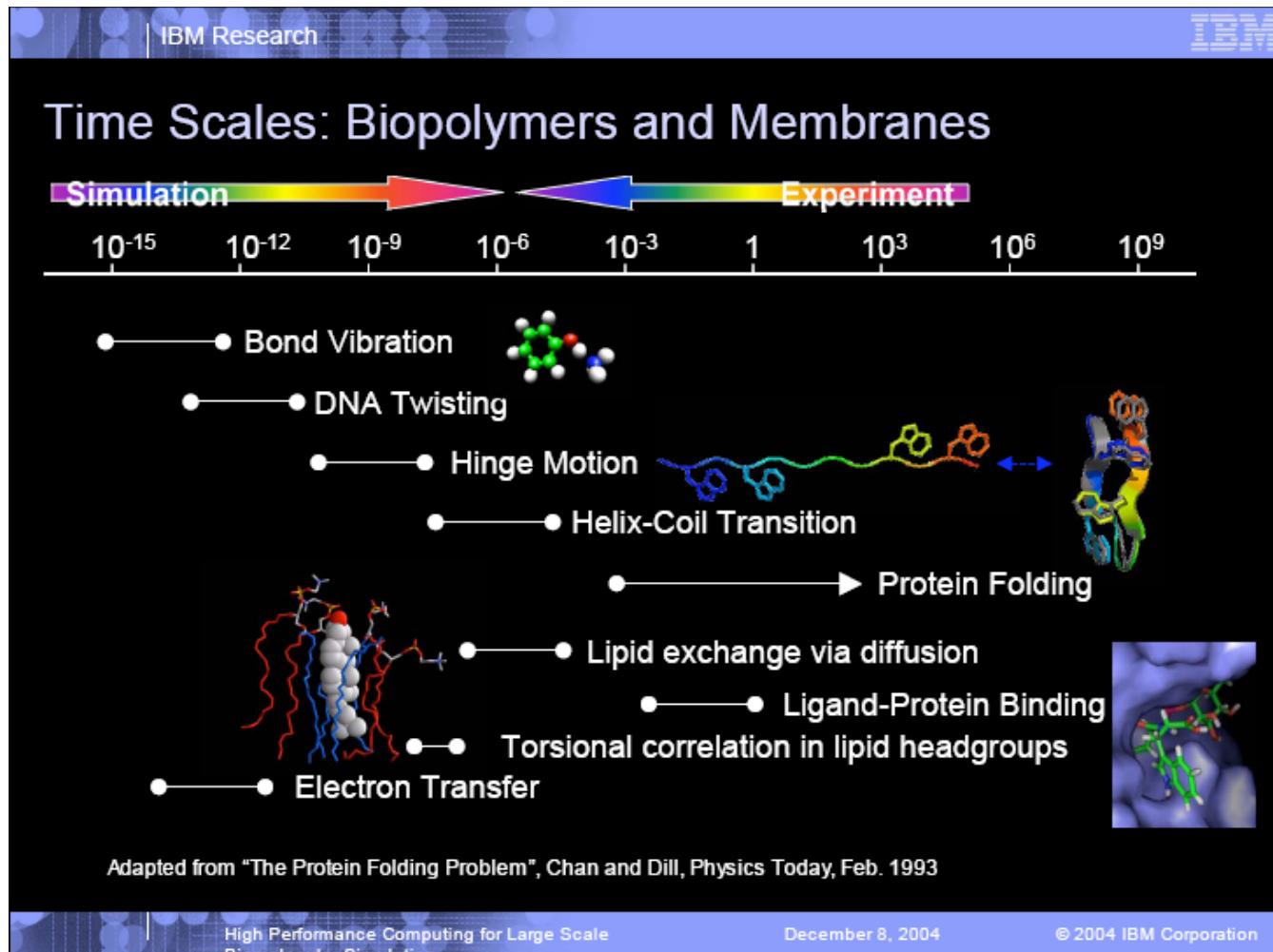


Modélisation Moléculaire

Méthodes d' exploration de la surface d' énergie
potentielle

Dynamique Moléculaire et
Analyse de trajectoire

Mouvements dans les protéines



Dynamique moléculaire

- Génère des configurations successives du système en intégrant les lois du mouvement de Newton
 1. Un corps se déplace en ligne droite à moins qu' une force agisse dessus
 2. La force est égale à la masse fois l' accélération
 3. A chaque action, il existe une réaction égale
- Trajectoire de dynamique moléculaire obtenue en résolvant la deuxième loi de Newton ($F=ma$):

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i}$$

avec

$$F_{x_i} = -\frac{\partial V}{\partial x_i}$$

Intégration de l' équation du mouvement

- Toutes les méthodes assument que les positions, vitesses, accélérations (etc...) peuvent être approximées par une série de Taylor:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \frac{1}{6} \delta t^3 b(t) + \frac{1}{24} \delta t^4 c(t) + \dots$$

$$v(t + \delta t) = v(t) + \delta t a(t) + \frac{1}{2} \delta t^2 b(t) + \frac{1}{6} \delta t^3 c(t) + \dots$$

$$a(t + \delta t) = a(t) + \delta t b(t) + \frac{1}{2} \delta t^2 c(t) + \dots$$

etc... (dérivée $n^{ième}$ de la position par rapport au temps)

- Quelques algorithmes communs:
 - Verlet, Leap-frog, Velocity-Verlet, Beeman...

L' algorithme de Verlet

- Nouvelles positions

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \dots$$

$$+ \quad r(t - \delta t) = r(t) - \delta t v(t) + \frac{1}{2} \delta t^2 a(t) - \dots$$

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t)$$

(termes d'ordres supérieurs négligés)

- Vitesses calculées après les positions (ou aux demi-temps) :

$$v(t) = \frac{r(t + \delta t) - r(t - \delta t)}{2\delta t}$$

ou

$$v(t + \frac{1}{2} \delta t) = \frac{r(t + \delta t) - r(t)}{\delta t}$$

L' algorithme "leap-frog" (1)

- Calcul des positions atomiques \vec{r}_i à tous les temps t et toutes les vitesses atomiques \vec{v}_i aux temps intermédiaires $t + \delta t/2$:

On a $v(t + \delta t) = v(t) + \delta t a(t)$ (termes d'ordres supérieurs négligés)

donc

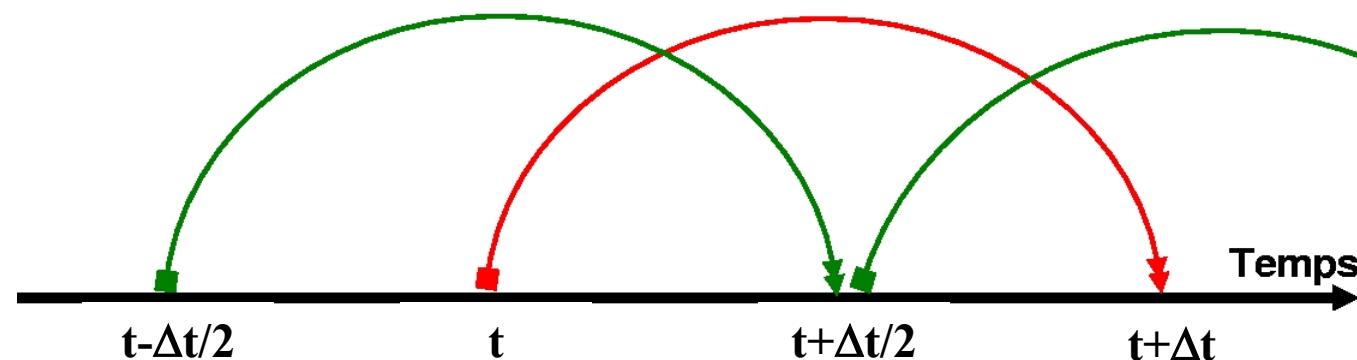
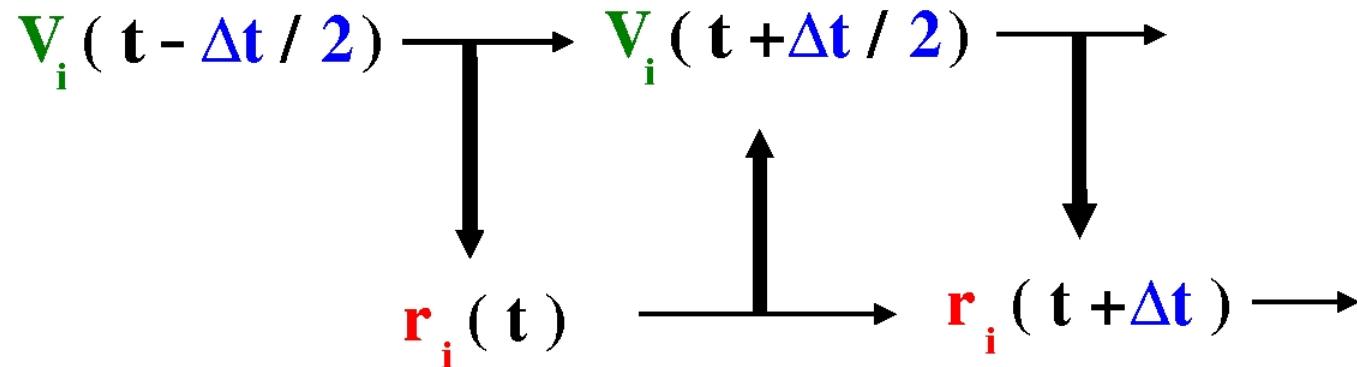
$$v\left(t + \frac{1}{2} \delta t\right) = v\left(t - \frac{1}{2} \delta t\right) + \delta t a(t)$$

On a $v(t + \frac{1}{2} \delta t) = \frac{r(t + \delta t) - r(t)}{\delta t}$

donc

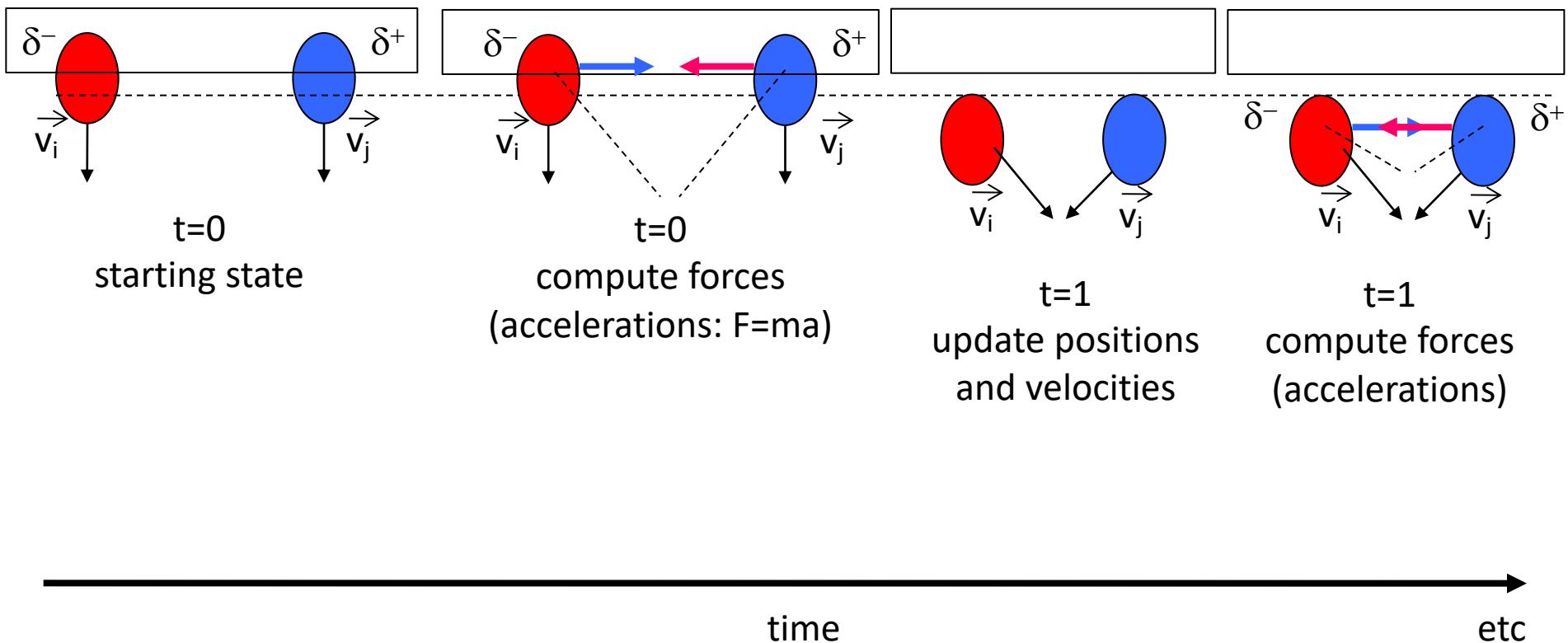
$$r(t + \delta t) = r(t) + v\left(t + \frac{1}{2} \delta t\right) \delta t$$

L' algorithme "leap-frog" (2)



Les positions et vitesses sont calculées alternativement comme si on sautait de l'une à l'autre (comme au saut de mouton !)

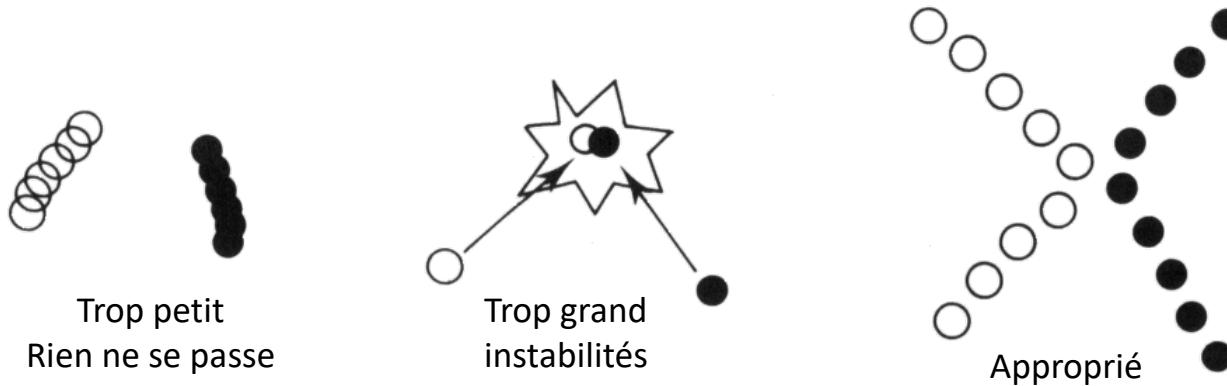
MD: a simple example



MD keeps memory of the previous step (velocity)

Choix du pas d' intégration

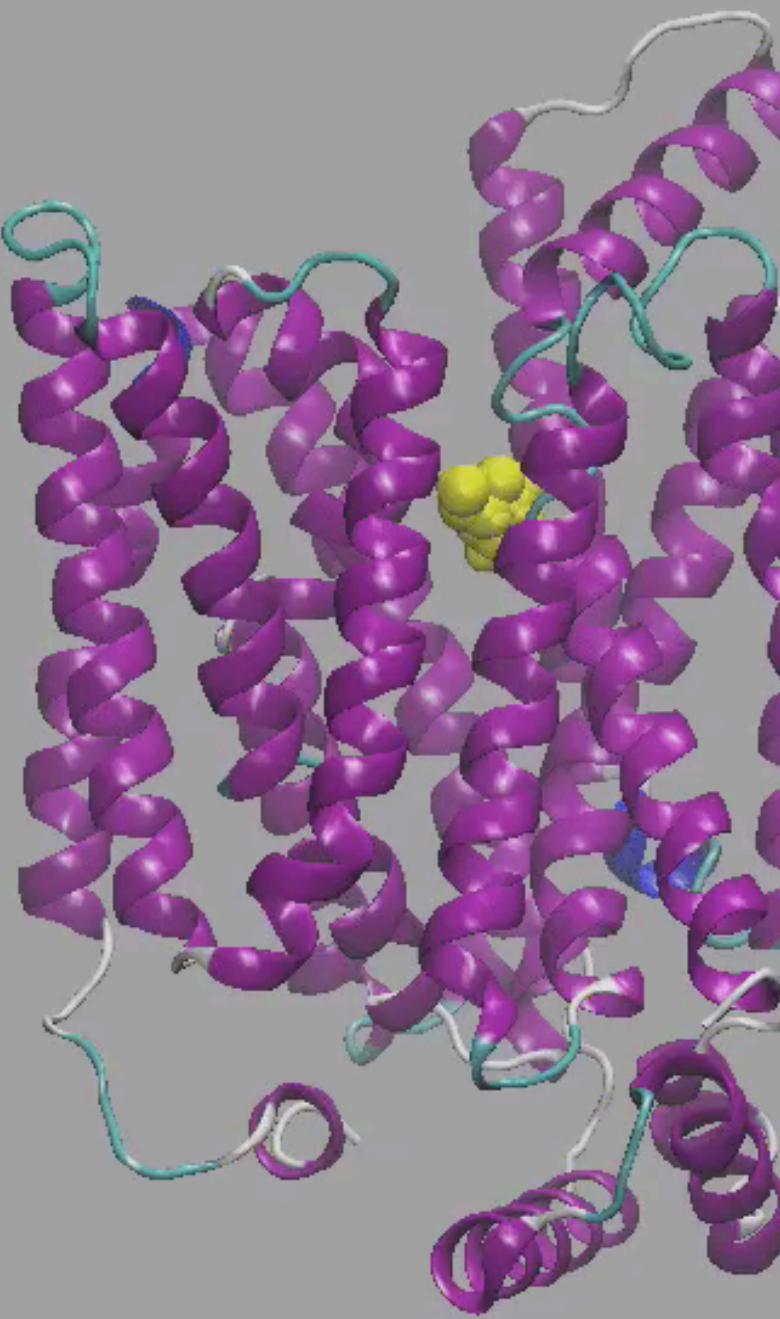
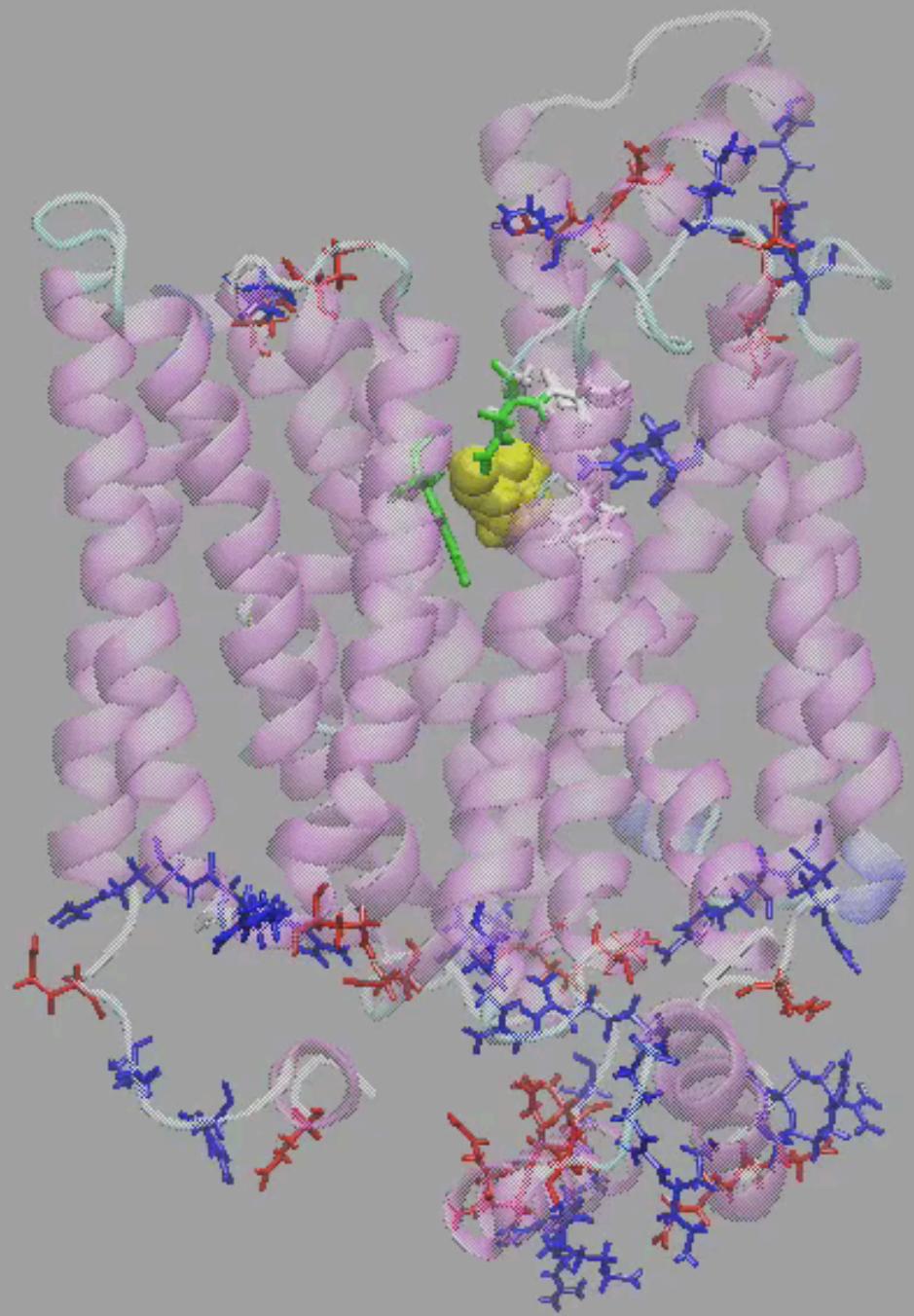
- Qu'est-ce qu'un pas d'intégration approprié ?



<u>Système</u>	<u>Types de mouvement</u>	<u>Pas d' intégration suggéré</u>
Molécules rigides	translation, rotation	5 fs
Molécules flexibles, Liaisons rigides	translation, rotation, torsions	2 fs
Molécules flexibles, Liaisons flexibles	Translation, rotation, torsions, vibrations	1 or 0.5 fs

Choix du pas d' intégration (2)

- Le pas d' intégration doit être plus petit que les mouvements de plus haute fréquence du système : $\delta t \ll \nu_{\max}$
- En “gelant” les mouvements de haute fréquence, des pas d' intégration plus importants peuvent être utilisés (gain en temps de calcul)
- Conditions:
 - Fréquences des mouvements gelés \gg fréquences des autres mouvements
 - Mouvements gelés faiblement couplés aux autres mouvements
- E.g. : longueurs de liaison
 - Pour les atomes d' Hydrogène: δt augmenté d' un facteur 2 (1 fs time step)
 - Toutes les liaisons : δt augmenté d' un facteur 3 (2 fs time step)
- Divers algorithmes proposés : SHAKE, LINCS, RATTLE, SETTLE...



