9 (Figure 2 B). These features render larger mobility of half1 than half2 at TMDs. Even though the ATP-binding residues are conserved between site 1 and site 2 (Figure 3), their surrounding residues are substantially different. WA1 and WB1 are surrounded by more charged residues (Figure 2 D and E) than WA2 and WB2 (Figure 2 D and E). Whereas LSGGQ2 is surrounded by more charged residues (Figure 2 C and F) than that of LSGGQ1 (Figure 2 C and F). In sum, our observations reveal a more hydrophilic site 1 (WA1 WB1 LSGGQ2) than site 2 (WA2 WB2 LSGGQ1) and indicate different thermodynamics and kinetics between the two sites, consistent with multiple findings in the literature37–39. A previous MD study40 was performed using the inward-facing conformation with ATP bound at each of the Walker A motifs but with both LSGGQ motifs completely dissociated. Even though the study was done at the near-opposite conformation and different ATP binding environment compared to our study, asymmetric structural dynamics between two NBDs was also observed. The asymmetry between the two halves of Pgp must therefore play a major role in the overall function and mechanism of the transporter.

The root-means-square fluctuation (RMSF) (Figure 4) reveals asymmetric dynamics of apo-Pgp in equilibrated states (Figure 4, blue line). The results suggested that MgATP-binding reduced flexibility and introduced more symmetry to the protein dynamics (Figure 4, red line). In general, apo-Pgp has larger RMSF than Pgp-MgATP across all regions. Certain locations are particularly noteworthy: ECL1 (Figure 4, peak A) has the largest RMSF of the entire structure for both apo-Pgp (3.5 Å) and Pgp-MgATP (2.8 Å). ECL4 (Figure 4, peak A) has the biggest RMSF of the TMs in half2. In apo-Pgp, ATP-binding site 1 (Figure 4 peak B, peak C, peak D, peak E) is more flexible than site 2 (Figure 4 peak B, peak C, peak D, peak E and peak F). Specifically, MgATP-binding reduces the flexibility of all the ATP-binding signature motifs at NBDs: Walker A (WA1: G423 NSGCGK429; WA2: G 1067 SSGCGK1072), Walker B (WB1: I547LLLDE552; WB 2 : I1192LLLDE1197), LSGGQ (LSGGQ 1: L527SGGQ531; LSGGGQ2: L1172SGGQ1176), A-loop (A-loop1: Y397P SR400; A-loop2: Y1040P TR1043). In addition, ATP-binding also reduced the flexibility of TMDs at various levels, which reveals an allosteric effect of nucleotide binding to the TMDs.

Bridging the Gap

Protein function
- Functional data
- Electrophysiology
- Substrate transport
- ...
- High resolution in time

Protein structure
- X-ray
- NMR
- Cryo-EM
- ~ “snapshots”
- High resolution in space

High resolution in “space” and “time”?
Brief History

- First MD study of proteins published in 1977
  ~60 residues, no solvent, ~9 ps

- 2019: full organelles, 139 million atoms, 0.5 μs

- Factors:
  - more structures determined
  - better algorithms
  - faster computer

(Nature, 1977)

(Cell, 2019)
Molecular Dynamics (MD): idea

- Classic mechanics (thinking of “Newton’s laws of motion”)

- Metaphor:
  If cycling at 15 km/h by Canal Rideau; keep a constant acceleration:
  - Predict how long to reach uOttawa main campus.
  - Predict where you are in 5 minutes.

- Do the same for all atoms in a protein system
Molecular Dynamics (MD): idea

Going to the next position:

\[ r(t+\Delta t) = r(t) + \Delta t \cdot v(t) + \frac{1}{2} [\Delta t^2 \cdot a(t)] \]

- \( r(t) \): position at “t”
- \( r(t+\Delta t) \): position after \( \Delta t \)
- \( v(t) \): velocity
- \( a(t) \): acceleration
Molecular Dynamics (MD): idea

**Acceleration:**

\[ F = m \times a \]
\[ F = -\frac{\Delta U}{\Delta r} \]

If we know U (potential energy), then we can calculate the force and the acceleration on each atom.
Molecular Dynamics (MD): workflow

