The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

SMR/1845-14

Conference on Structure and Dynamics in Soft Matter and Biomolecules: From Single Molecules to Ensembles

4 - 8 June 2007

Computational studies of the early steps of protein aggregation using the OPEP coarse-grained energy model

Philippe DERREUMAUX

Institut de Biologie Physico Chimique CNRS
13 rue Pierre et Marie Curie
75005 Paris
FRANCE

Computational studies of the early steps of protein aggregation using the OPEP coarse-grained energy model.

Philippe Derreumaux
Laboratoire de Biochimie Théorique
UPR9080 CNRS, IBPC, Université Paris VII

Amyloid-fibril formation is often described by a polymerization-nucleation process. Once a nucleus is formed (highest free energy), maturation into fibrils is rapid. Often we see amorphous aggregates and annular species prior to the formation of protofibrils.

Structural characterization of the early oligomers (e.g tetramers and dodecamers of Aβ with the highest toxicity) is a challenging task from both the experimental and numerical fronts.

They are transient and in dynamic equilibrium between various oligomeric species . The lag phase is salt, concentration and sequence-dependent, and spans a time scale of hours unreachable by all-atom MD simulations (1-2fs timestep).

In the case of Aβ, monomer is described by random coils and the atomic fibril model is still under refinement and depends on agitation (strains for PrP), so standard pathway techniques cannot be used.

To better understand assembly through computer simulations, we must resort to coarse-grained models.

Four issues using coarse-grained simulations on amyloid-forming peptides: KFFE, Aβ(16-22), β2m(83-89), Aβ(11-25), NFGAIL (IAPP fragment), Aβ(1-40) and Aβ(1-42)

1. What is the dynamics of self-assembly and the free energy surfaces for small oligomers (2- to 4-mers: building blocks for larger oligomers) ?

2. What are the mechanisms for oligomer growth and the paths from random states to fibrils? (4- to 12-mers)?

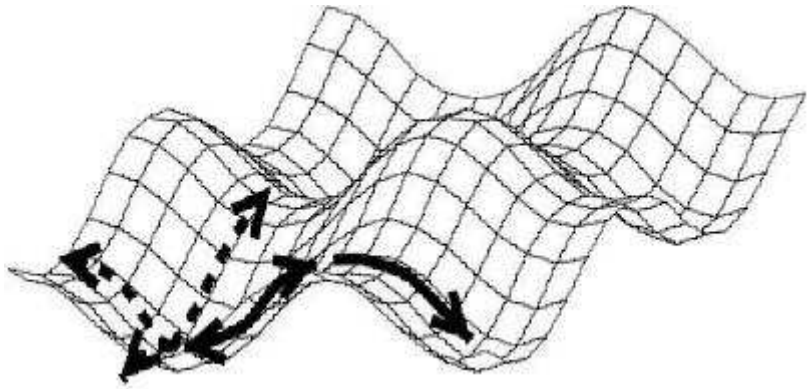
3. What are the equilibrium structures of full-length Aβ dimers?

4. How do N-methylated peptides block fibrillogenesis?

Exploring conformations, phase space:

ART-Nouveau OPEP simulations

- **Activation :**
  - Bring a conformation outside its minimum;
  - Direction is chosen randomly (in a 3N dimensional space);
  - Follow this direction until eigenvalue becomes negative;
  - Push along the corresponding eigenvector until total force is approaching to zero (Saddle point);