- On oublie la relation temporelle entre les états
- On s'intéresse à la diversité des états accessibles dans des conditions thermodynamiques données

Notion de Mécanique Statistique

- Mécanique Statistique :
 - étude des systèmes macroscopiques d' un point de vue moléculaire
 - permet de faire le lien entre les simulations microscopiques (DM) avec des propriétés macroscopiques
- Quelques définitions:
 - état thermodynamique : état macroscopique dépendant de paramètres thermodynamiques (température, pression, nombre de particules...)
 - état microscopique : jeu de coordonnées et de vitesses pour un système donné (= un point dans l' espace de phase)
 - espace de phase : espace multidimensionnel (6N dimensions: 3N coordonnées, 3N vitesses) caractérisant l' ensemble des états microscopiques possibles d' un système
 - ensemble thermodynamique : ensemble des points dans l' espace de phase satisfaisant un certain état thermodynamique

Ensembles Thermodynamiques

- Les différents ensembles thermodynamiques
 - micro-canonique : ensemble NVE (N =nombre de particules, V =volume, E =énergie totale du système)
 - canonique : ensemble NVT (T = température)
 - isotherme-isobare : ensemble NPT (P = pression)
 - grand-canonique : ensemble μ VT (μ = potentiel chimique)
- L' ensemble NPT
 - devient le plus commun pour les dynamiques moléculaires de systèmes protéiques
 - nécessaire pour les simulations de lipides et de protéines membranaires

DM à température constante

- La température T est reliée à l'énergie cinétique E_{kin} du système et donc aux vitesses. D'après le théorème d'équipartition de l'énergie:

$$E_{kin} = \sum_{i=1}^{N_{atoms}} \frac{1}{2} m_i v_i^2 = \frac{k_b T}{2} (3N - N_c)$$

$N_c=3$ si on contraint le moment linéaire total
 $(3N-N_c)$ = nb de degrés de liberté

- La température peut être contrôlée en modifiant les vitesses v_i des atomes et molécules du système, e.g.:
 - Ajustement des vitesses : les vitesses sont ajustées de manière à avoir $T=T_{ref}$ à chaque pas
 - Pas de fluctuations de température !
 - Algorithme de Berendsen ("Weak coupling") : les vitesses sont ajustées à un taux proportionnel à la différence de température
 - Fluctuations de température, plus réaliste
- Autre algorithme: Nose-Hoover

DM à température et pression constantes

- Dynamique moléculaire standard : nombre de particules N, volume V et énergie totale E_{tot} du système sont conservés
$$E_{\text{tot}} = E_{\text{potential}} + E_{\text{kinetic}}$$
- Les approximations (cut-offs, intégration) peuvent mener à des instabilités et donc à des variations d'énergie et/ou de température
- Etudes de systèmes qui ne sont pas à l'équilibre (gradient P ou T)



DM à pression constante

- La pression est reliée au volume du système et aux interactions entre particules
- Gaz idéal : $PV = Nk_b T$
- Système réel : $PV = Nk_b T - \frac{1}{3k_b T} \sum_{i=1}^N \sum_{j=i+1}^N r_{ij} f_{ij}$
 - Le second terme est le viriel décrivant la contribution due aux forces entre particules
- La pression peut être contrôlée :
 - en faisant varier le volume du système et en ajustant la position des particules les unes par rapport aux autres
 - algorithme de Berendsen (similaire à celui pour la température)

Gestion des bords

monde réel \Leftrightarrow nombre de molécules \sim nombre d'Avogadro (6.02 10^{23}).

Peu de molécules sont influencées par les bords (e.g. parois, interface air-solvant...).



simulations par Ordinateur
 \Leftrightarrow nombre de particules limité (limites de temps et de mémoire).

Beaucoup de molécules seront à proximité des bords.

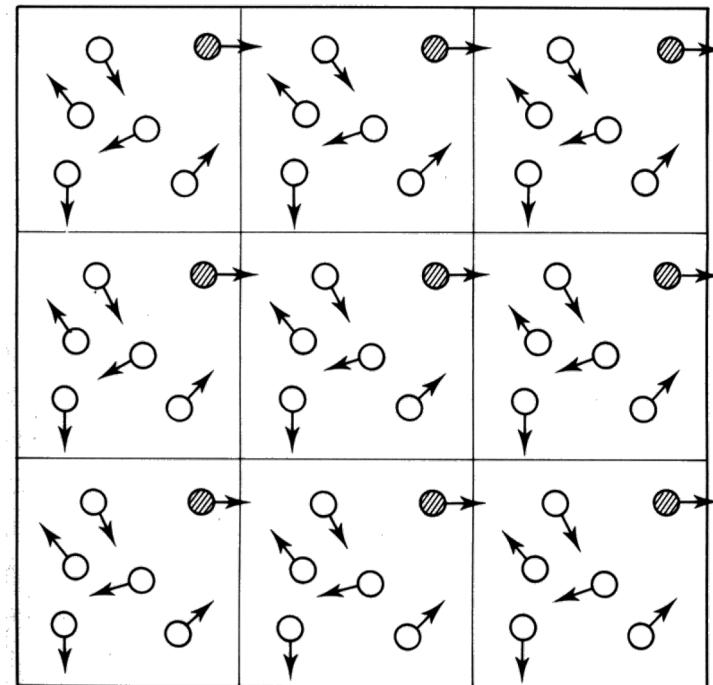
Pour pouvoir prédire les propriétés macroscopiques d'un solvant ("bulk properties") il est nécessaire de traiter le problème des bords.

Gestion des bords : vide

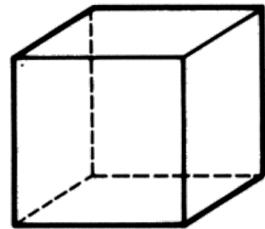
- Pas de solvant : interface avec le vide !
- Problèmes:
 - Simulations non réalistes
 - La tension de surface tendra à rendre le système plus compacte (sphérique) afin de diminuer la surface moléculaire
 - Déformations de molécules non sphériques e.g. ADN
 - Pas d' écrantage des charges par le solvant
- Solutions possibles
 - Modèle de solvant implicite (e.g. constante diélectrique dépendante de la distance)
 - Ajout de forces stochastiques pour mimer le solvant sur les atomes de surface
 - Simuler explicitement les molécules de solvant (solvant explicite)

Conditions périodiques aux limites

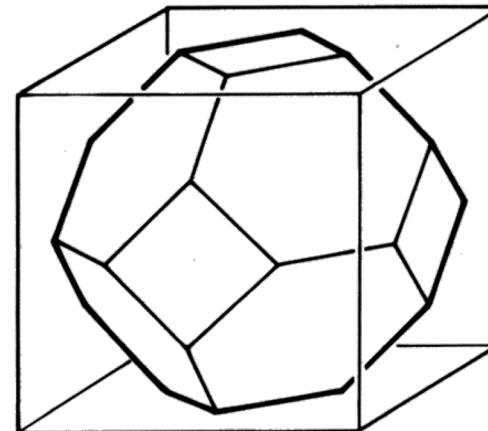
- La boîte de simulation est répliquée dans toutes les directions de l'espace
- Les particules quittant la boîte centrale sont réintroduites du côté opposé
- Les interactions sont calculées avec les particules des cellules voisines



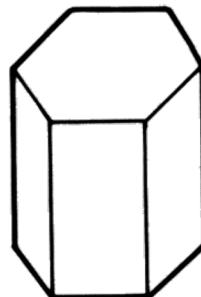
Formes de boîte



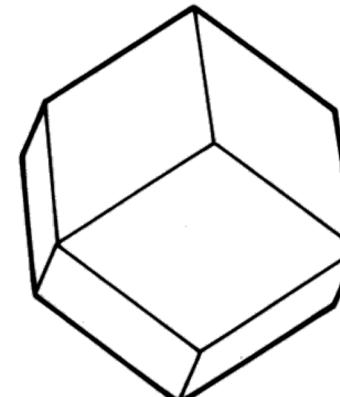
Cube



Truncated octahedron



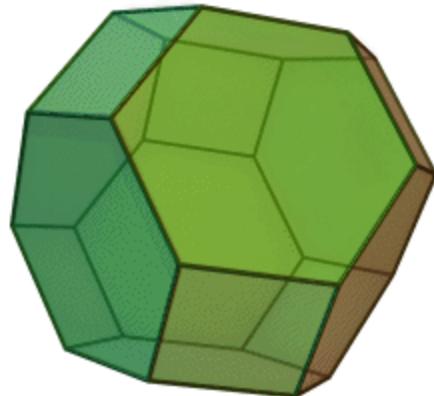
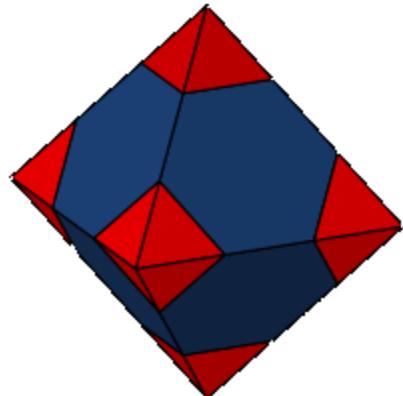
Hexagonal prism



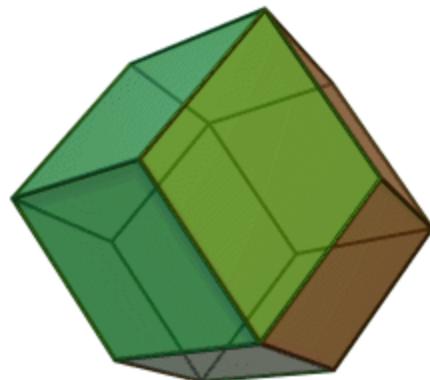
M2BI Rhombic dodecahedron

Formes de boîte (2)

Truncated octahedron



Rhombic dodecahedron



Etapes de dynamique moléculaire

- Sélection d'un modèle d'interaction
- Sélection des conditions frontières
- Sélection des conditions initiale (positions, vitesses . . .)
- Sélection de l'ensemble (NVE, NVT, NPT . . .)
- Sélection de la température, densité/pression .
..
- Sélection de l'intégrateur, thermostat, barostat
..
- Effectuer une simulation jusqu'à ce que l'équilibration soit atteint puis une phase de production pour obtenir des moyennes thermodynamiques, des positions, et des vitesses
- Analyse des résultats

Initialisation: positions initiales des particules, vitesses

Calcul des forces à partir des positions particules

Résolution des équations de Newton (intégration)

Analyse (écriture sur disque)

Mise au point du système

- Besoins
 - Coordonnées de départ (e.g. PDB, modèle par homologie, chaîne étendue, conformation aléatoire)
 - Boîte de solvant pré-équilibrée
- Démarrage de la dynamique moléculaire :
 - Coordonnées connues
 - Vitesses aléatoires à une température donnée (e.g distribution de Maxwell-Boltzmann)
 - Tous les paramètres nécessaires (cutoff, pas d' intégration, ...)

Importance des conditions initiales

- Le résultat d' une dynamique moléculaire dépend des conditions initiales :
 - Conformation de départ
 - Différents paramètres
 - Le jeu de vitesses initiales qui suit une distribution de Maxwell-Boltzmann:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{1/2} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right]$$

$p(v_{ix})$ = probabilité d' un atome i de masse m_i d' avoir une vitesse v_{ix} dans la direction x à la température T

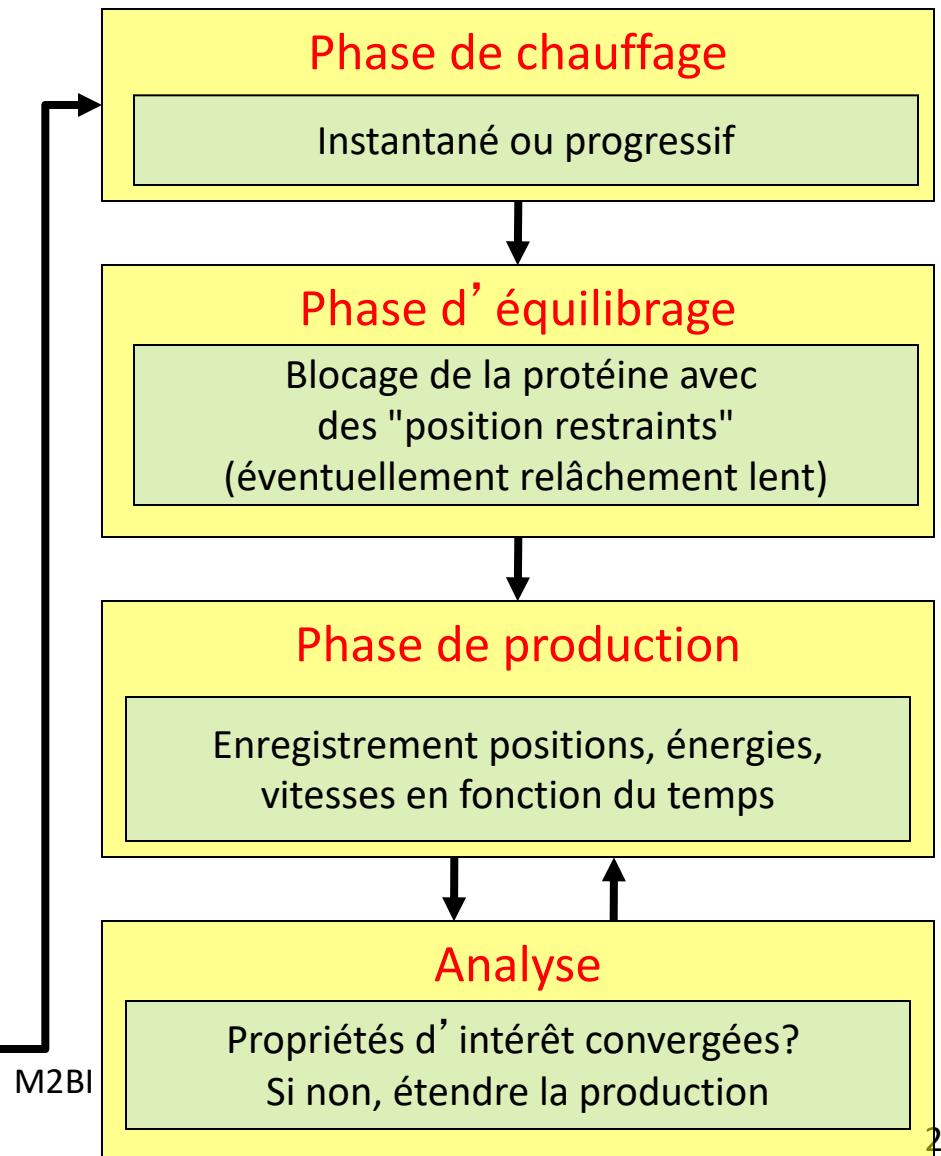
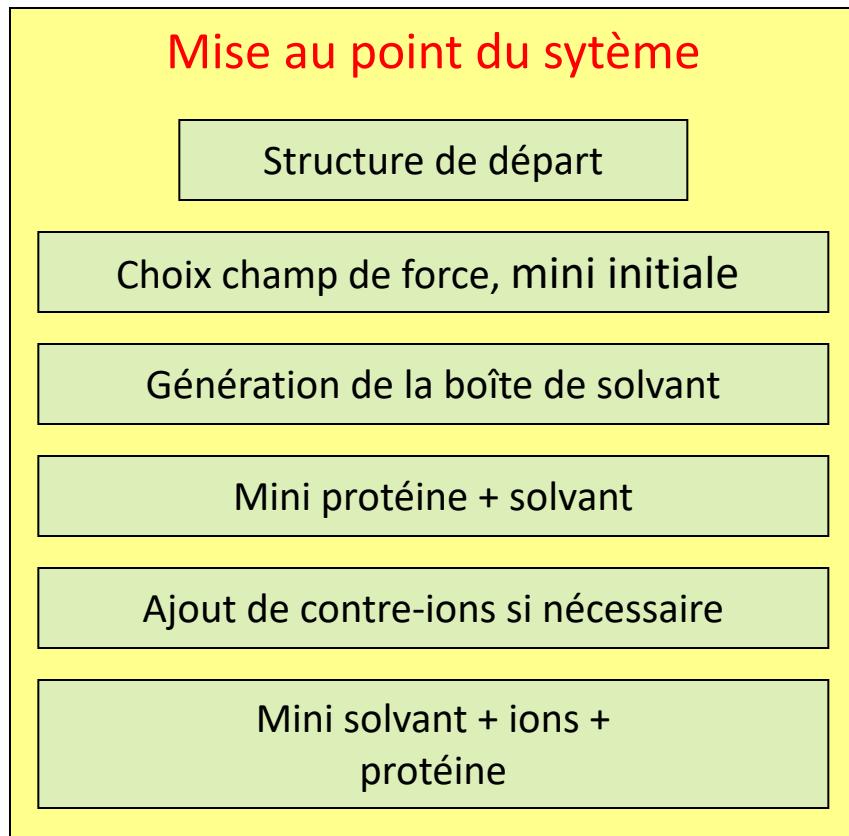
Distribution obtenue à partir d' un générateur de nombre aléatoire

- En changeant une de ces conditions, on explorera une autre zone de l' espace conformationnel (trajectoire différente)
⇒ Importance de la graine aléatoire

PLAN

1. Gestion des bords
2. Algorithme de Dynamique Moléculaire
3. Protocole de Dynamique Moléculaire
4. Analyse de trajectoire

Exemple de mise au point d'une simulation de protéine



PLAN

1. Gestion des bords
2. Algorithme de Dynamique Moléculaire
3. Protocole de Dynamique Moléculaire
4. Analyse de trajectoire

Séries temporelles

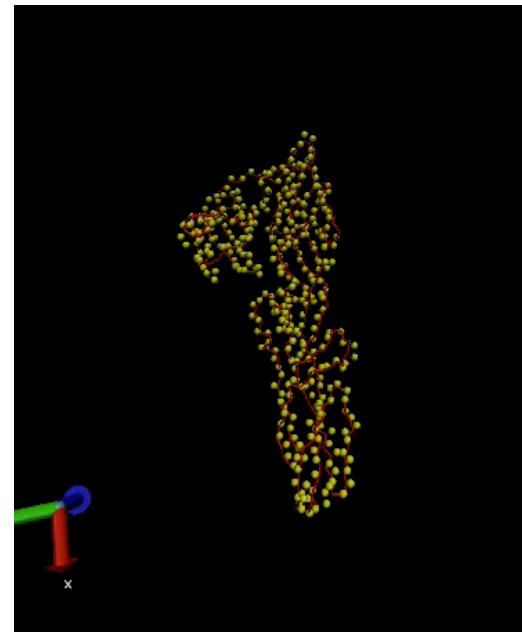
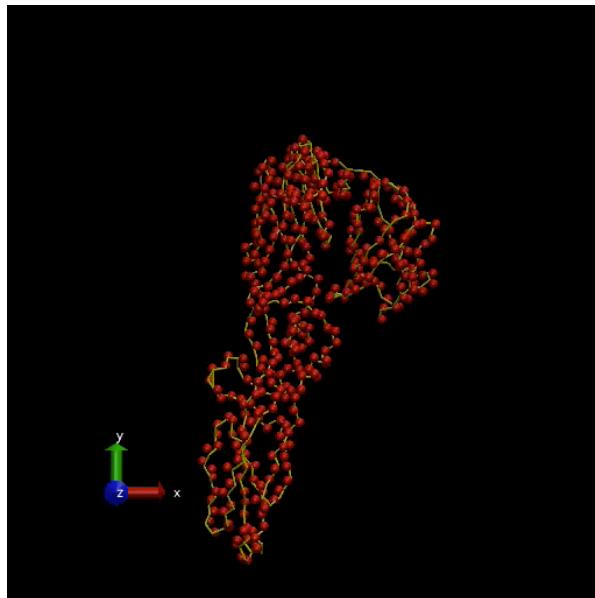
- Soit A une grandeur d'intérêt (e.g. une énergie, une distance...)
- Les variables dépendent des coordonnées et/ou des vitesses : $A(t) = f(r_1(t), \dots, r_N(t), v_1(t), \dots, v_N(t))$
- Séries temporelles : variable dynamique le long de la trajectoire

A(t):	f(0)	f(1)	f(2)	...f(nsteps)
t:	t=0	t=1	t=2	...t=nsteps

- Moyenne : $\langle A \rangle = \frac{1}{nsteps} \sum_{t=1}^{nsteps} A(t)$
- Fluctuations : $\sigma_A = \sqrt{\frac{1}{nsteps} \sum_{t=1}^{nsteps} (A(t) - \langle A \rangle)^2}$

Dengue E envelop

- CG (Martini) molecular dynamics: Monomer



Analyses de base (obligatoires!)

- Visualisation de la trajectoire (mouvements mis en jeu)
 - ngmx (outil GROMACS) : simple, peu flexible (pas très beau !)
 - VMD (<http://www.ks.uiuc.edu/Research/vmd/>) : plus poussé, meilleur graphisme, sans doute le plus adapté (lecture de trajectoire GROMACS, CHARMM, AMBER...), (plus joli !)
- Analyse des énergies du système
 - Énergie potentielle (décomposition en ses différents termes)
 - Énergies cinétique et totale ($E_{\text{tot}} = E_{\text{pot}} + E_{\text{kin}}$)
 - Température, Pression (Volume / Densité) selon l' ensemble thermodynamique
- Ces valeurs ont-elles convergées (fluctuations autour d' une valeur moyenne)?
 - Oui → on peut continuer l' analyse
 - Non → la ligne de base diverge : tenter d' en comprendre la raison (continuer la simulation ou l' abandonner et changer le protocole)

Exemple : simulation d' un peptide d' abductine

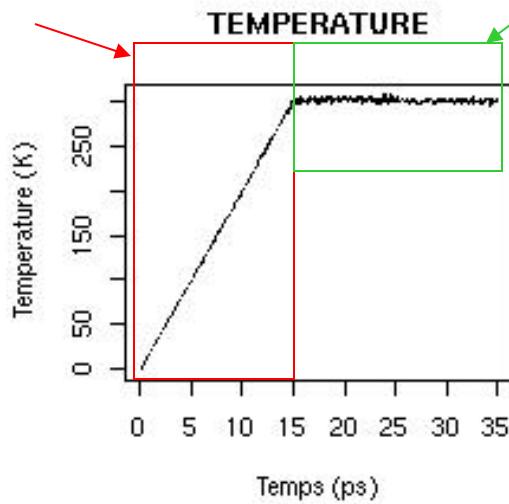
- Abductine :
 - protéine élastique (ligament de coquille St-Jacques)
 - Application possible dans le domaine des biomatériaux
 - Élasticité d' origine moléculaire
 - Séquence très riche en Gly, Phe, Met

1 MNAYICLAAC LIAAVSAAGY GGGAGSMGGT GGMGGGMNAG
41 GF GGGMGGGMG GGKGGFGGIG GF GGGMGGGMG GGPGGFGGMG
81 GF GGGMGGGKG GF GGGMGS GMG G**FGGMGGNA** **FGGMGGNA**
121 **G**F GGGMGGQGG FGGKGY

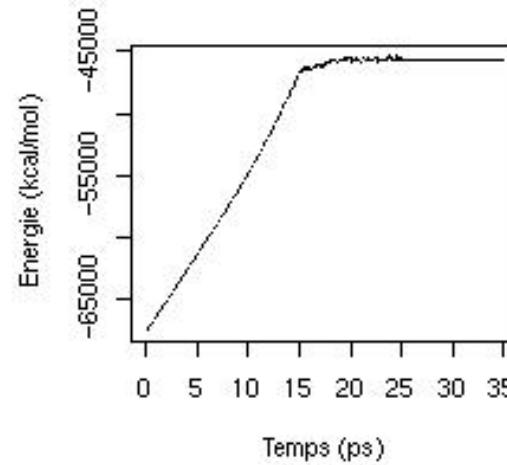
- Peptide représentatif simulé: FGGMGGGNAG
- Logiciel CHARMM, Ensemble NVE, T=300K
- Cutoff électrostatique (switch) et van der Waals (shift): 14 Å