In general, how do we do MD simulation?

a. Find the coordinates of a known protein model from the database.
b. Choose a force field to generate energy potential for further calculation.
c. Calculate the force that results from the theoretical potential energy.
d. Find out how molecules speed up with the obtained force.
e. Calculate the speed of the molecule and where the protein move into.
Molecular Dynamics (MD): workflow

Another way to see the MD workflow:

a. Find a model template and artificially add necessary ingredients that suit the physiological condition of the target protein. This includes protonation states, salts, water, etc.

b. Prepare the simulation system by selecting the best protocol, aka force field.

c. Run the simulation using a cluster of computers.

d. Process the data and predict the where the segment of interest moves to.
Molecular Dynamics (MD): force field

What determines “force field”?

- Atoms: different in size, softness, mass, charge, ...

- Bonds: different in lengths, stiffness, ...

- Electrons: implicitly accounted for covalent bonds.
Molecular Dynamics (MD): force field

What is a force field used for?

- Used for large molecules or conformational studies
- Not used to break or form chemical bonds
- Empirical, so no one is most correct.

Requires:
- Energy equation to describe $U$ as a function of atomic coordinates
- Constant parameters to be used in the energy equation
- Atom types to establish constant parameters, charges, masses, etc.
Molecular Dynamics (MD): force field

Selection of force field is like deciding what kind of potential energy to use:

a. Covalent bonds & bond angles

b. Torsion angles

c. Van der Waals interaction

d. Electrostatic force / charge-charge interaction
Molecular Dynamics (MD): force field

- **Do’s and Don’ts**
  - Never compare energies from different force fields, unless absolute energy is known
  - Never mix parameters, unless tested
  - Do simulations in the conditions similar to those used to obtain the force field
  - For new ligands, need a full set of parameters (all you can)
Molecular Dynamics (MD): time scale

Biological timescales

- Simulation $\Delta t$: 1-2 fs
  - Too fast:
  - Too slow:
  - Good

$\Rightarrow$ 0.5 to 1 million steps to reach 1 ns (!)

Kumar and Balbach, Biochim. Biophys. Acta 2015
Molecular Dynamics (MD)

- Local motions
- Loop dynamics
- Folds
- Ligand binding/unbinding

- Vibrational motion
- Rotational motion

- Helical folds
- Hairpin folds
- Loop dynamics
- Ligand binding/unbinding
- Protein folding

Enhanced sampling MD simulations

Classical MD simulations

Time scales:
- $10^{-15}$ fs
- $10^{-12}$ ps
- $10^{-9}$ ns
- $10^{-6}$ ms
- $10^{-3}$ ms
- 1 s
- >1 s
Figure 2. Applications of Molecular Dynamics Simulations

Here we illustrate some of the most common applications of MD simulations.

Structural and dynamic studies: Studying conformational flexibility and stability

(Structural Determination *in silico*)

(Hollingsworth & Dror, Neuron, 2018)
Structural Determination *in silico*

**Perturbations:** Observe response following controlled change to system

- **Add or remove ligand**
  - Add ligand
  - Remove ligand
  - Example: Figure 3

- **Mutation or modification**
  - E.g. Point mutation
  - Mutate Tyr to Phe

- **Alter protonation state**
  - E.g. Histidine protonation
  - Protonate
  - Deprotonate

- **Apply mechanical force**
  - E.g. Pulling ligand out of binding pocket

- **Apply an external potential**
  - Apply transmembrane potential

- **Alter protein environment**
  - E.g. Alter membrane composition

*(Hollingsworth & Dror, Neuron, 2018)*
**Structural Determination in silico**

**Processes:** *Observe a dynamic process over time*

- **Transport across a membrane**
  - Binding of molecule
  - Translocation
  - Release of molecule
  - Example: Figure 4

- **Conformational change**
  - Initial state
  - Remove ligand
  - Conformational change
  - Final state
  - Example: Figure 5

- **Protein folding**
  - Observe folding pathway

- **Ligand binding and assembly pathways**
  - Initial binding of ligand
  - Movement to binding site
  - Final ligand-bound receptor