Analyse des énergies

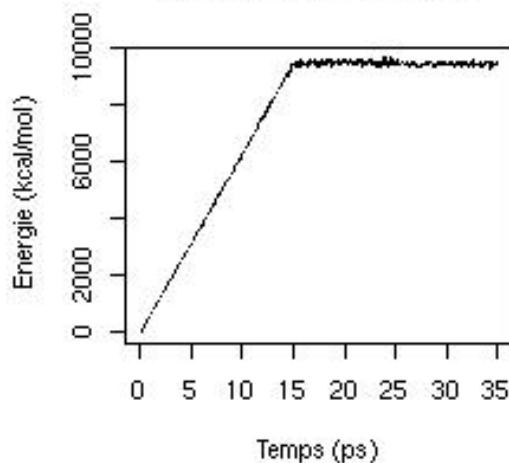
1) chauffage



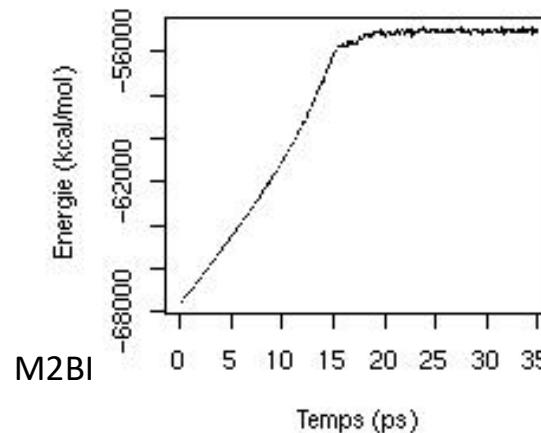
ENERGIE TOTALE



ENERGIE CINETIQUE

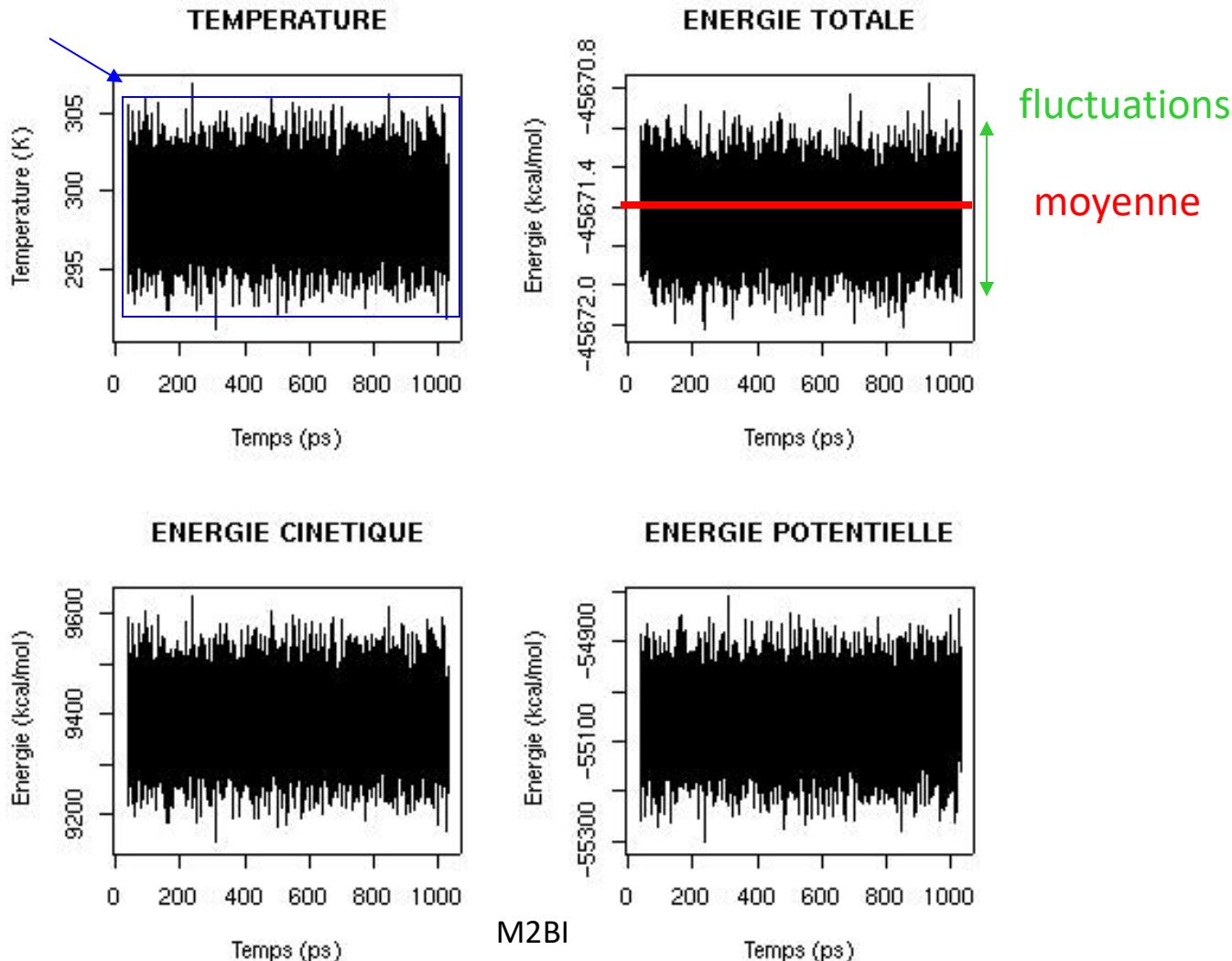


ENERGIE POTENTIELLE



Analyse des énergies (2)

3) production



Mouvements dans les protéines (1)

- Les mouvements dans les protéines couvrent un large spectre
 - d'énergie ($0.1 - 100 \text{ kcal.mol}^{-1}$)
 - de temps ($10^{-15} - 10^4 \text{ s}$)
 - d'amplitudes ($0.1 - 100 \text{ \AA}$ voire plus)
- Les mouvements sont interdépendants et couplés
 - Mouvements très rapides (vibrations de liaison)
 - Mouvements à petite échelle
 - Mouvements à moyenne échelle
 - Mouvements de grande ampleur

Root Mean Square Deviation

- Écart quadratique moyen (RMSD) :
 - caractérise la différence géométrique entre 2 conformations différentes (d'une même molécule ou d'un même système moléculaire) après alignement
- 2 phases :
 - Alignement optimal des 2 conformations
 - Calcul du RMSD :

$$RMSD(a, b) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left\| \vec{r}_i(a) - \vec{r}_i(b) \right\|^2}$$

N = nombre d' atomes

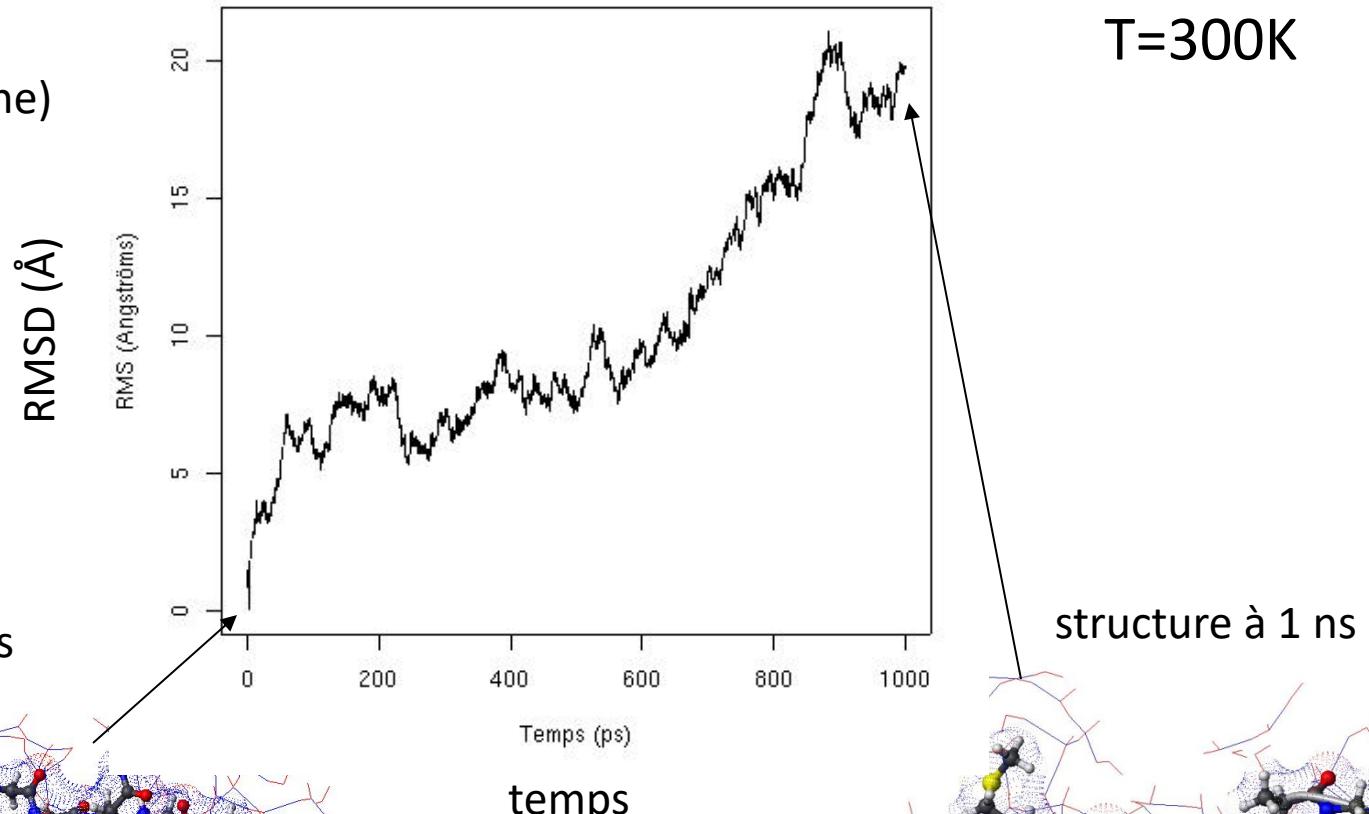
a et b = 2 conformations différentes

Vecteur r_i (a) = position de l' atome i dans la conformation a

vecteur r_i (b) = position de l' atome i dans la conformation b

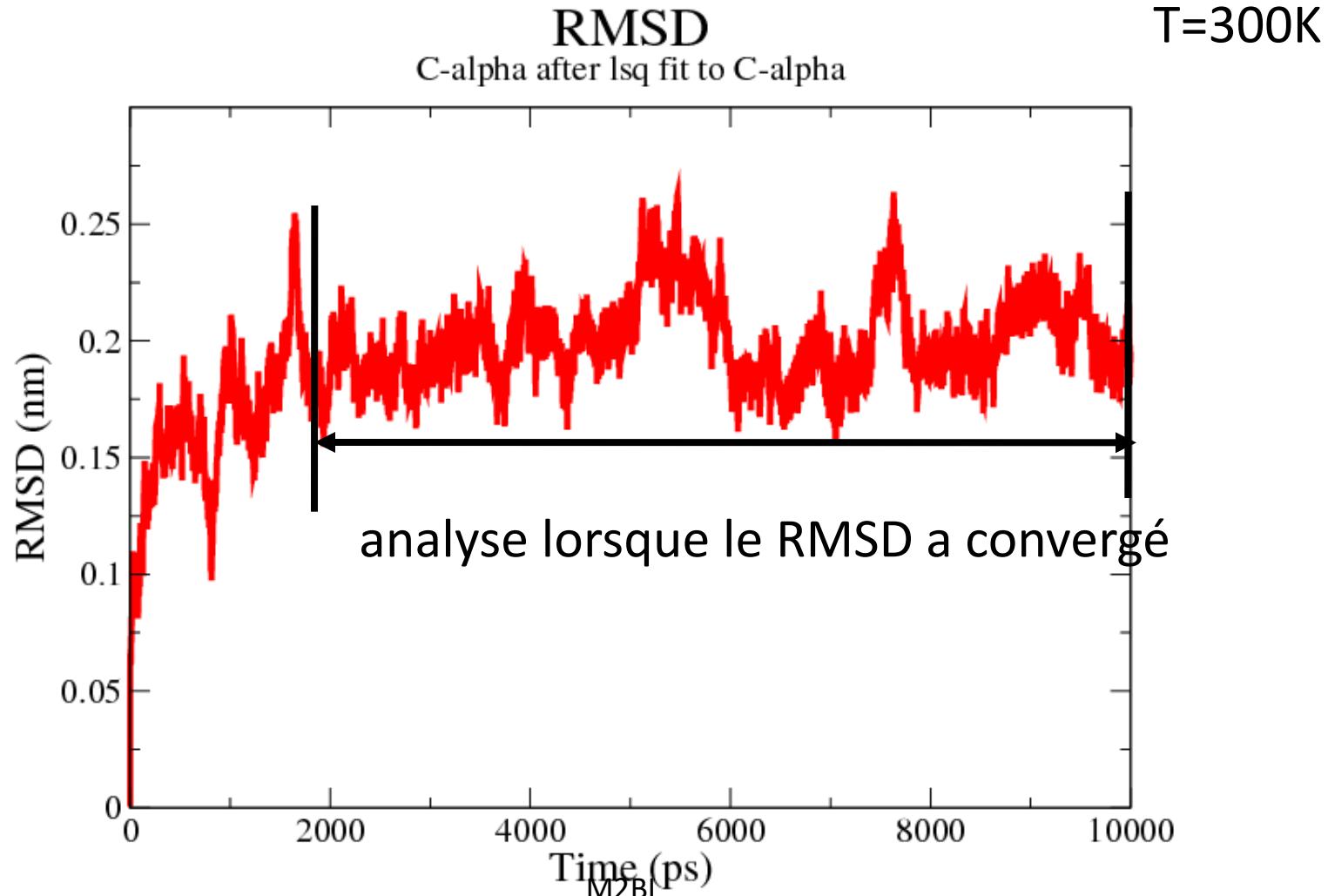
Évolution du RMSD par rapport à la conformation de départ : 1) Peptide

(peptide d'abductine)



M2BI

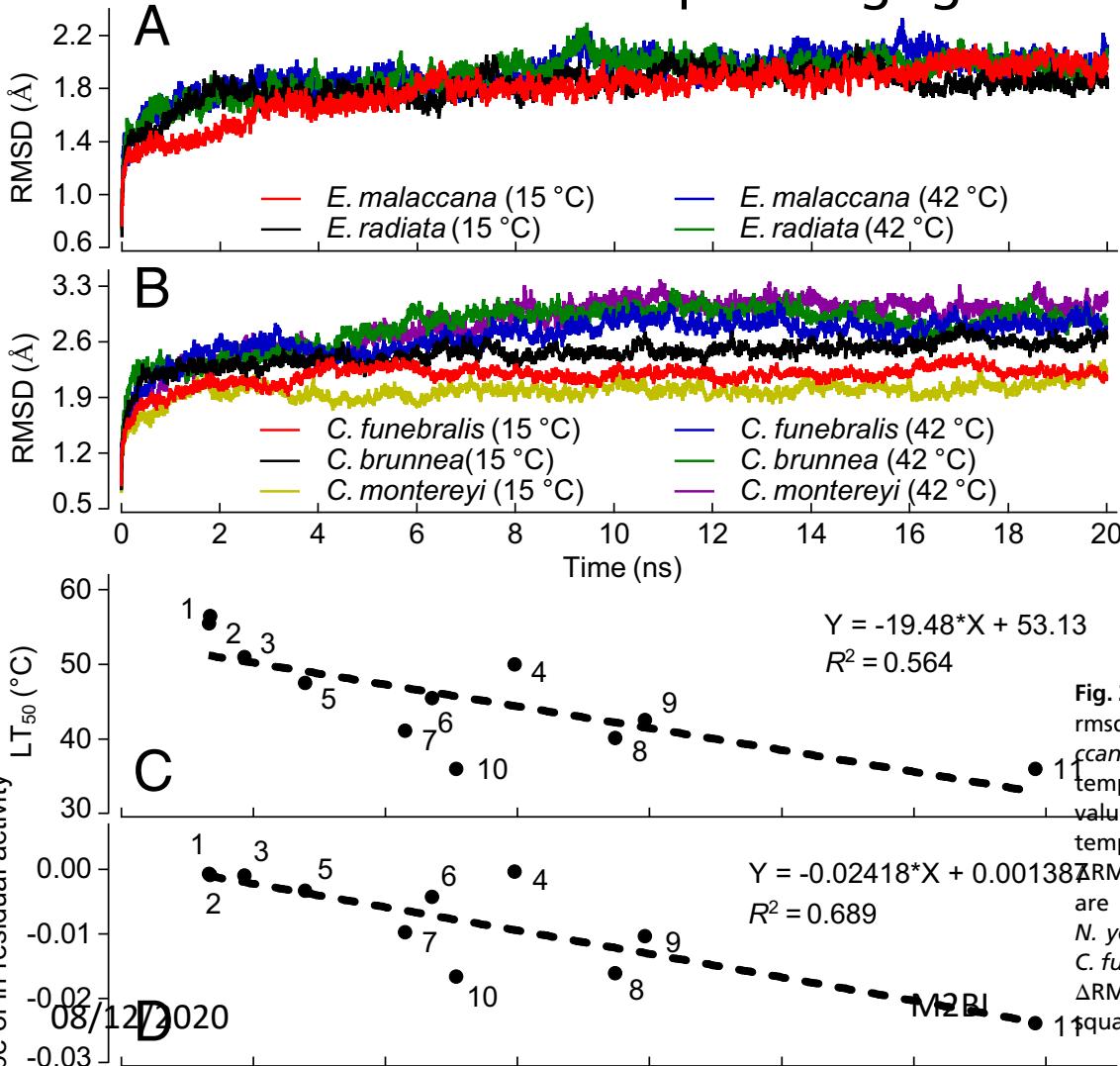
Évolution du RMSD par rapport à la conformation de départ : 2) Protéine



Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs

Yun-wei Dong^{a,1}, Ming-ling Liao^a, Xian-liang Meng^b, and George N. Somero^{c,1}

www.pnas.org/cgi/doi/10.1073/pnas.1718910115



Significance

Analysis of structural and functional properties of cytosolic malate dehydrogenase (cMDH) orthologs of five genera of marine molluscs adapted to an ~60 °C range of temperatures reveals the role of protein flexibility in adaptation to temperature. Molecular dynamics simulation (MDS) analysis reveals protein-wide as well as local adaptation in flexibility. Sequence regions involved in binding and catalysis show significant interspecific, temperature-related differences in flexibility. MDS analysis is shown to provide a powerful means of examining adaptive change in protein evolution at different temperatures.

Fig. 2. MDS analysis of backbone atom movements in cMDHs. (A and B) The rmsd of backbone atom positions for cMDHs from *Echinolittorina malaccana*, *E. radiata*, *C. funebralis*, *C. brunnea*, and *C. montereyi* at simulation temperatures of 15 and 42 °C ($n = 3$). (C) The Δ RMSD (the difference of rmsd value between 42 and 15 °C) over the equilibration state (10–20 ns) vs. lethal temperature (LT₅₀) for cMDH orthologs of 11 species of molluscs. (D) The Δ RMSD vs. the rate of thermal denaturation for 11 cMDH orthologs. Species are numbered as follows: 1, *E. malaccana*; 2, *E. radiata*; 3, *N. albicilla*; 4, *N. yoldii*; 5, *L. keenae*; 6, *L. scutulata*; 7, *L. austrodigitalis*; 8, *L. digitalis*; 9, *C. funebralis*; 10, *C. brunnea*; and 11, *C. montereyi*. The relationship between Δ RMSD and LT₅₀ or rate of thermal denaturation was analyzed by a least-squares linear regression analysis model.

Fluctuations atomiques

- Positions atomiques moyennes

$$\langle \vec{r}_i \rangle = \frac{1}{nsteps} \sum_{t=1}^{nsteps} \vec{r}_i(t) \quad (\text{avec } nsteps = \text{temps total de simulation})$$

- Fluctuations atomiques (Root Mean Square Fluctuations)

$$RMSF_i = \sqrt{\frac{1}{nsteps} \sum_{t=1}^{nsteps} \left\| \vec{r}_i(t) - \langle \vec{r}_i \rangle \right\|^2}$$

$$RMSF_i = \sqrt{\frac{3B_i}{8\pi^2}} \quad (\text{avec } B_i = \text{facteur B de cristallographie})$$

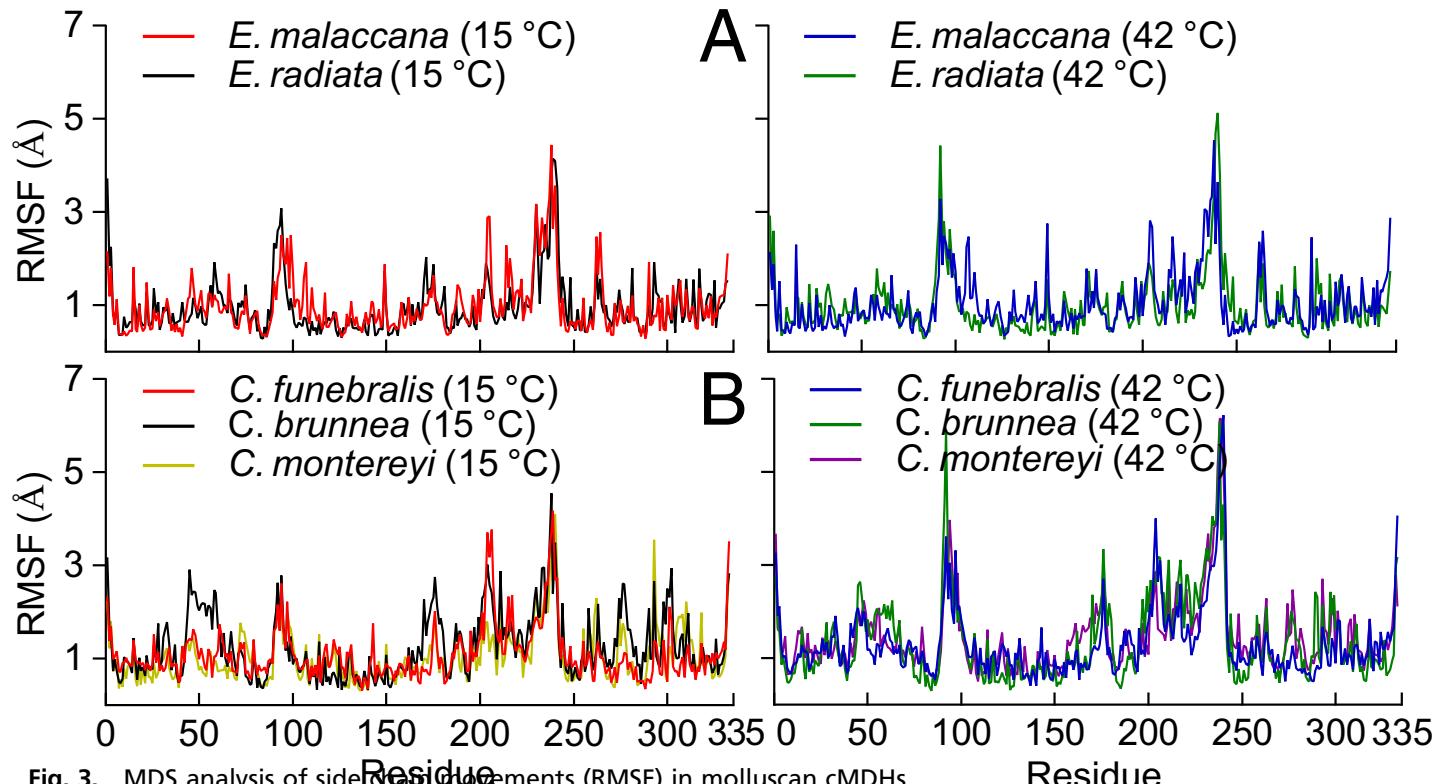
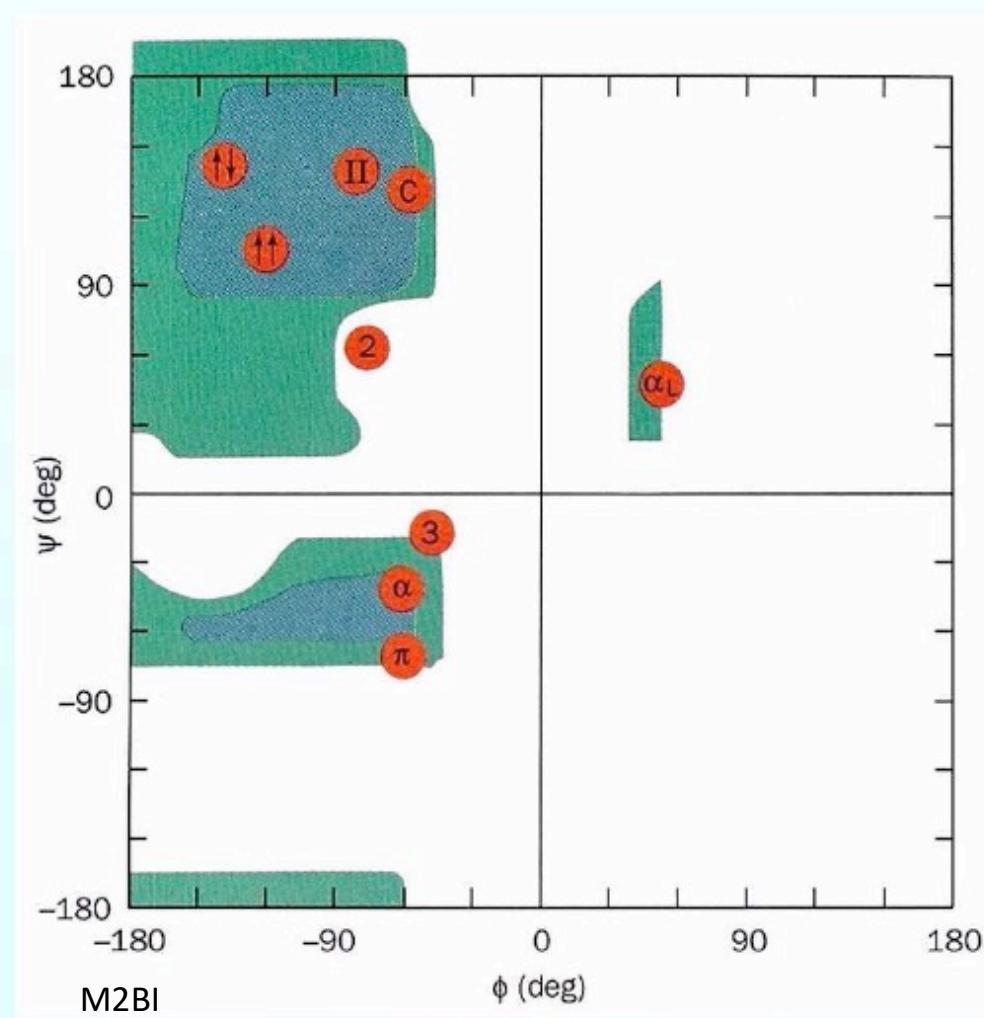
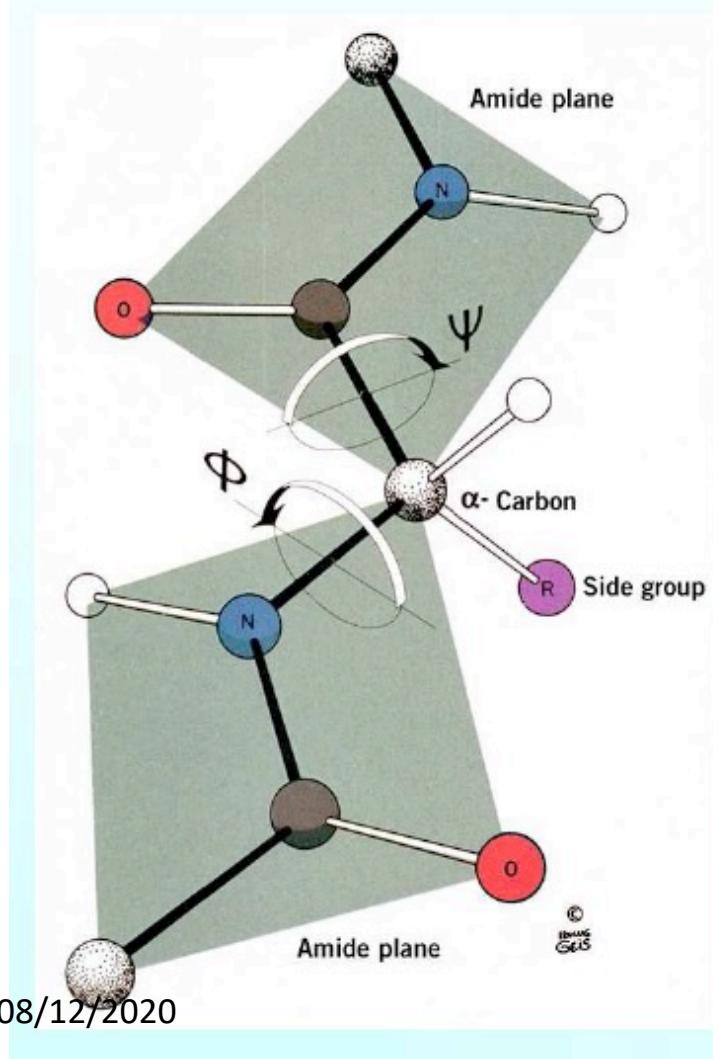


Fig. 3. MDS analysis of side-chain movements (RMSF) in molluscan cMDHs. RMSF spectra for congeners of *Echinolittorina* (A) and *Chlorostoma* (B) at simulation temperatures of 15 and 42 °C ($n = 3$). Values for average RMSF across the full sequence are given in Table S1. (C and D) Number of residues showing a change in RMSF ($\Delta\text{RMSF} > 0.5 \text{ \AA}$) vs. the species' lethal temperature (LT_{50}) (C) and the rate of thermal denaturation (D) for the cMDH orthologs. (E and F) Number of residues occurring in MR regions (residues 90–105 and 230–245) showing $\Delta\text{RMSF} > 0.5 \text{ \AA}$ vs. LT_{50} (E) and the rate of thermal denaturation (F) for the cMDH orthologs. The numbers assigned to each species are given in the Fig. 2 legend. The relationships between the number of residues showing significant change and the LT_{50} or rate of thermal denaturation were analyzed by least-squares linear regression.

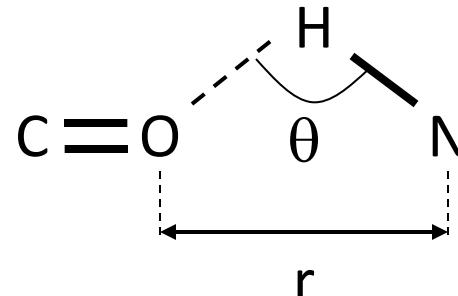
Angles dièdres ϕ/ψ et carte de Ramachandran (angles χ)



Liaisons hydrogène

- Critères géométriques:

- $r < 3.5 \text{ \AA}$
- $\theta > 120^\circ$



- Critères énergétiques

- modèle purement électrostatique C(+q1), O(-q1), N(+q2), H(-q2)

$$E = q_1 q_2 (r(ON)^{-1} + r(CH)^{-1} - r(OH)^{-1} - r(CN)^{-1}) \cdot f$$

($f \sim 332$ facteur de conversion en kcal.mol^{-1} , $q_1 = 0.42 \text{ e}$, $q_2 = 0.20 \text{ e}$)

- Liaison H si : $E < -0.5 \text{ kcal.mol}^{-1}$
- Critères utilisés dans DSSP (voire suite)

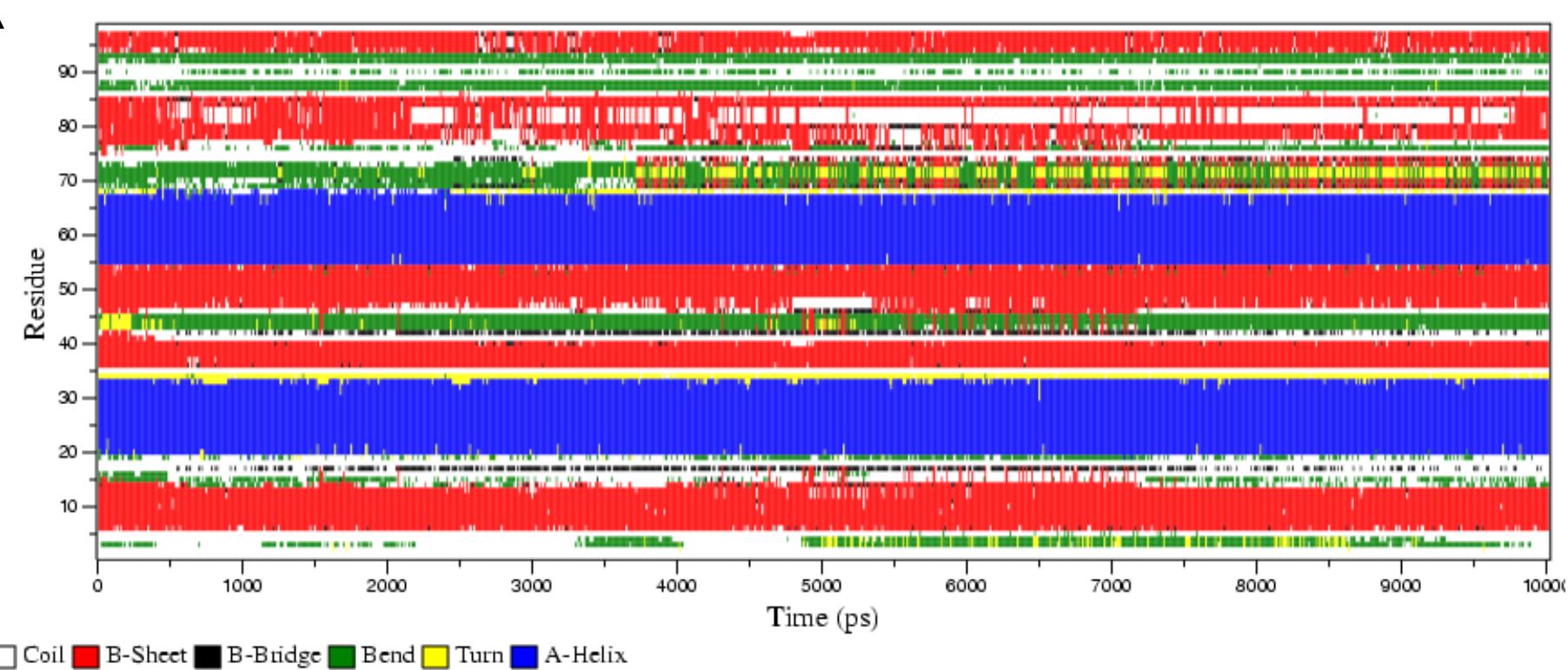
Structures secondaires

- Évolution des structures secondaires au cours du temps
- Algorithmes d' évaluation des structures secondaires:
 - DSSP (Kabsch & Sander): liaisons H
 - STRIDE (Frishman & Argos): liaisons H et ϕ/ψ
 - P-SEA (Labesse et al.): critères de distances C α -C α
 - P-Curves (Sklenar et al.) : paramètres hélicoïdaux
 - etc...

Évolution des structures secondaires

séquence

Secondary structure



Rayon de giration

- Décrit la compacité d' une molécule

$$Rgyr = \sqrt{\frac{1}{N} \sum_{i=1}^N \left\| \vec{r}_i - \vec{r}_{CM} \right\|^2}$$

N = nombre d' atomes
vecteur r_i = position de l' atome i

- Position du centre de masse

$$\vec{r}_{CM} = \frac{1}{M} \sum_{i=1}^N m_i \vec{r}_i$$

N = nombre d' atomes
 m_i = masse de l' atome i

$$M = \sum_{i=1}^N m_i$$

vecteur r_i = position de l' atome i
M = masse de la molécule

Accelerated Molecular Dynamics Simulations of Protein Folding

Journal of Computational Chemistry 2015, 36, 1536–1549

Yinglong Miao,^{*[a]†} Ferran Feixas,^{*[b,c]†} Changsun Eun,^{*[a]†} and J. Andrew McCammon^[a,b,c]

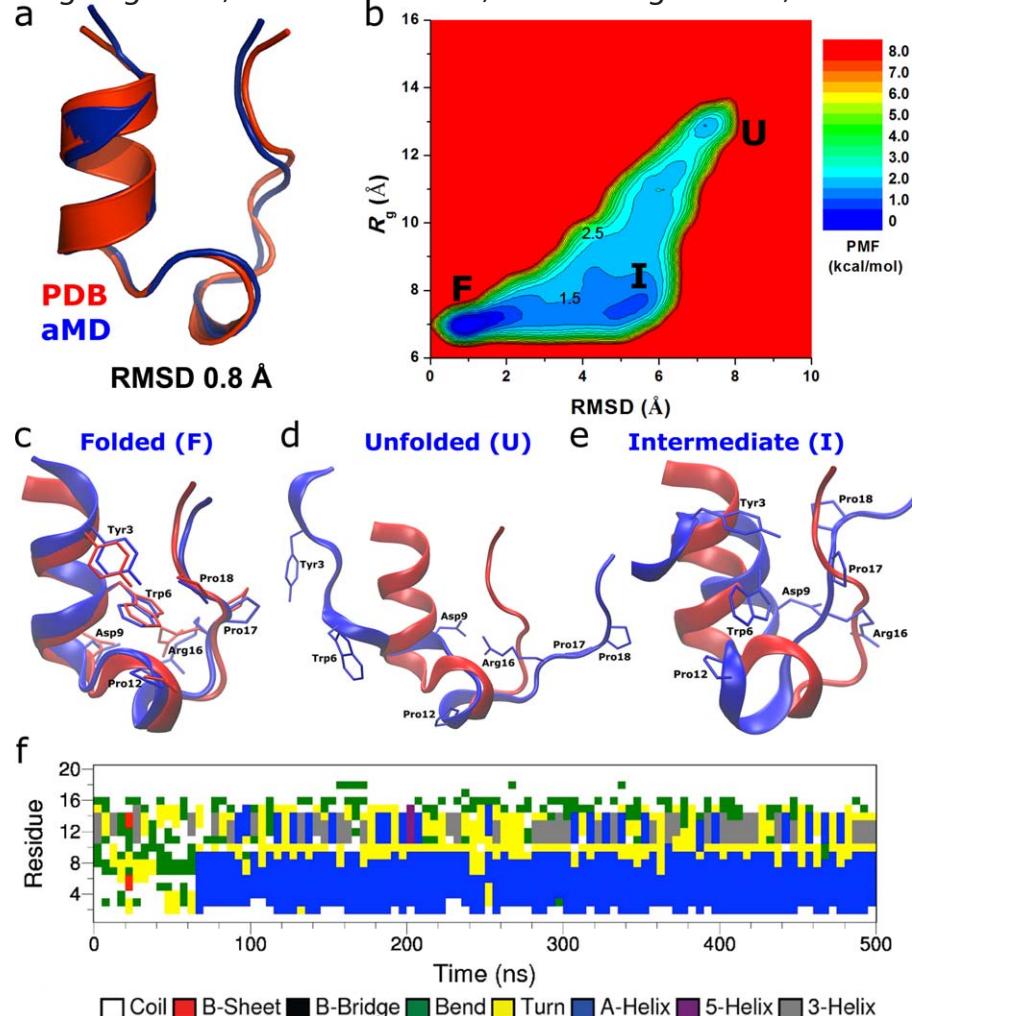


Figure 2. Folding of Trp-cage simulated via aMD: a) comparison of simulation-folded Trp-cage (blue) with the PDB (2JOF) native structure (red) that exhibits 0.8 Å RMSD at $t = 120$ ns in Sim1 (Fig. S4), b) two-dimensional (RMSD, R_g) free energy profiles calculated by reweighting the four 500 ns aMD simulations combined, structural representations of the c) folded ("F"), d) unfolded (U) and e) intermediate (I) states (blue) aligned to the native structure (red), and f) time evolution of the protein secondary structure during the 500 ns aMD simulation containing the folded structure shown in (a), in which the α -helix (blue) starts to form at ~ 65 ns.

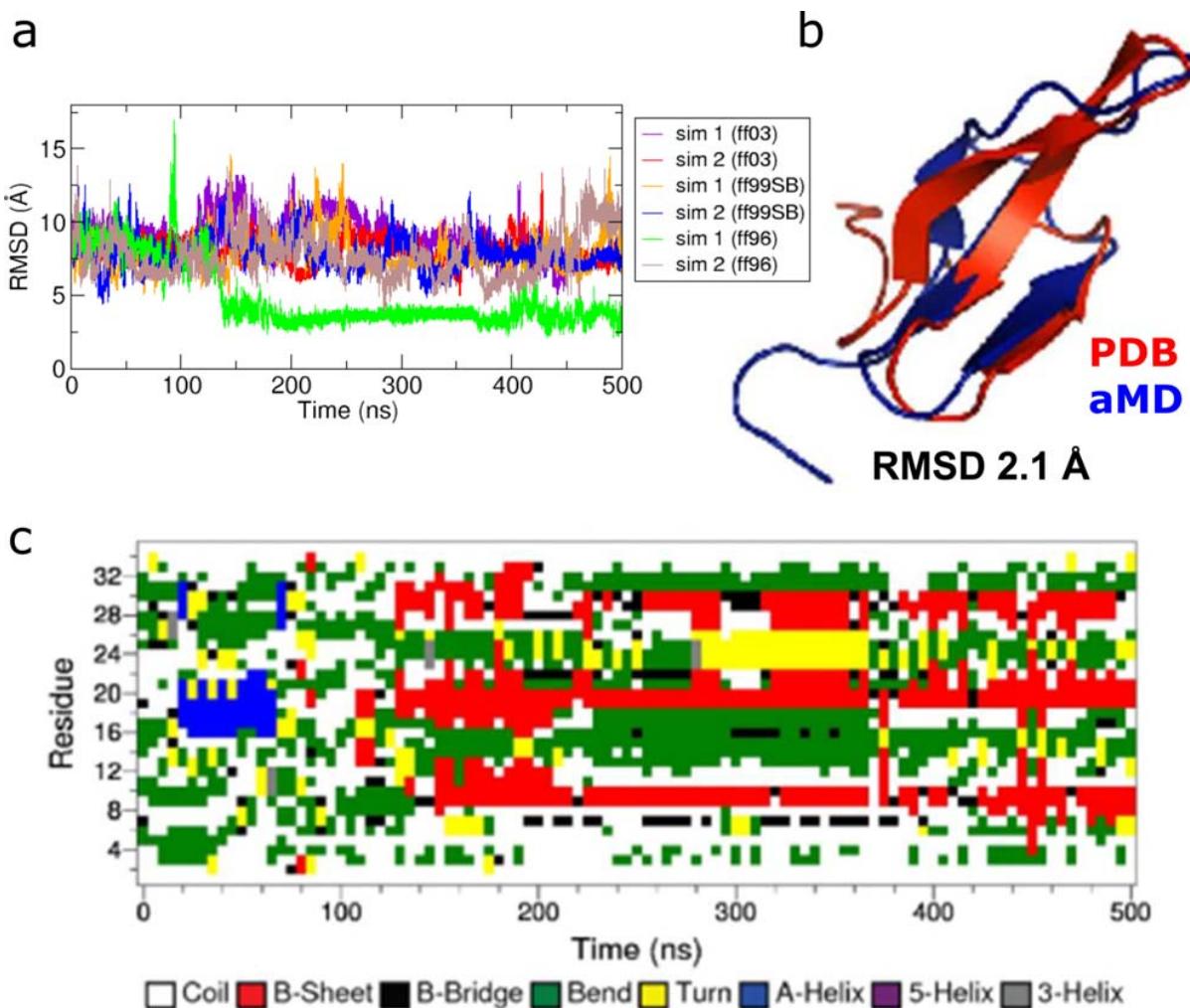


Figure 4. Folding of WW domain simulated via aMD: a) RMSD plots for the simulations with different AMBER force fields, b) comparison of simulation-folded WW domain (cyan; blue) with the PDB (2F21) native structure (red) that exhibits 2.1 Å RMSD for the residues between 5 and 30 out of total 35 residues at $t = 432$ ns, c) time evolution of the protein secondary structure during the 500 ns simulation run (Run 1) containing the folded structure shown in b). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

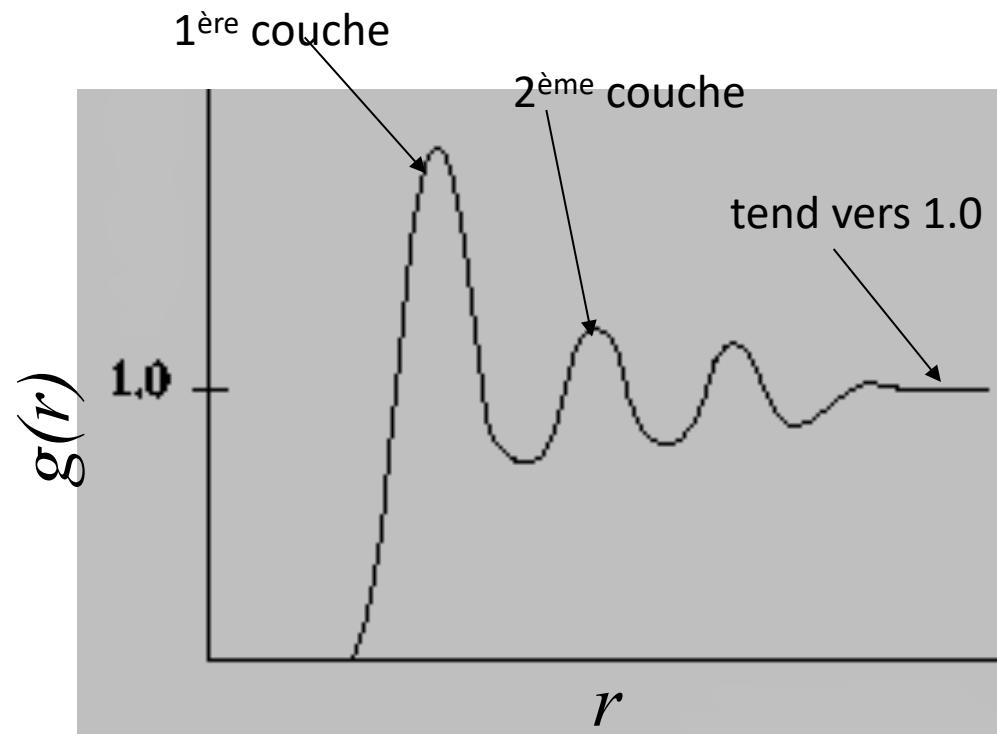
Distribution radiale

- Définit la position des particules les unes par rapport aux autres (mesurable par diffraction de neutrons ou de rayons X)