*(Hollingsworth & Dror, Neuron, 2018)*
Molecular Dynamics (MD)

- Advantages
  - High resolution in space and time
  - Precise simulation conditions: conformations, $\pm$ ligands, ...
  - Cheap: mutations, protein-ligand, protein design, ...
  - Structure-function relationship

- Limitations
  - Validation: need experimental data
  - Timescale and sampling
  - Quality of starting structures
  - Force fields
  - No bond making/breaking, as it depends on protonation states
Case Study: P-glycoprotein (drug-resistance)

Figure 6 | Conformational change illustrated with landmark residues. Landmark residue distances at each regions of the protein are measured using the average of the last 50 ns trajectories from each simulation. The protein is shown as cartoon representation with half in pink and half2 in ice blue. The Cα of the landmark residues are drawn as spheres in green (D-L). P atoms of lipid bi-layer are drawn in semi-transparent sphere in gold. MgATP is drawn in licorice with H in white, N in blue, C in cyan, O in red, P in gold and Mg2. Orange arrows point to the directions of conformational change. The conformational changes are monitored using residue pair from two pseudo-symmetric counterparts: Extracellular gate (A) uses A79-T736 (ECL1-ECL3), G207-G850(TM3-TM9), G325-E968(ECL2-ECL4), drug binding pocket (B) uses higher end F71(TM1)-F728(TM7) pair and lower end Q343 (TM6)-Q986 (TM12) pair, ATP-binding sites (C) uses K429-S1173 (site 1) and K1072-S528 (site 2). The predominant conformation observed from all three runs of apo-Pgp (E, G, I and K) and predominant conformation observed from 30 ns-100 ns of run1, 30-50 ns of run2 and 30-200 ns of run3 of Pgp-MgATP (F, H, J and L) are different in various aspects.

Case Study: P-glycoprotein (drug resistance)

LETTER

Extended Data Figure 8

| Collective variables and structural features used to obtain the outward-facing state. a–g, Description of the collective variables (CVs) used to obtain a reliable outward-facing state of Pgp (a–d) and tracking important structural features to verify the stability of the outward-facing state (e–g). a, Orientation quaternion (β) describing the angle between the two bundles of TM helices that separate to form the outward-facing state. Distance between K185 and D993, a charged residue pair located within the translocation chamber. b, CVs used to form accurate NBD-based interactions, which include NBD–ATP interactions and X-loop interactions. Walker A (WA1 and WA2), Walker B (WB1 and WB2) and LSGGQ (L1 and L2) motifs are shown in purple, yellow and green new cartoon representations, respectively. Y397/1040 from the A-loop (white) and ATPs (cyan) are shown in stick representations, whereas Mg ²⁺ ions and Cα carbons of X-loop residues are displayed as grey and white beads, respectively. c, Metrics used in evaluating the basic structural elements that are key to any outward-facing ABC exporter, namely, dimerized NBDs (dNBD), closed cytoplasmic (α), and opened extracellular/periplasmic (β) sides. d, Sim1 (light colours) failed to maintain these basic structural requirements within 10 ns, whereas Sim2 (dark colours) results in a stable outward-facing state for up to 300 ns. Solid and dotted horizontal lines represent the corresponding values in inward-facing and outward-facing conformations, respectively, based on crystal structures of Pgp (PDB accession 4M1M) and MsbA (PDB accession 3B60). e, Description and time series of centre of mass distance between extended TM helical regions of TM3 (V164–V175) and TM10 (E887–E898) shown in orange, and TM4 (A244–A255) and TM9 (T806–D817) shown in pink, describing the tight closing of cytoplasmic side. f, Description and time series of centre of mass distance between the residues forming the top half of TMDs that open at the extracellular side (shown with pink and orange beads). g, Salt bridge interaction between K185 (TM3) and D993 (TM12). These calculations are compared between five different simulations.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Case Study: ABCG5/G8 (sterol efflux)

(Lee et al, Nature, 2016)
Case Study: ABCG5/G8 (sterol efflux)

(Xavier et al, IJMS, 2020)
Case Study: ABCG5/G8 (sterol efflux)

(Xavier et al, IJMS, 2020)