densité moyenne à une distance r d'une particule donnée

$$g(r) = \frac{\rho(r)}{\rho}$$

densité moyenne du fluide



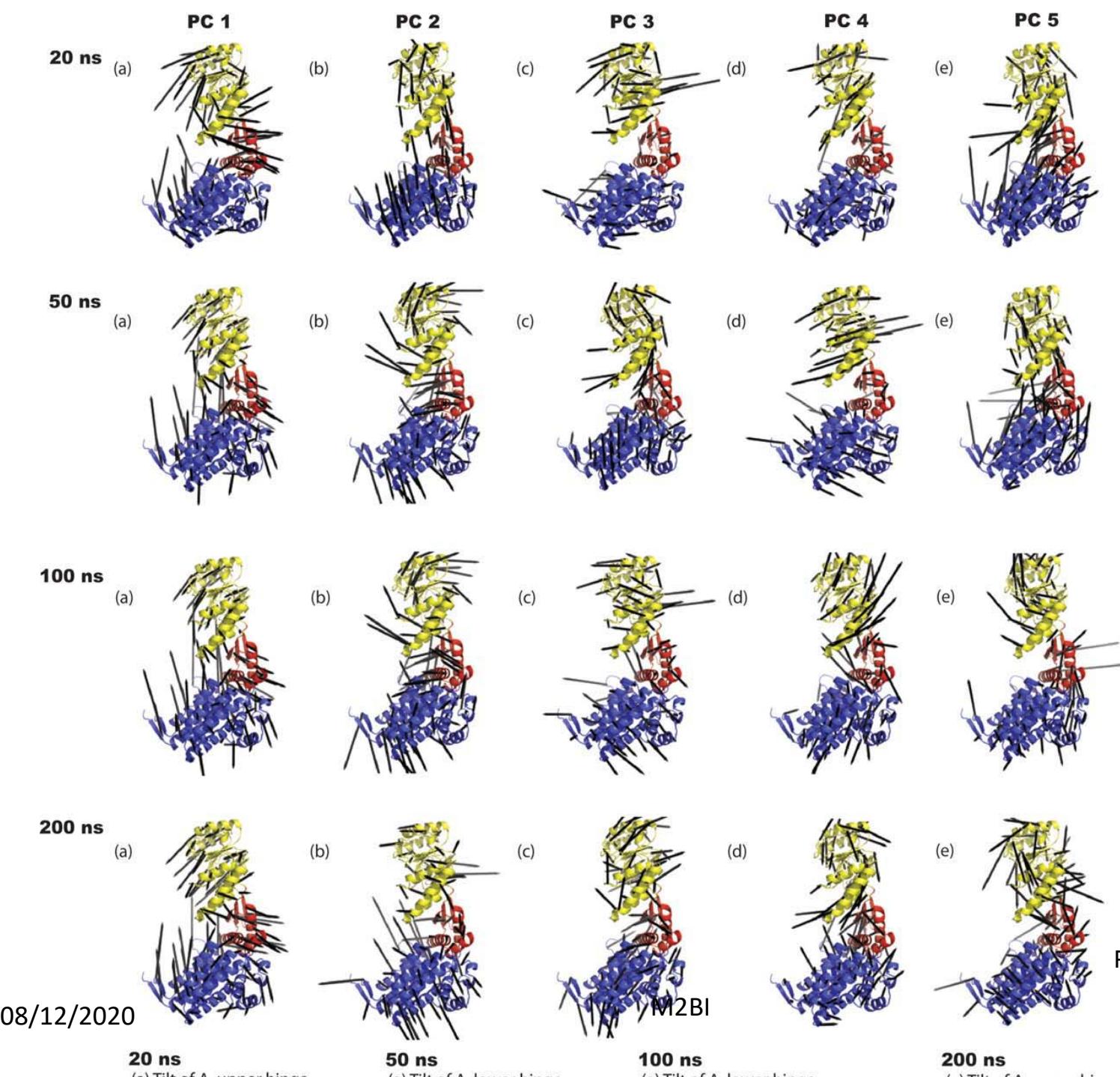
Analyses en composantes principales

- Etape 1: Calcul de la matrice de covariance C:

$$c_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle,$$

- Etape 2: Diagonalisation de cette matrice

$$\mathbf{A}^T \mathbf{C} \mathbf{A} = \lambda,$$



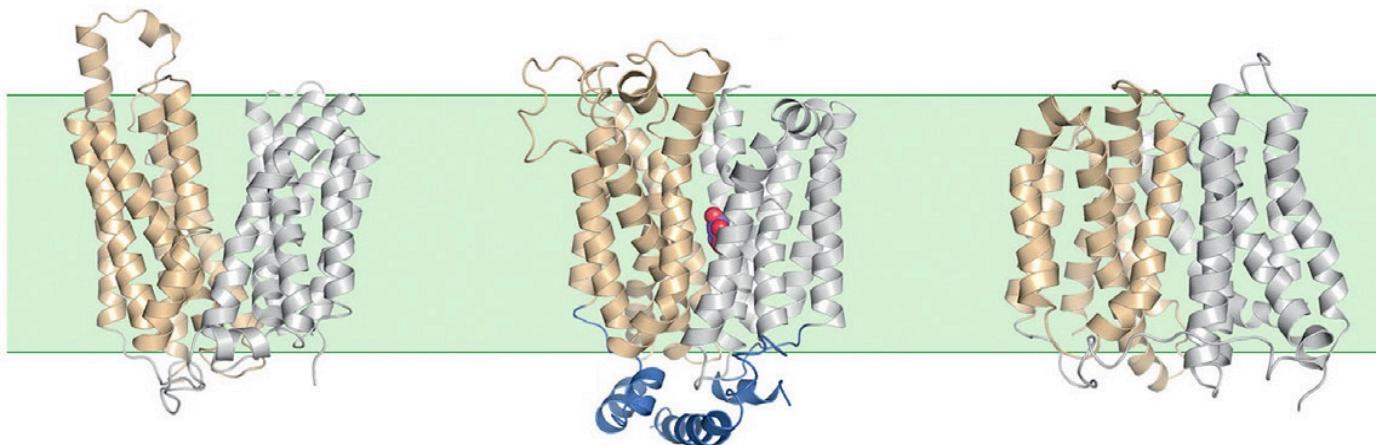
An example

Glucose Transporter GluT1

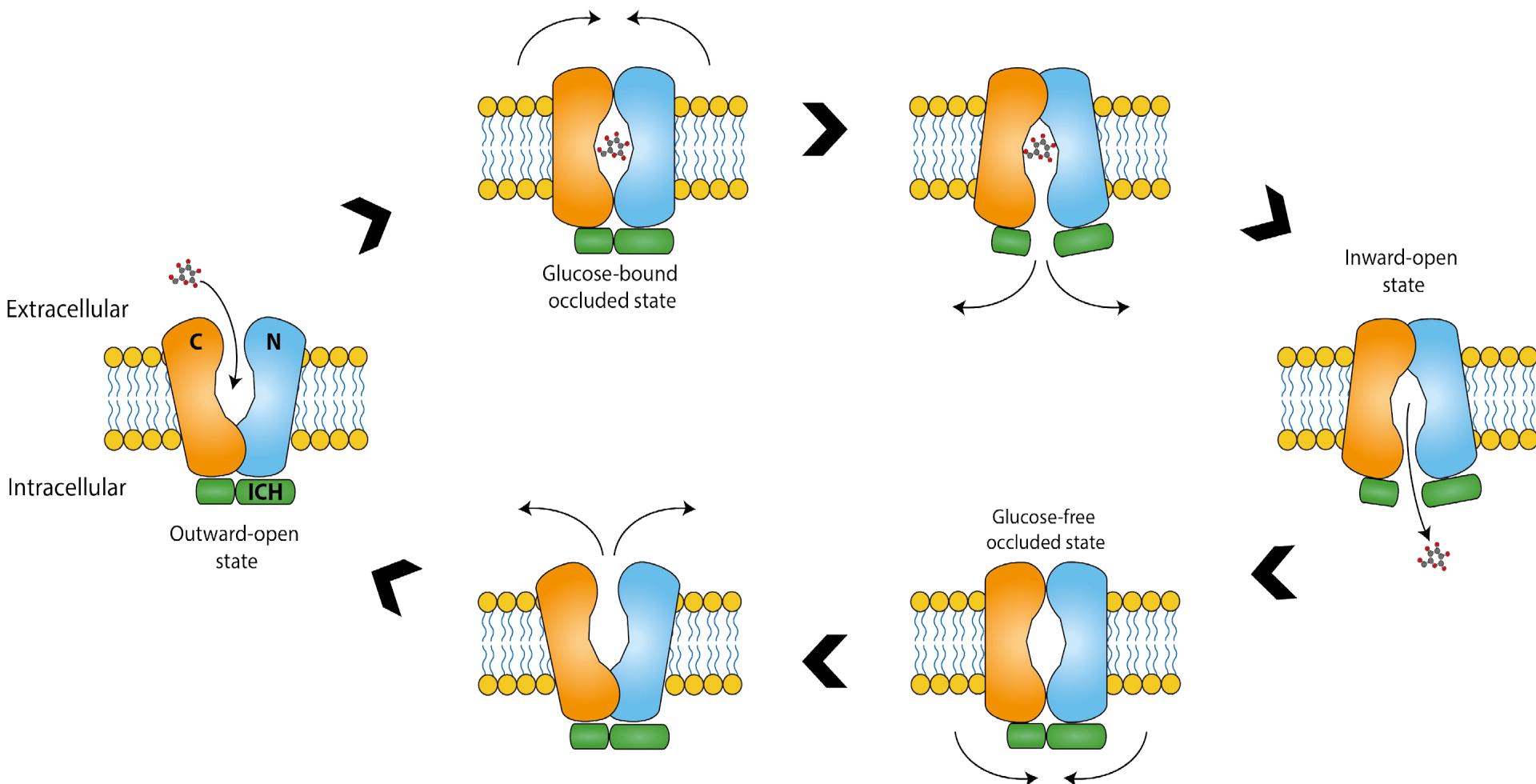
- First characterized glucose transporter
- Mediates cellular uptake of glucose into erythrocytes and endothelial cells of the blood-brain barrier
- Elevated expression levels of GLUT1 in **malignant cells**

Major facilitator superfamily

- One of the largest families of membrane transporters
 - Present ubiquitously in bacteria, archaea, and eukarya
 - Include members that can function as uniporters, antiporters, or symporters
 - Catalyze the transport of a wide range of compounds
 - All members share a common structural fold: **MFS fold**
- 12 TMs** organized into **2 discretely folded domains**, the N and C domains, each containing 6 consecutive TMs

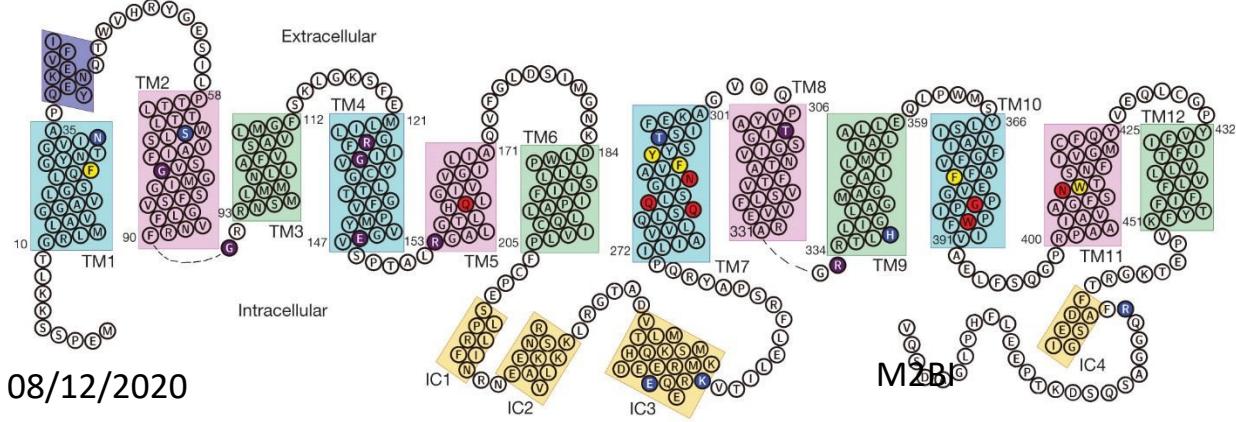
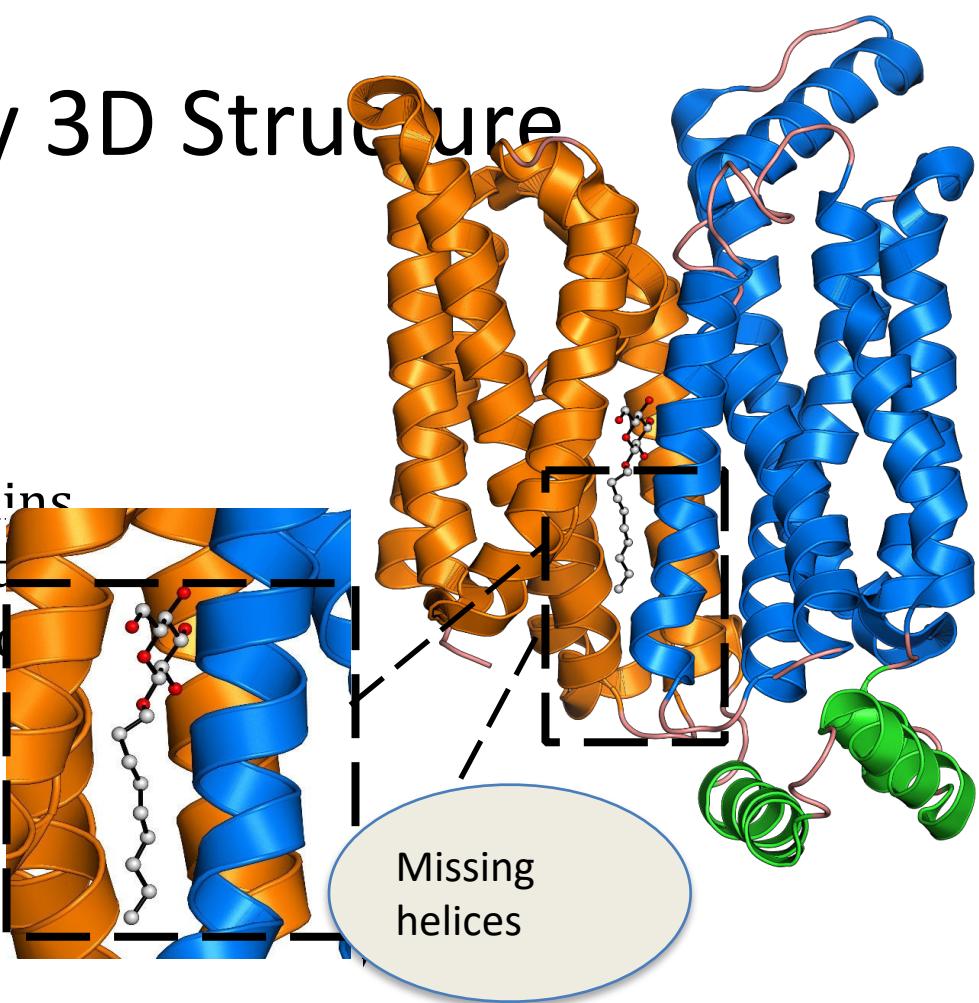


Model of solute transport



Atomic X-ray 3D Structure

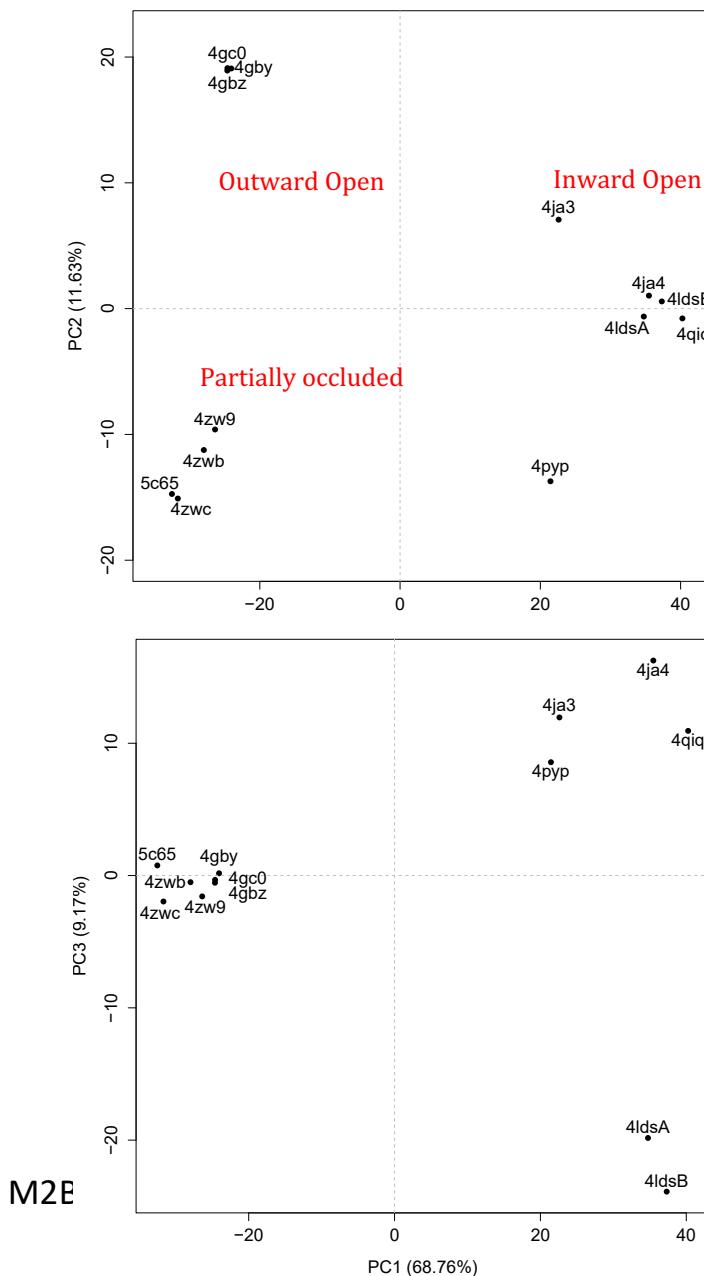
- 446 residues /492
- 2 mutations: N45T, E329Q
- 12 transmembrane helices organized into two distinct domains N and C, plus 3 short helices IC1 to IC3 located on the cytoplasmic side
- Ligand: β -nonylglucoside
- Inward-open state

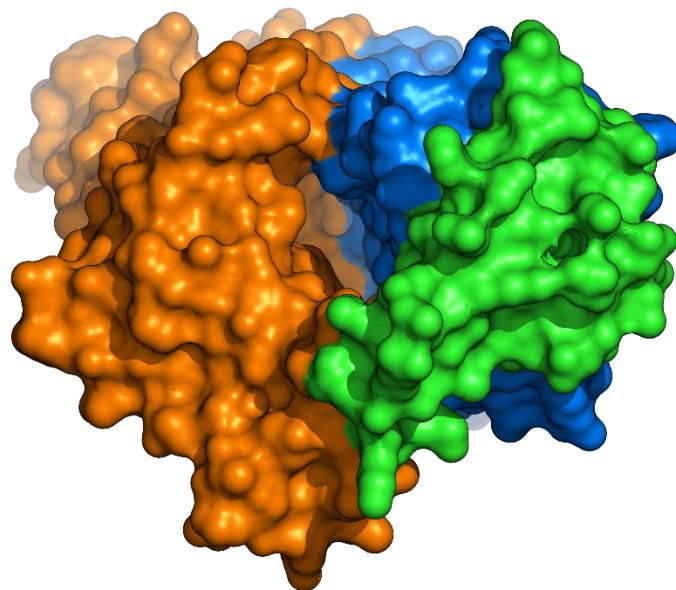
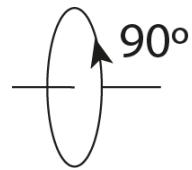
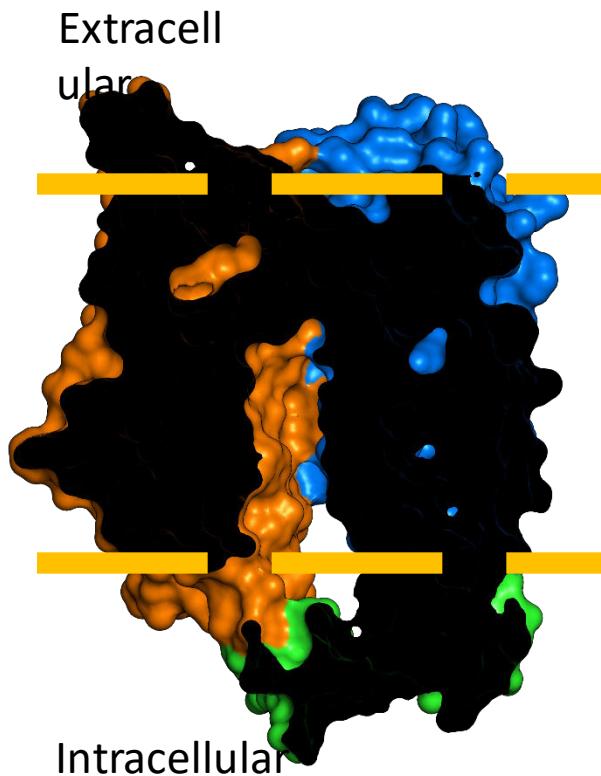


Principal components analysis of X-ray structure

Protein	Function	PDB ID(s)	Resolution (Å)	Conformation	Ligand in cavity
GLUT1	D-glucose transporter	4PYP	3.17	Inward open	β-Nonylglucoside
GLUT3	D-glucose transporter	5C65	2.65	Outward open	None
		4ZW9	1.50	Partly occluded, outward-facing	β-D-Glucose; α-D-Glucose; 1-Oleoyl-R-glycerol
		4ZWB	2.40	Partly occluded, outward-facing	Maltose
		4ZWC	2.60	Outward open	Maltose
XyIE	D-xylene-proton symporter	4GBY	2.81	Partly occluded, outward-facing	β-D-Xylose
		4GBZ	2.68	Partly occluded, outward-facing	β-D-Glucose
		4GC0	2.60	Partly occluded, outward-facing	6-bromo-6-deoxy-β-D-Glucose
		4JA3	3.80	Partly occluded, inward-facing	None
		4JA4	4.20	Inward open	None
		4QIQ	3.51	Inward open	None
GlcP	D-glucose-proton symporter	4LDS	3.20	Inward open	None

08/12/2020

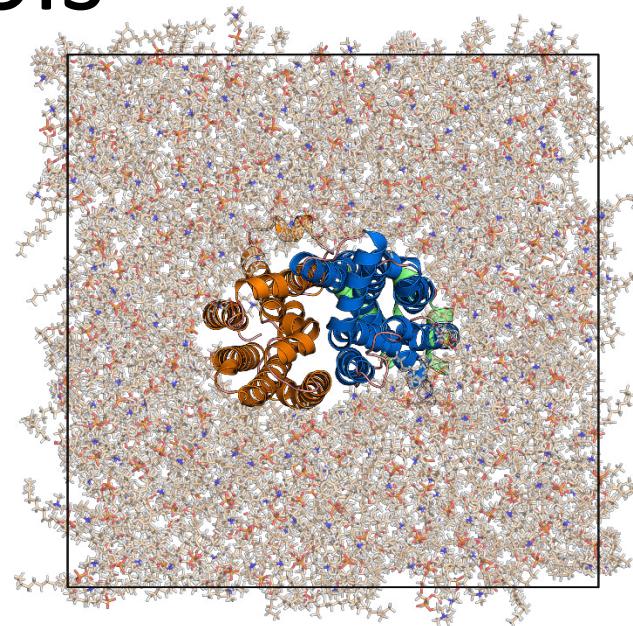
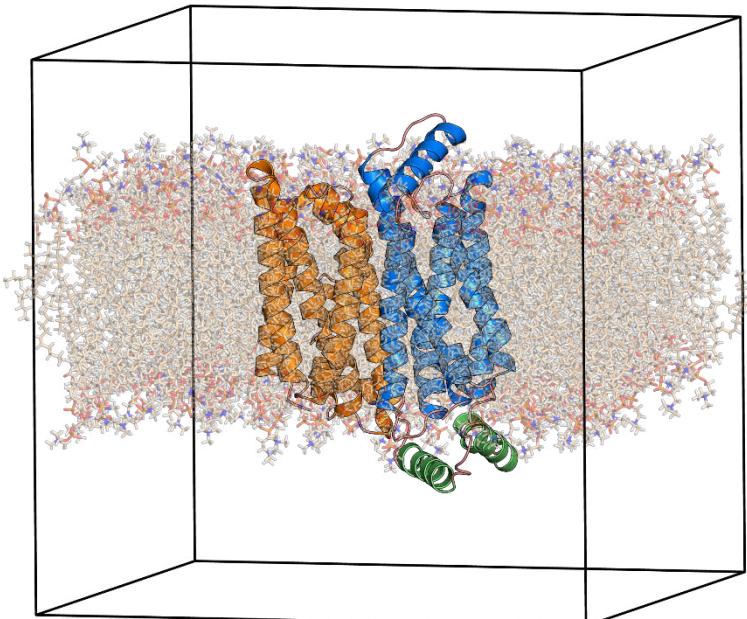




M2BI

Protocols

- 4PYP with or without ligand
- Mutant or wild type, with and without Cter
- POPC membrane
- K⁺; Cl⁻



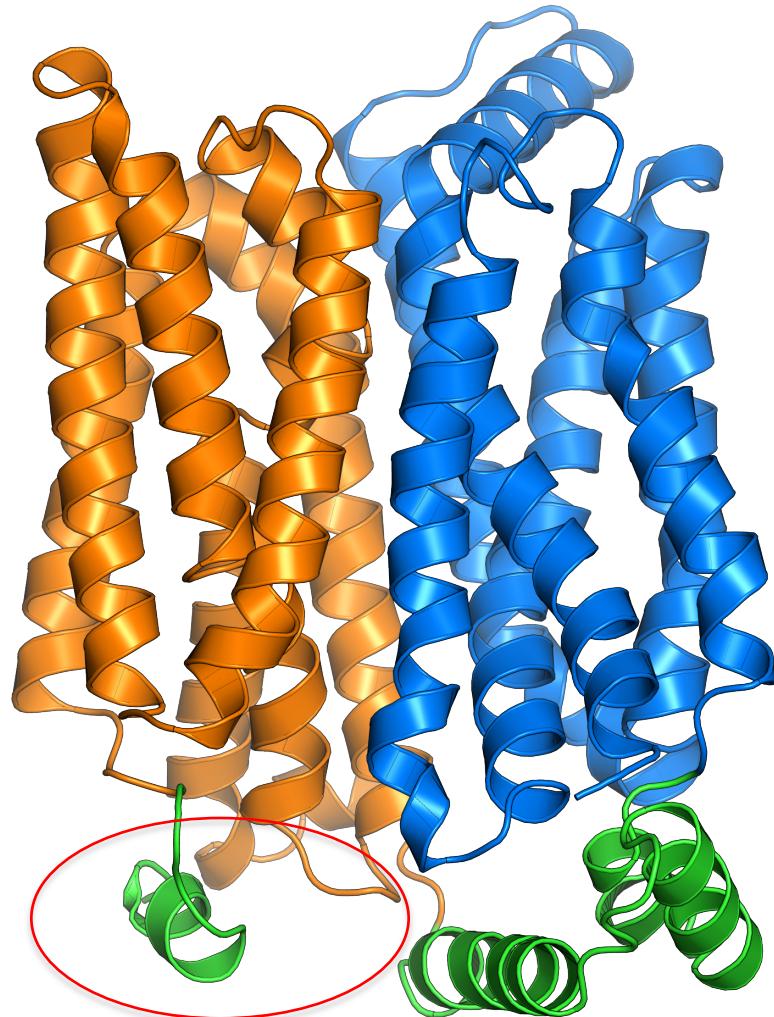
- CHARMM36 force field
- TIP3p water model
- NPT ensemble at 300 K, 1 bar
- 25 ns equilibration
- Up to 1.5 μ s production

Simulations

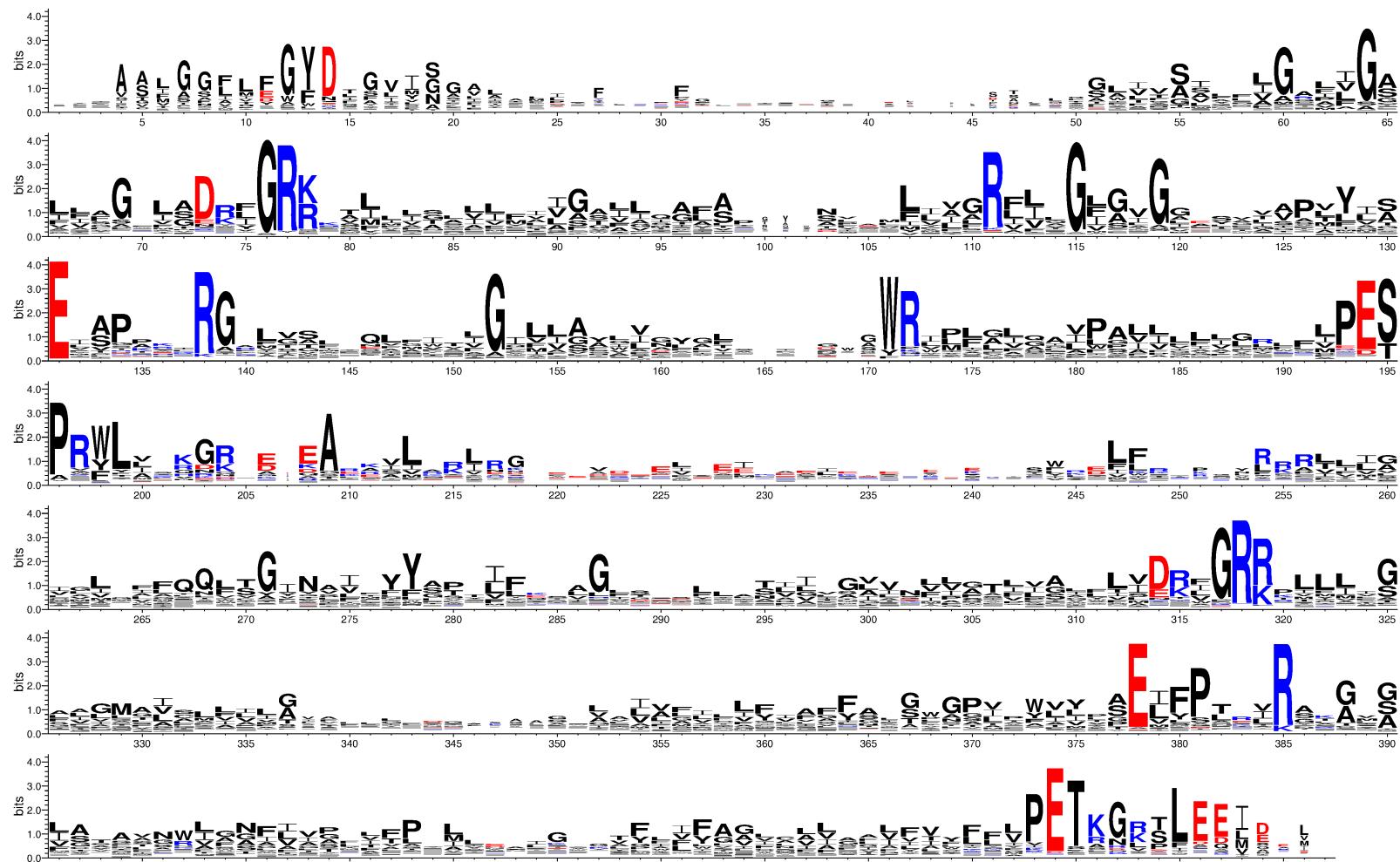
Simulation	Time (ns)	Glucose	Mutations	C-ter
glucose_wt_1	400	yes	WT	no
glucose_wt_2	500	yes	WT	no
glucose_original_1	100	yes	N45T, E329Q	no
glucose_original_2	200	yes	N45T, E329Q	no
glucose_original_3	500	yes	N45T, E329Q	no
wt_1	1500	no	WT	no
wt_2	1500	no	WT	no
original_1	1000	no	N45T, E329Q	no
original_2	1000	no	N45T, E329Q	no
wt_cter_1	1500	no	WT	yes
wt_cter_2	1000	no	WT	yes
wt_cter_3	1000	no	WT	yes
outward_occluded_1	1000	no	WT	yes

Construction of the missing region

- Model of a portion of the missing C-ter region based on bacterial Xylose proton symporter structure (4GBZ)
- Several positions were tested for the modelled C-ter
- Several protein conformations were also tested



Residue conservation



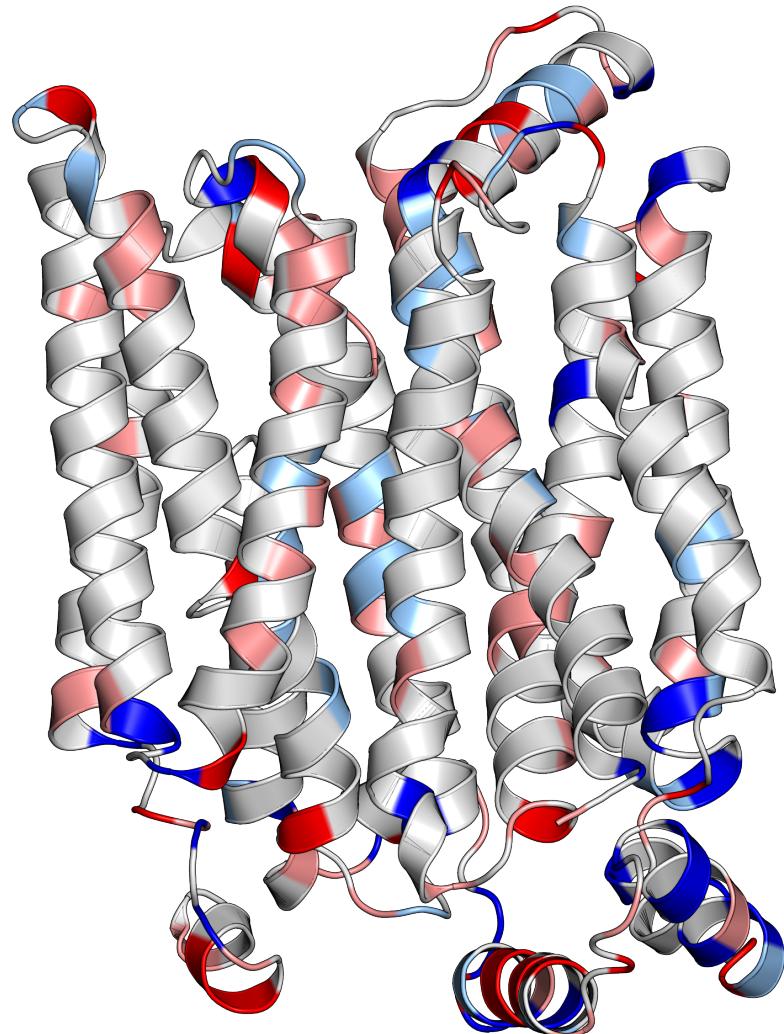
C-ter region

Positively charged residues are in blue

Negatively charged residues are in red

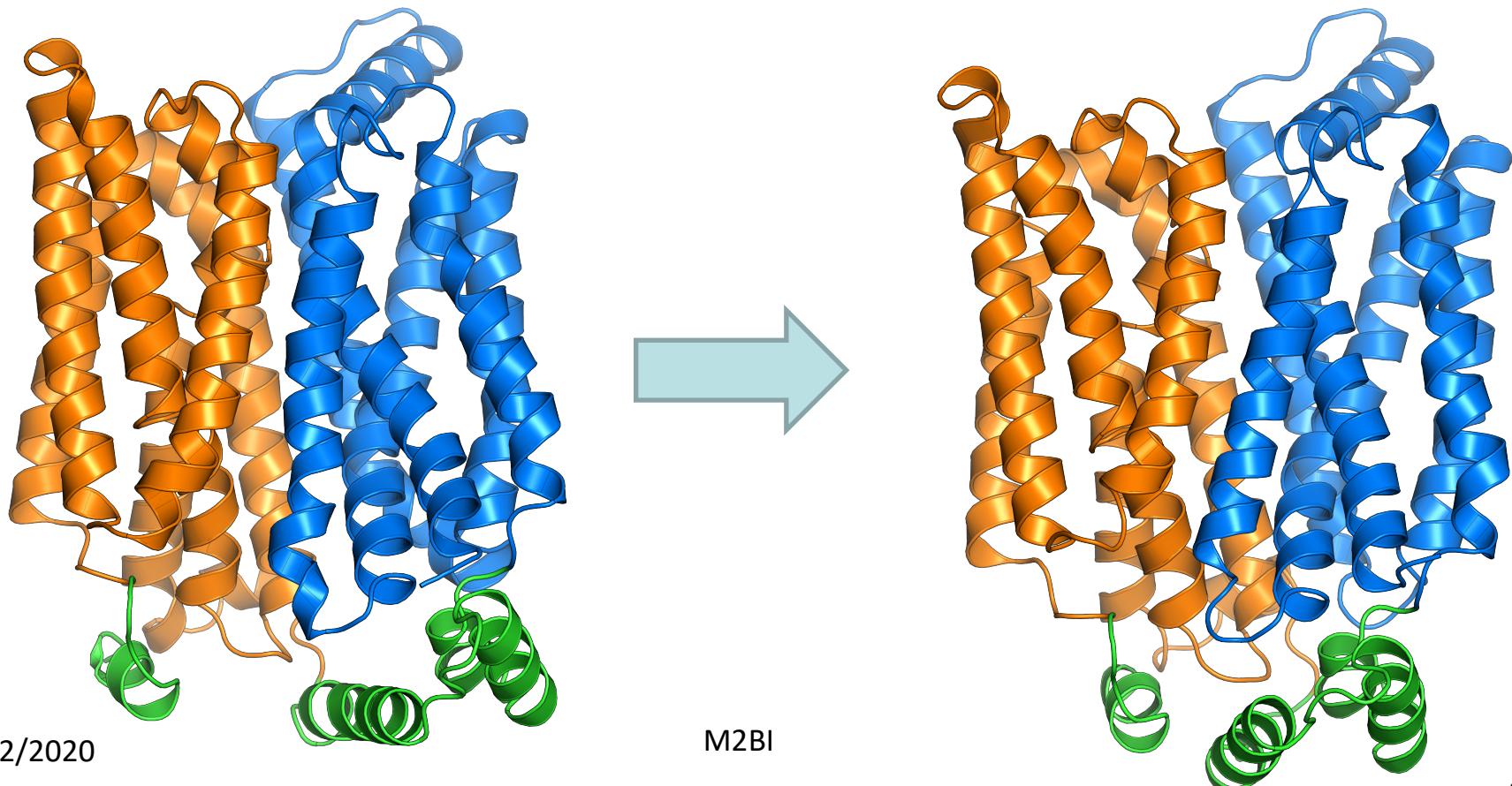
Other polar residues are in light blue or pink

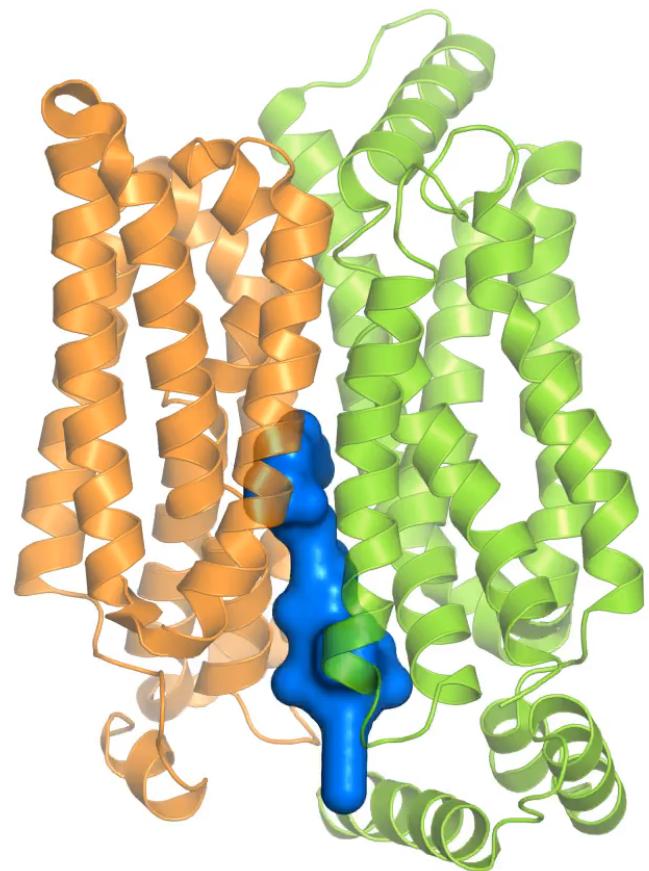
Salt-bridges between the charged residues ?



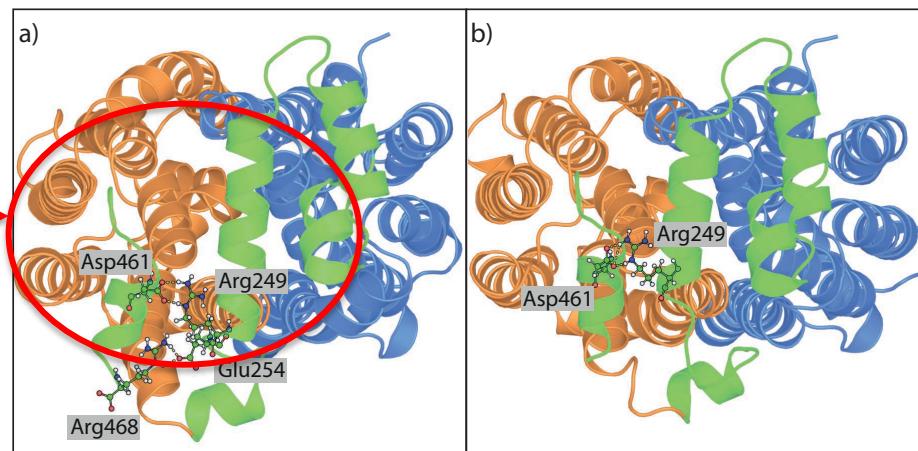
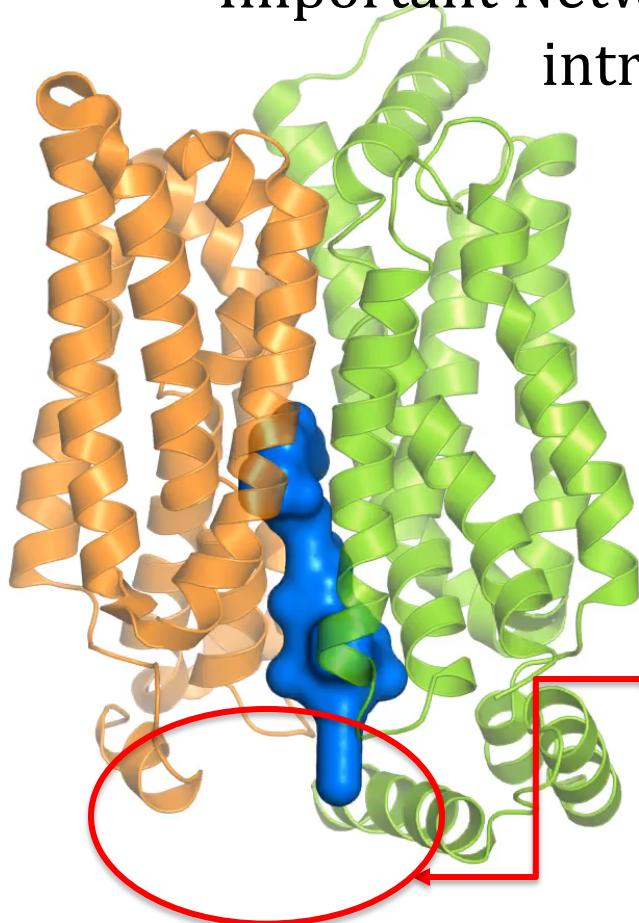
Simulations of the Apo State

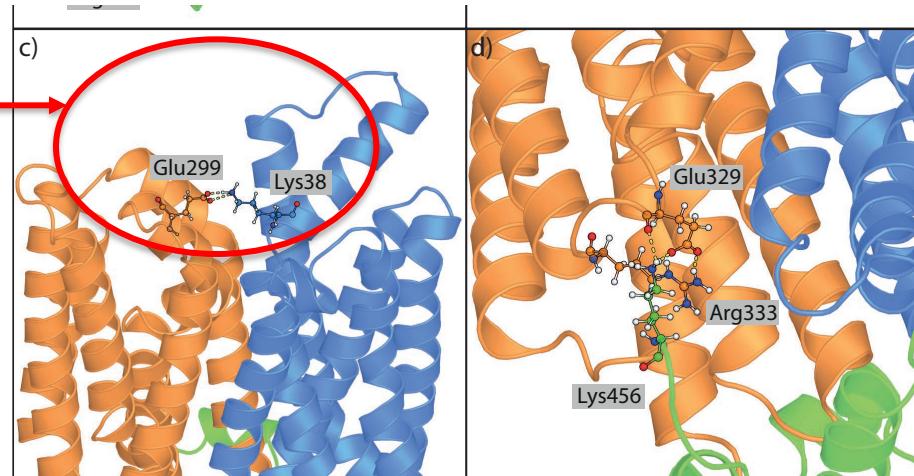
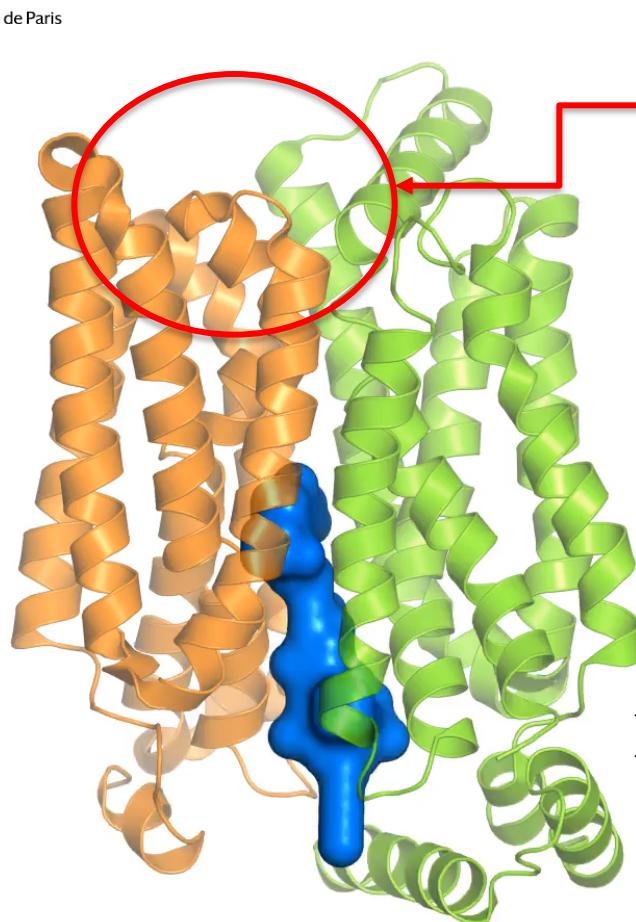
- Models with C-ter reached an occluded state during 1 μ s MD simulations





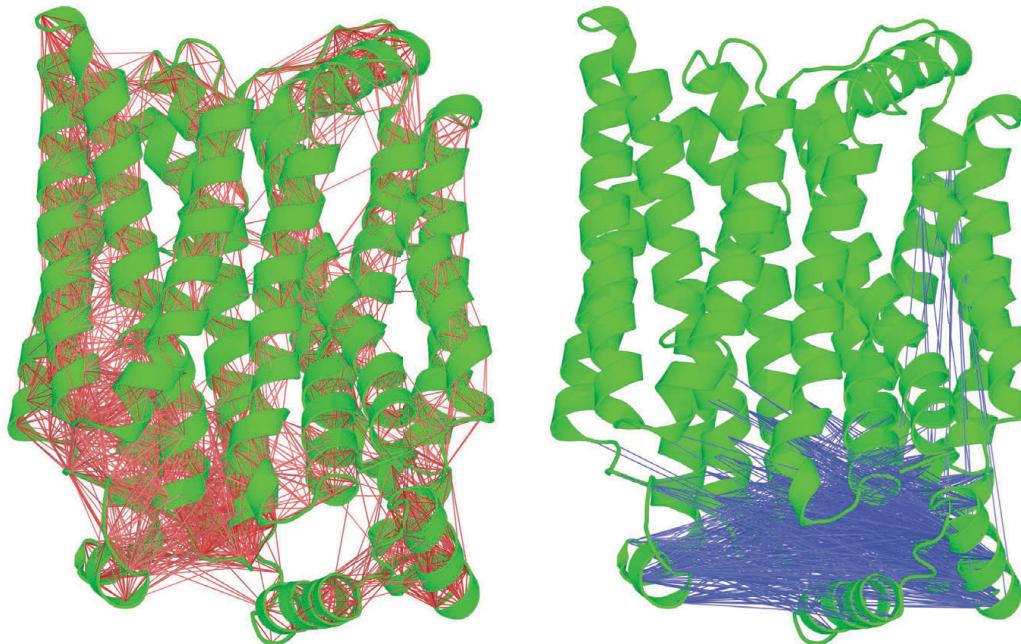
Important Network of Salt Bridges in the intracellular side



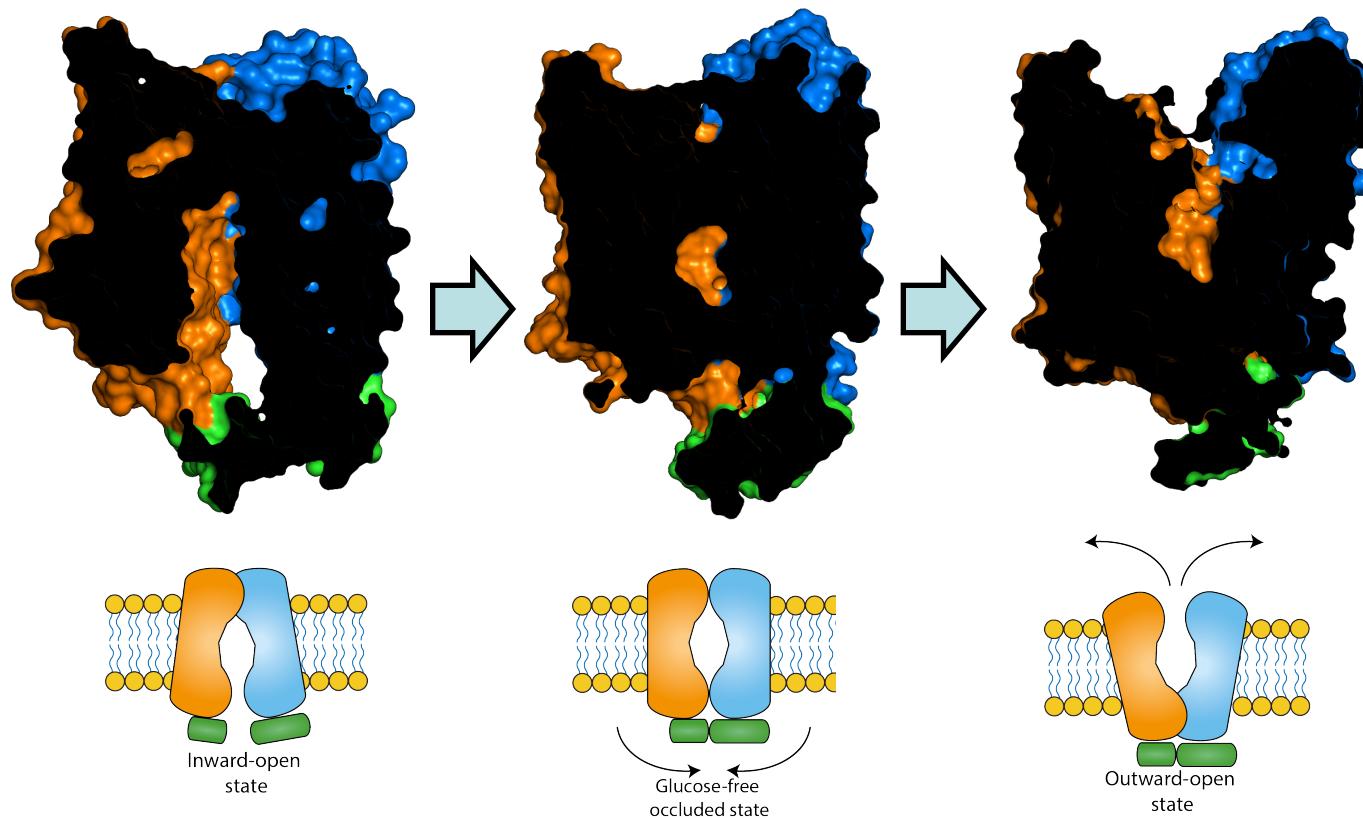


But also on the extracellular side

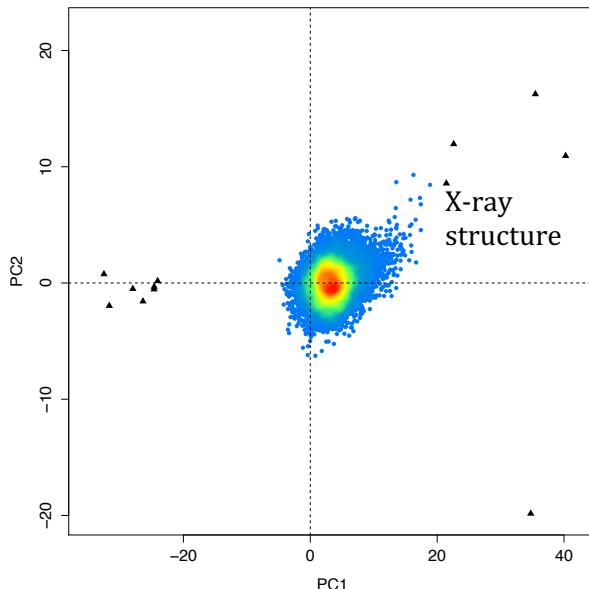
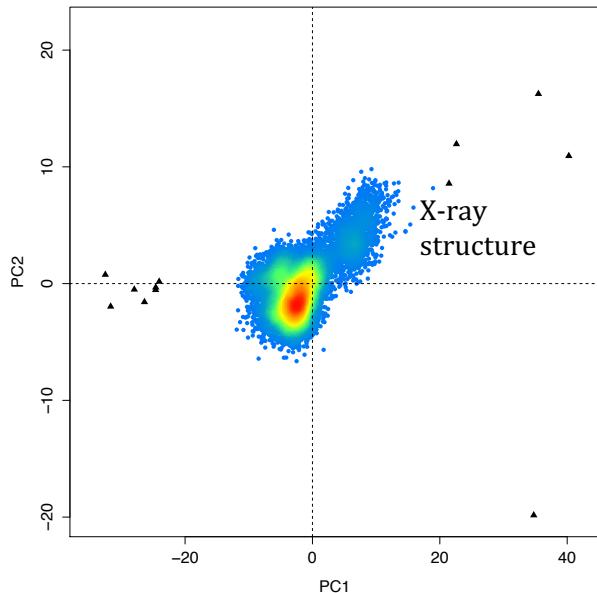
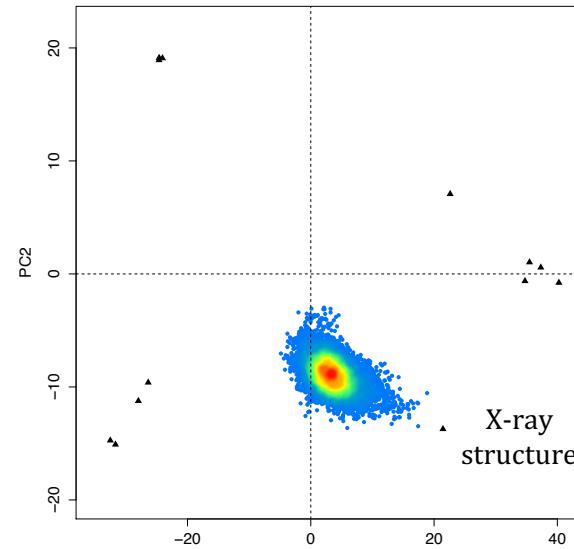
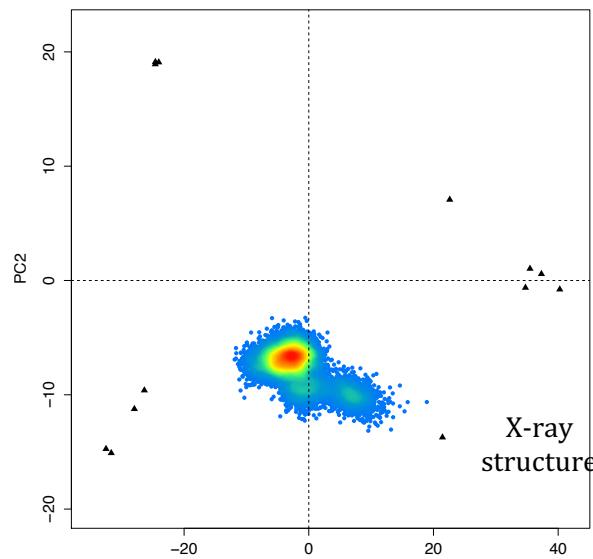
Cross correlations motion connecting intracellular and extracellular sides



Opening towards the extracellular side



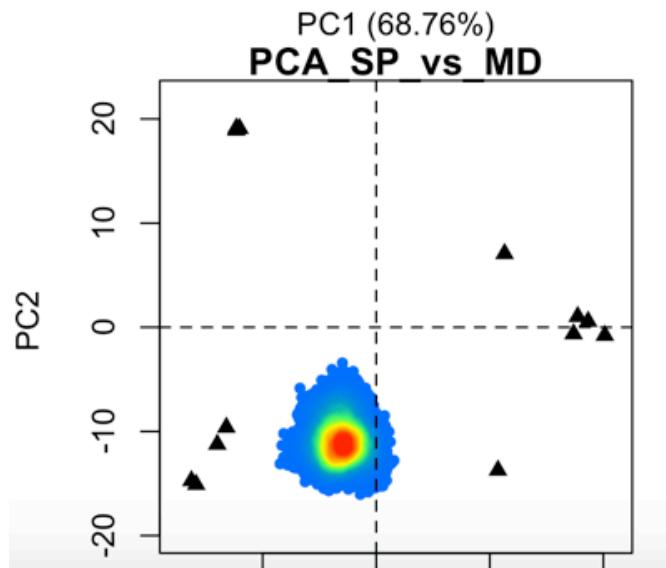
Projections of MD structures on X-ray PCA



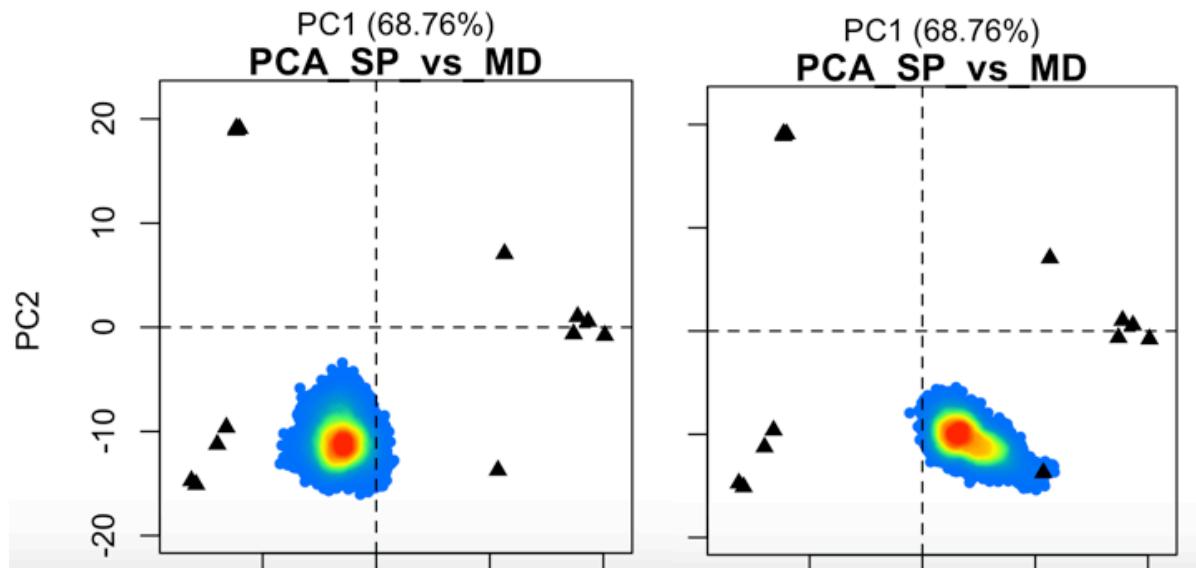
Glucose Transport

- Starting point: An outward-open conformation selected from the previous MDs.
- Docking exploration with Autodock
- Selection of different poses.
- Different sets of MD simulations ~100ns each.
 - > Entrance then exit of the glucose
 - > For some MDs, transfer of glucose in deeper position

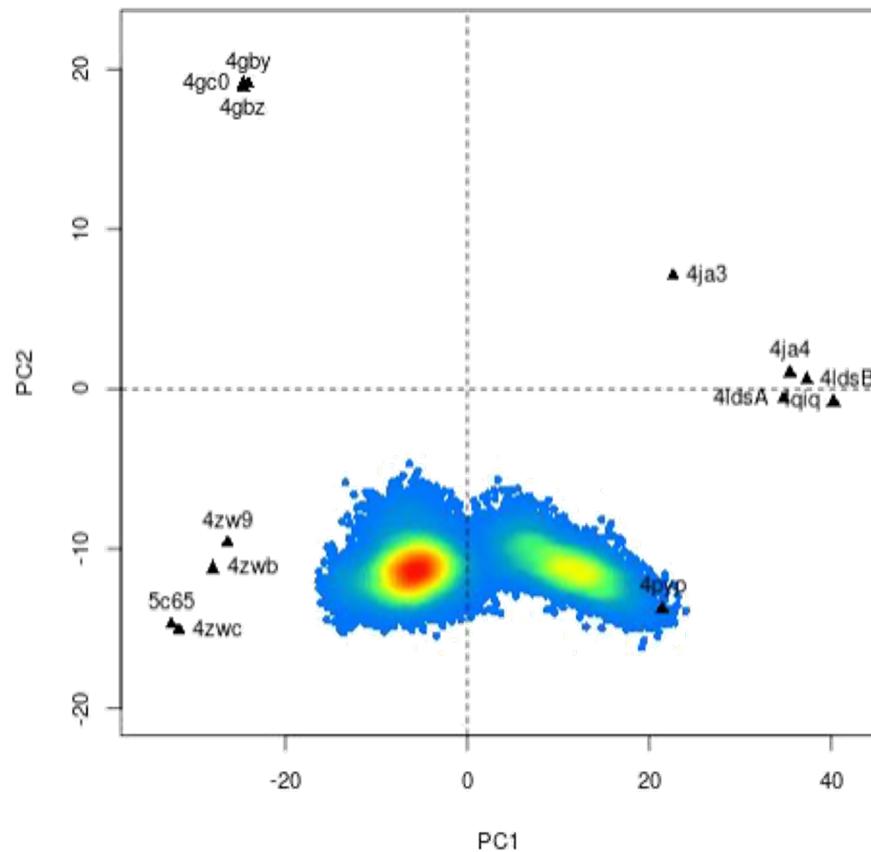
Projections of trajectories in MFS conformational space

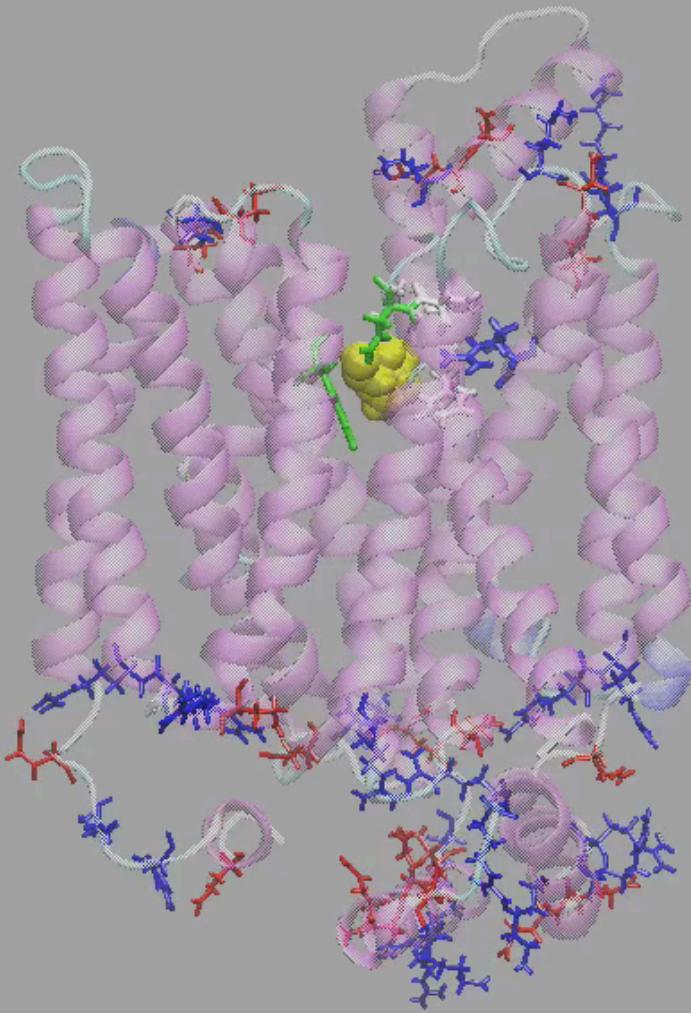


Projections of trajectories in MFS conformational space

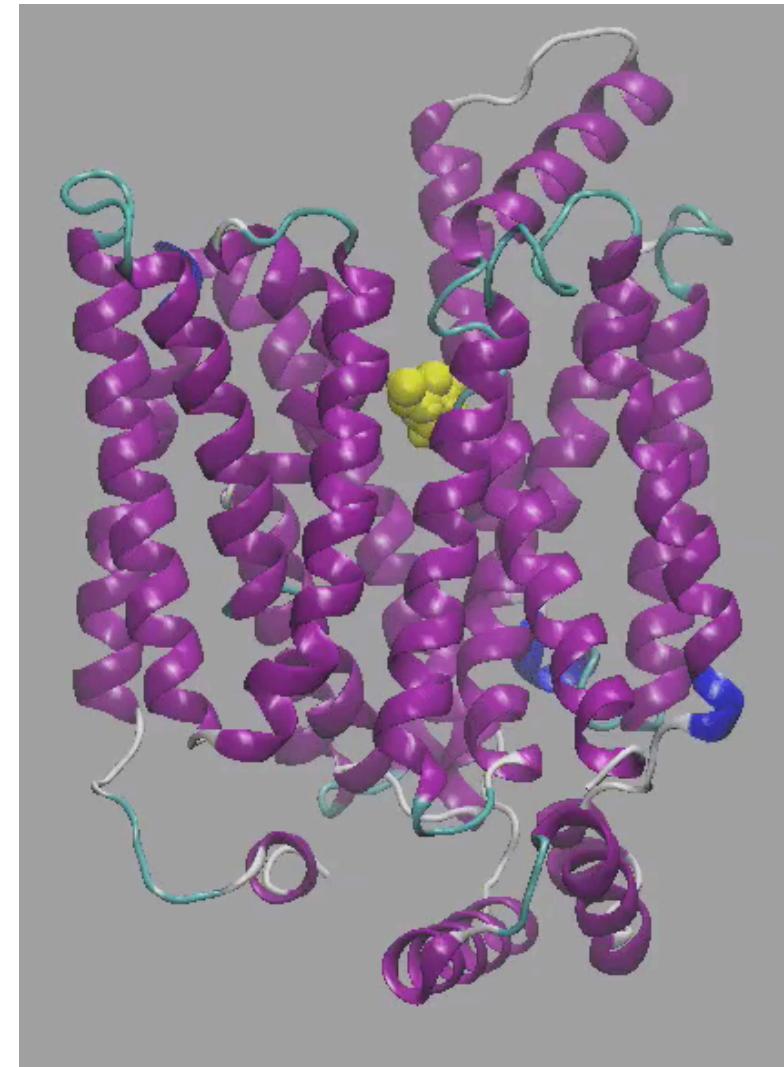


Projections on the MFS conformational space

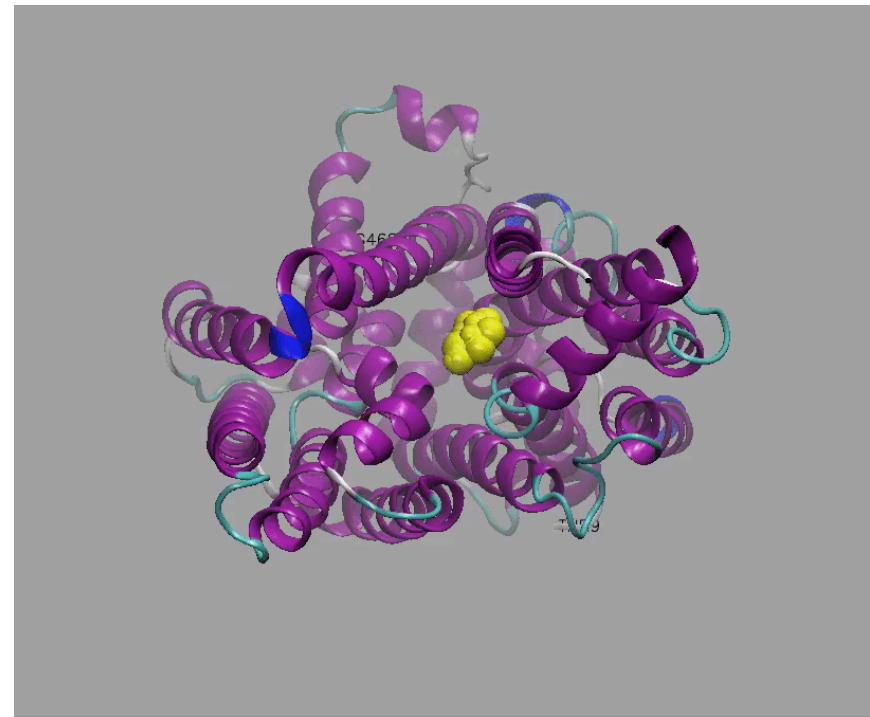




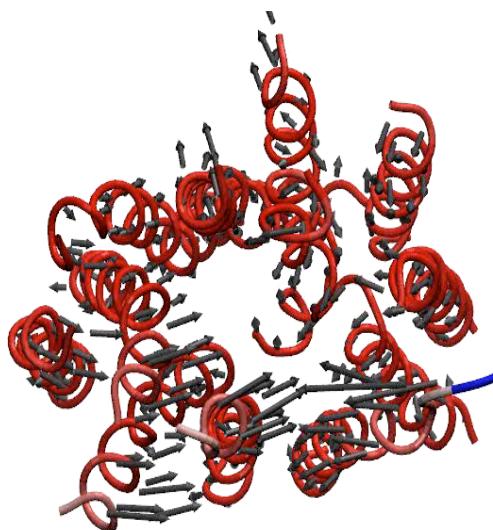
08/12/2020



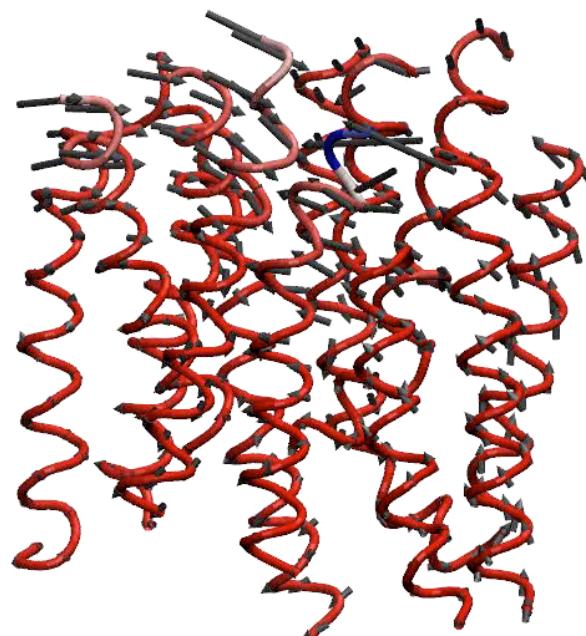
M2BI



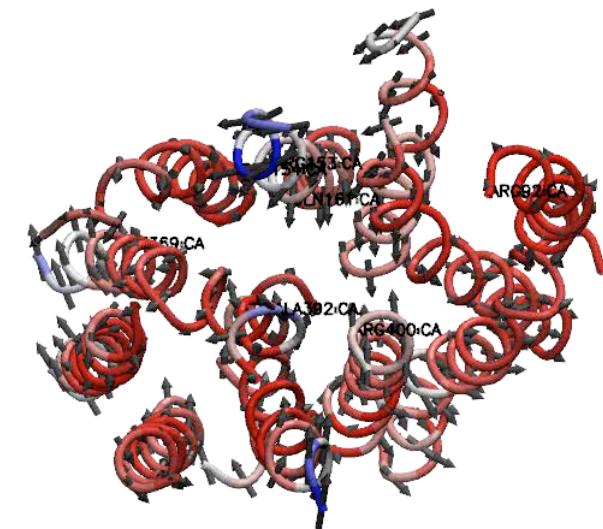
PCA direction of motions



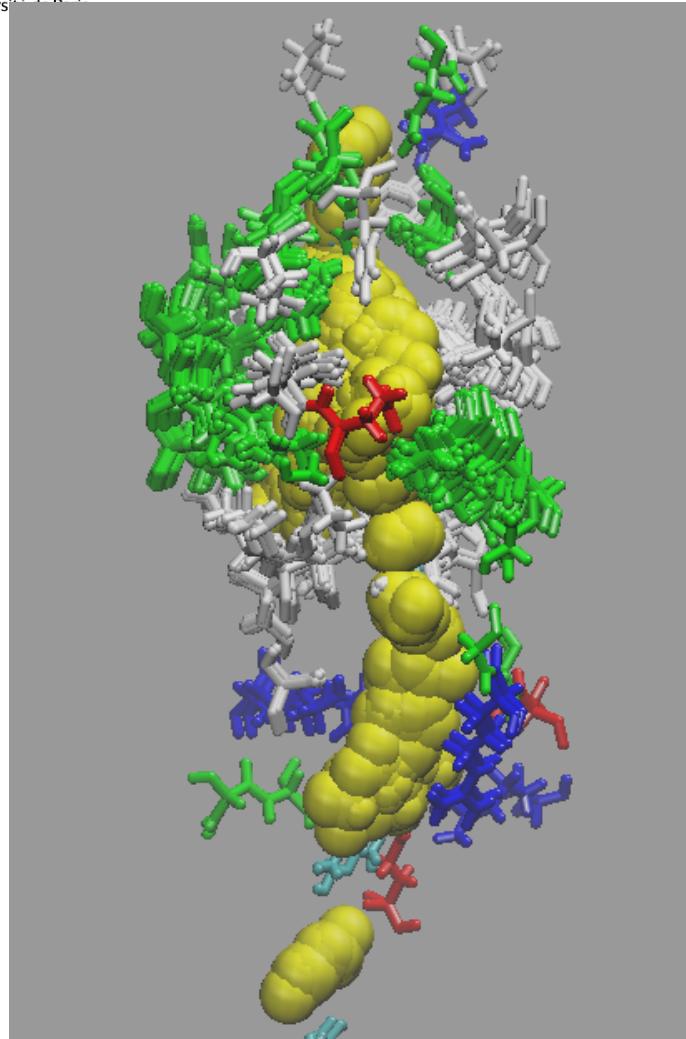
PC1



M2BI



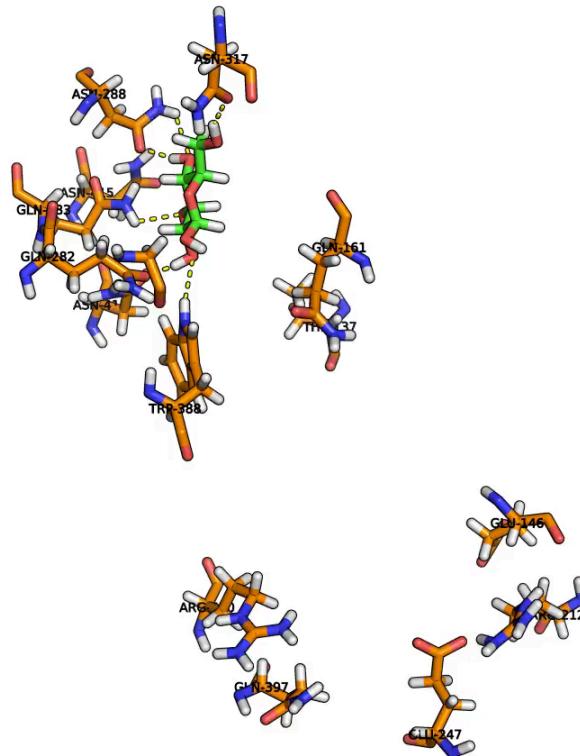
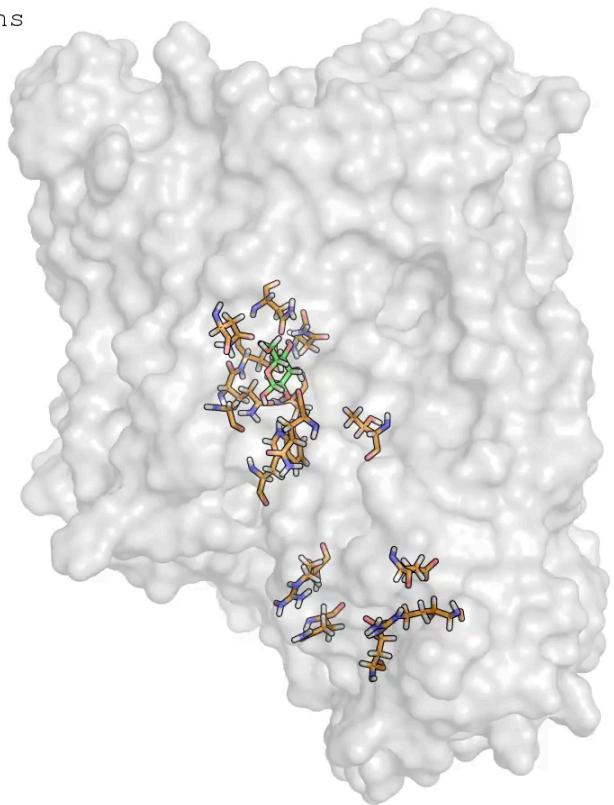
PC3

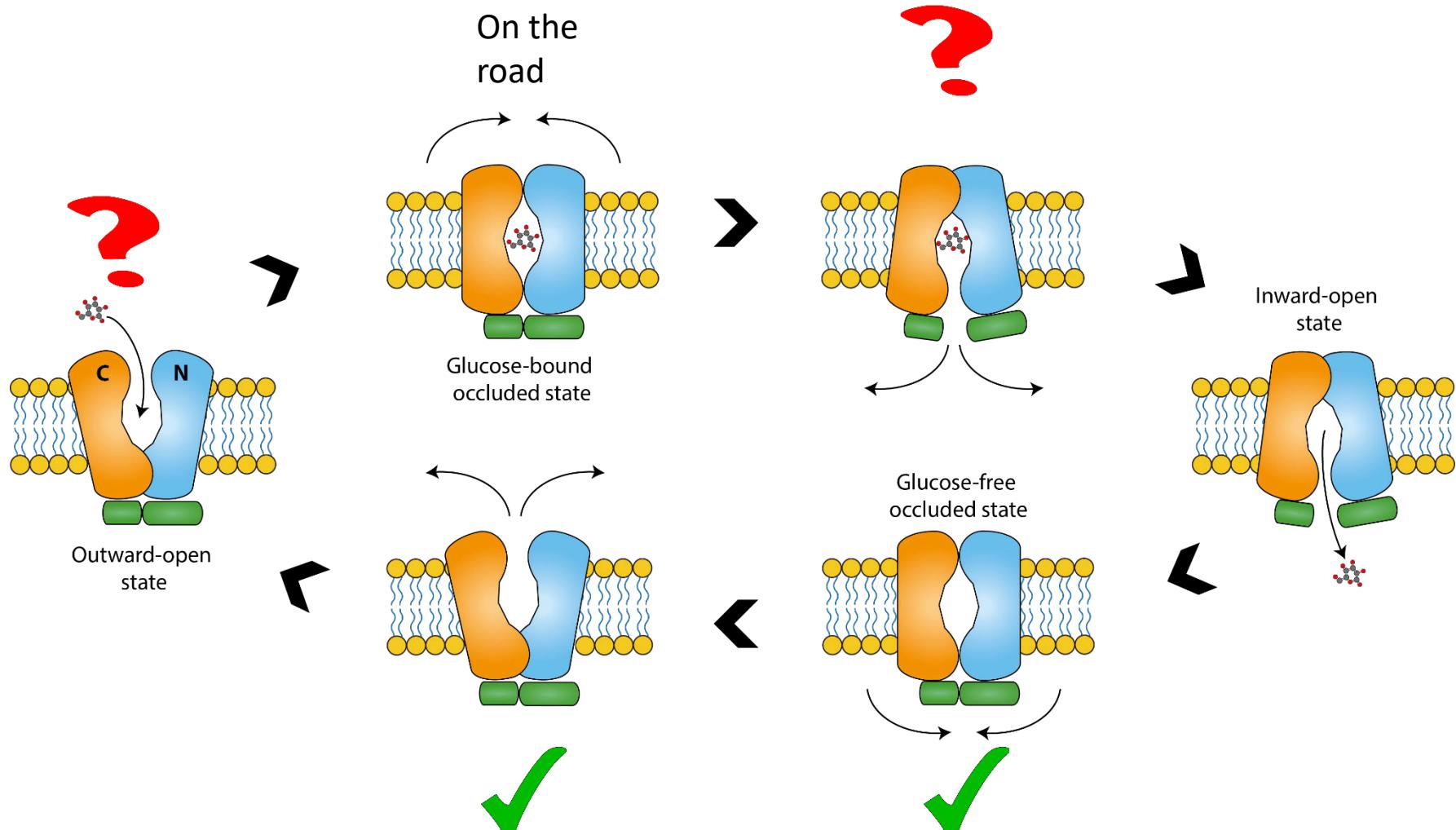


Amino acids that contact glucose along the time course of the simulation

Glucose Transport in the holo Conformation

0.00 ns





Discovery of a Novel Binding Trench in HIV Integrase

- Julie R. Schames,* Richard H. Henchman, Jay S. Siegel, Christoph A. Sotriffer, Haihong Ni, and J. Andrew McCammon, J. Med Chem 2004

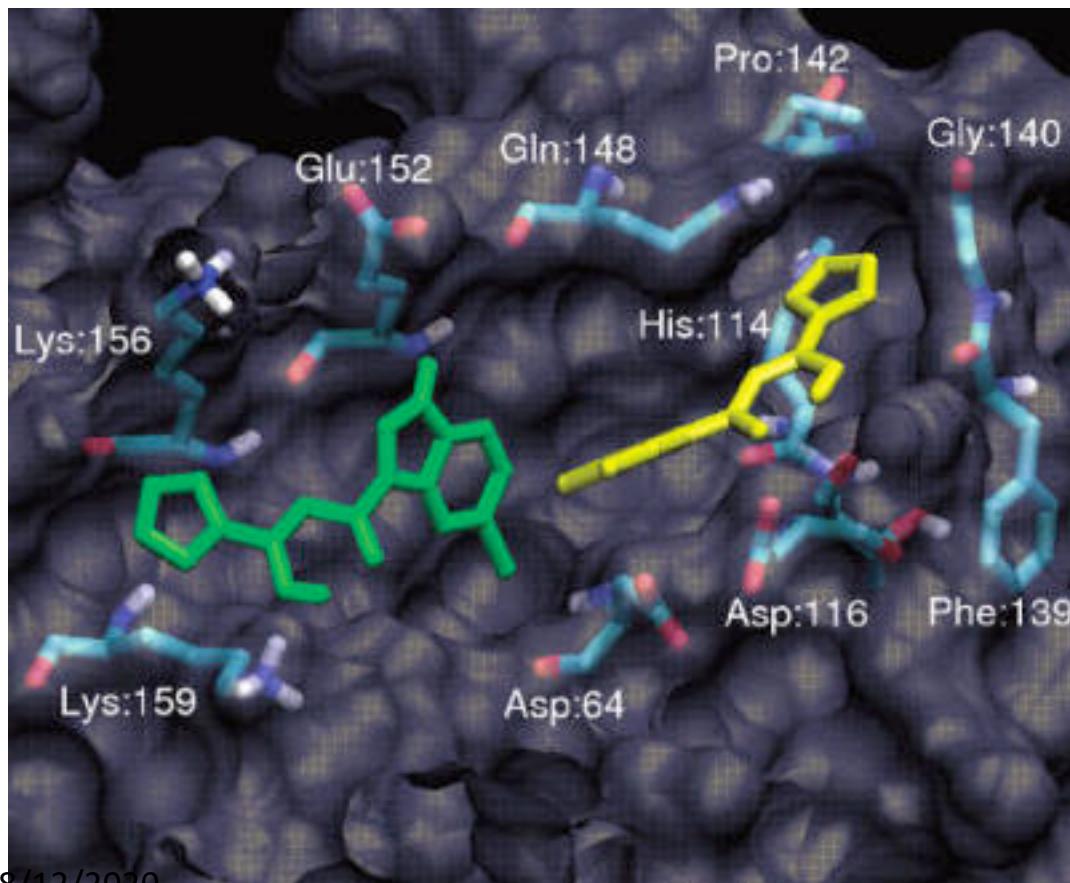
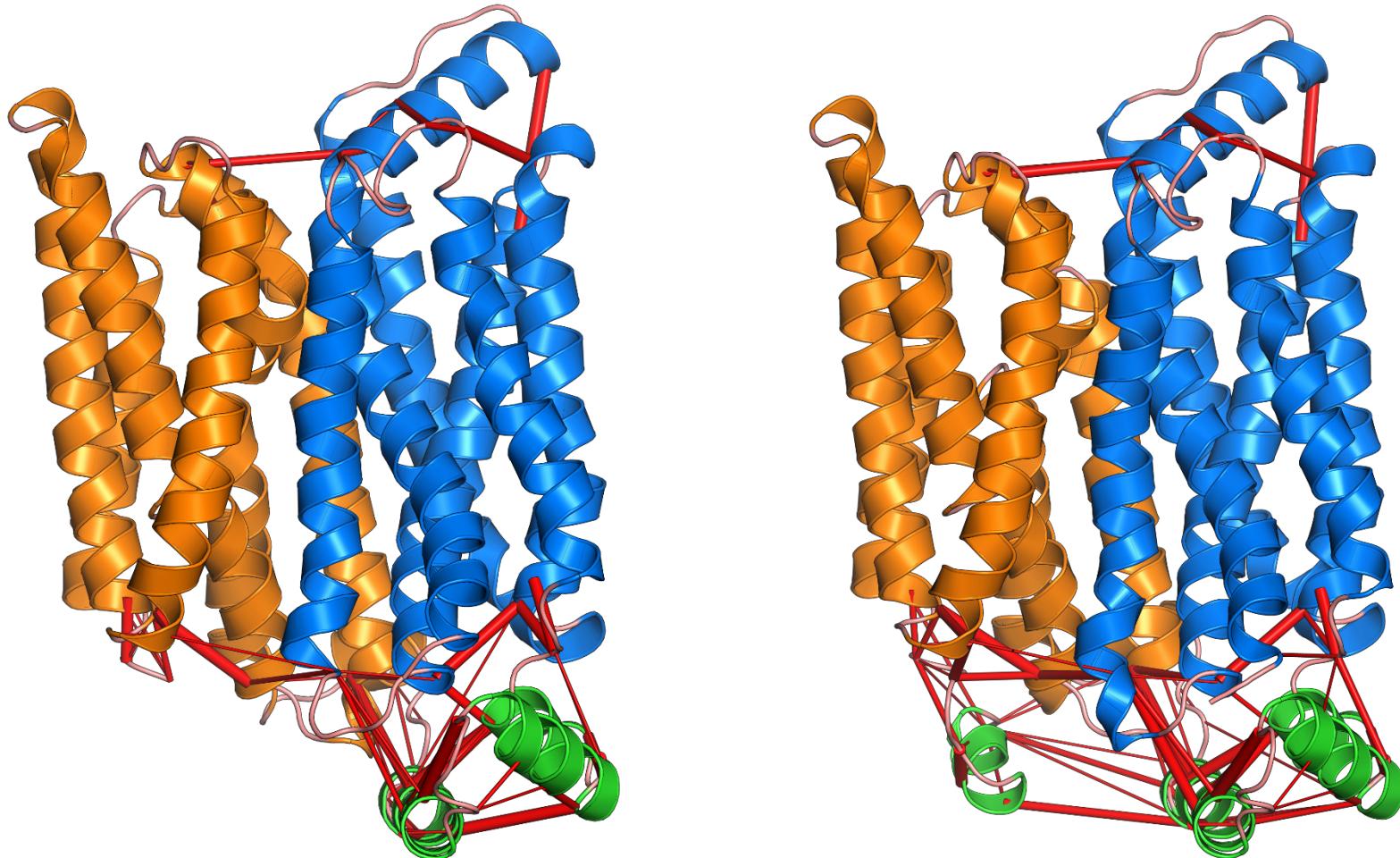


Figure 1. The two predominant docking conformations of 5CITEP to an open MD snapshot of integrase. The ligand in green shows 5CITEP in the orientation similar to the crystal structure of the complex. The ligand in yellow shows 5CITEP in its “flipped” orientation. Residues lining both ligand positions are highlighted.

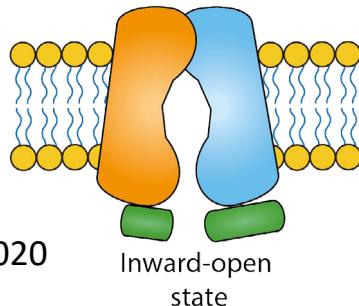
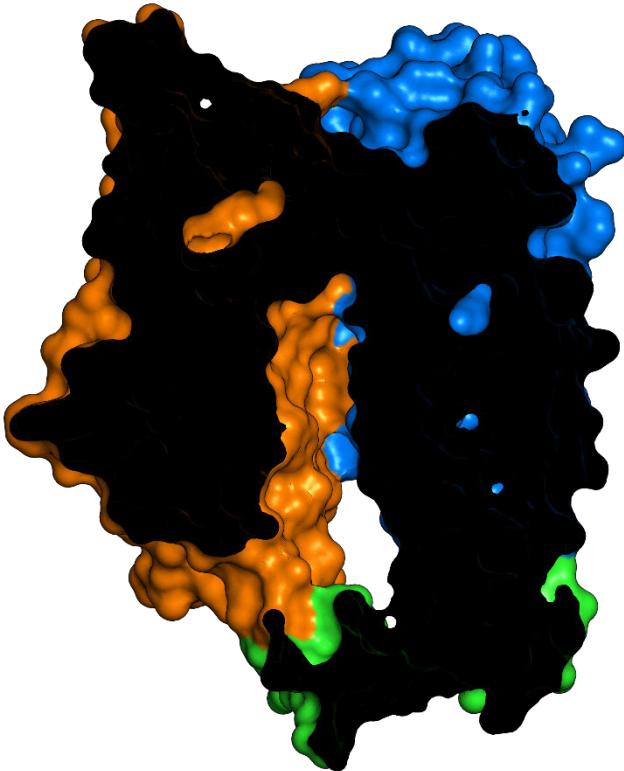
FIN

Role of C-ter region

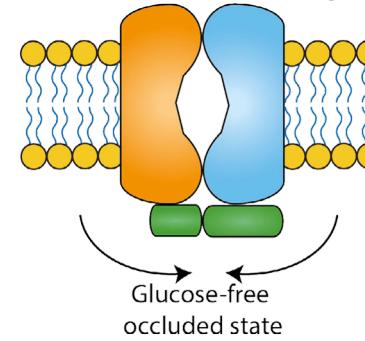
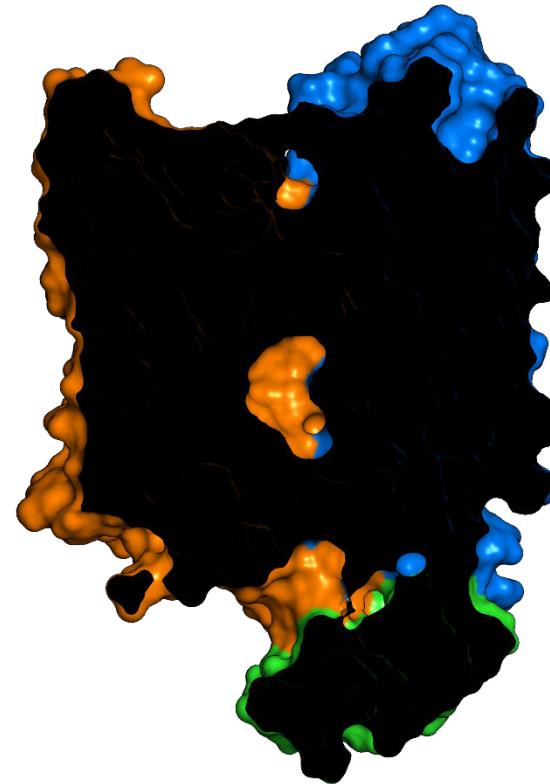
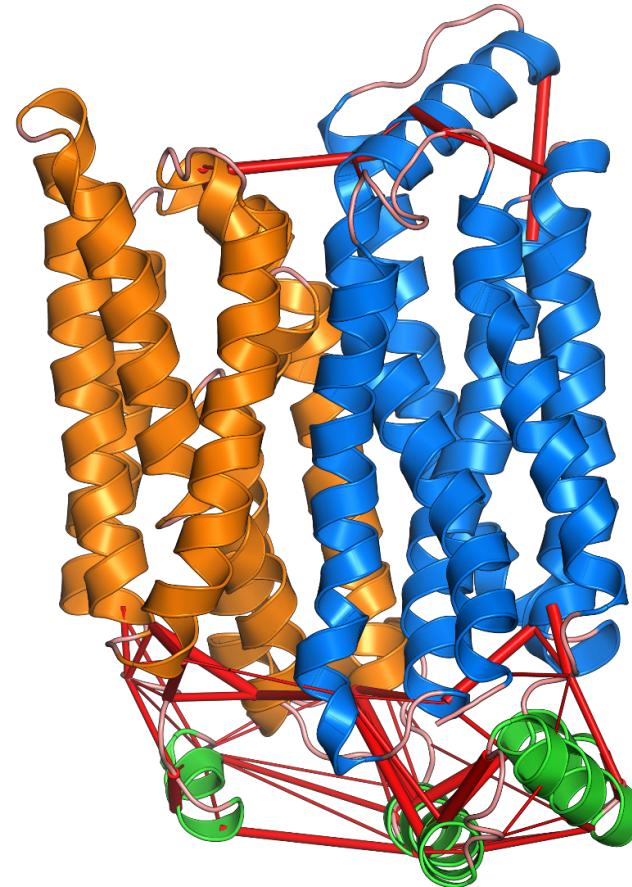


Salt bridges are represented by **red cylinders** with thickness proportional to the **persistence** of the interaction.
08/12/2020

Apo State : Conformational change



08/12/2020



85

In some cases: Outward-open state