Malek, Mousseau, PRE, 2000;
Wei, Mousseau, Derreumaux *JCP, 2002*

- **Relaxation** :
  - Minimize with *Steepest Descent*

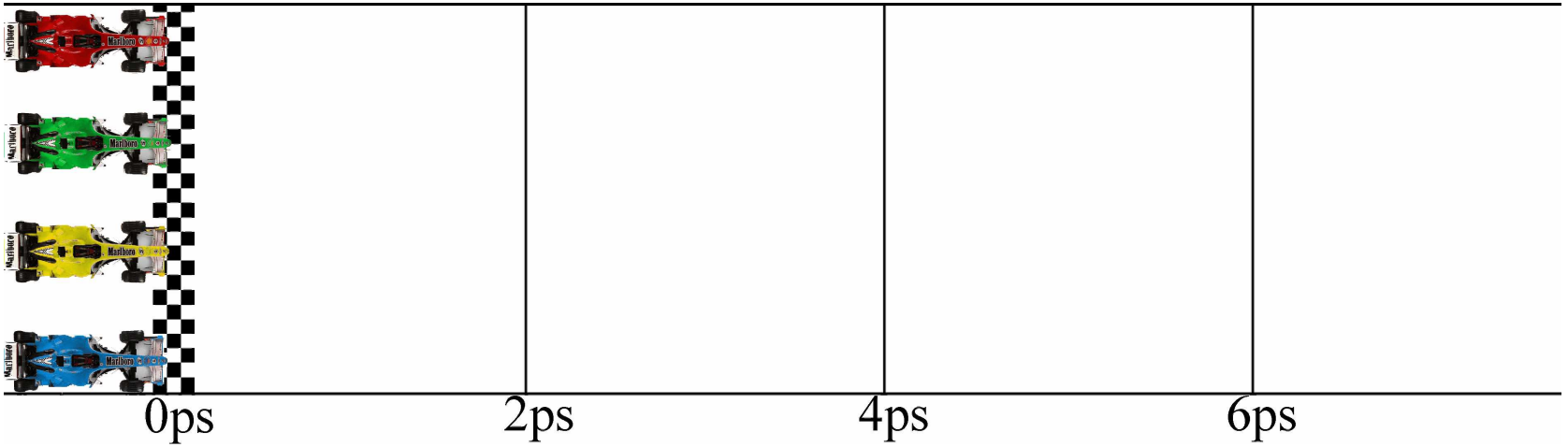
- Conformation is accepted/rejected by **Metropolis criterion.** *not real T, no time scale, but moves of any complexity, physical paths*

MD-OPEP simulations

- **Molecular Dynamics at T constant:**
  - Berendsen bath (0.1 ps);
  - timestep (1 – 2 fs); RATTLE

Derreumaux, Mousseau, J. Chem. Phys 126, 065101-7 (2007).

REMD-OPEP simulations



X replicas between T1 and Tx, exchange swap: 20 ps

$$P(1 \leftrightarrow 2) = \min\left(1, \exp\left[\left(\frac{1}{K_B T_1} - \frac{1}{K_B T_2}\right)(U_1 - U_2)\right]\right)$$

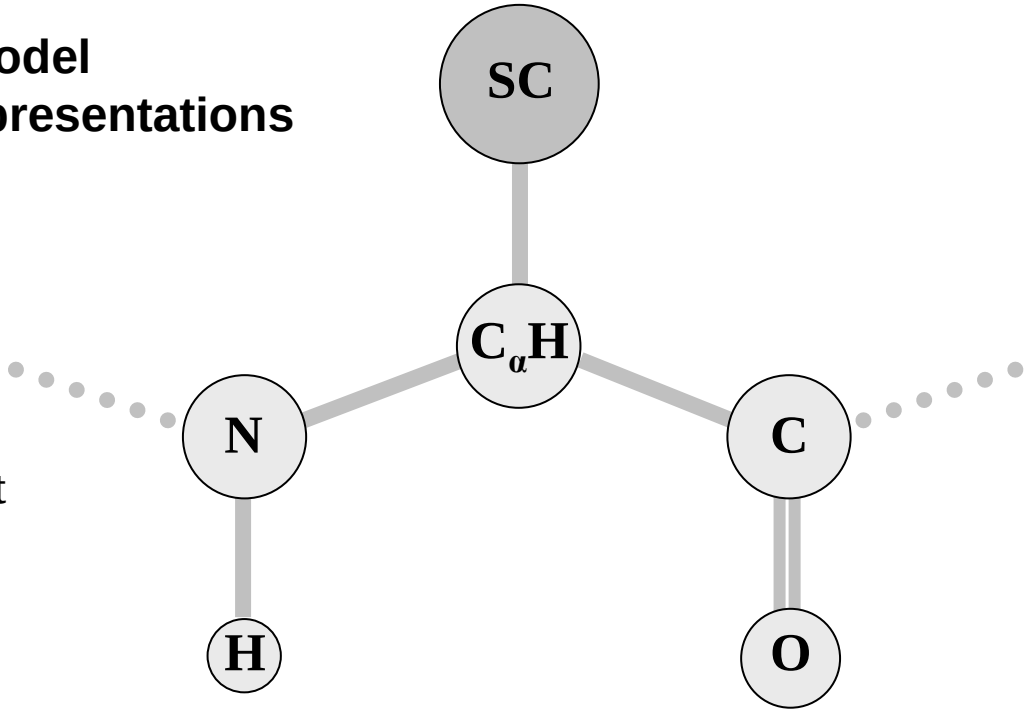
Dong, Derreumaux, Mousseau, in preparation

OPEP force field

(**O**ptimized **P**otential for **E**fficient structure **P**rediction)

Coarse-grained off-lattice model between Cα and all-atom representations

Generic model to be used for structure prediction, protein folding and protein aggregation of any sequence: H, O and Sc included, parameters are not tuned according to the problem, rugged energy surface.



Derreumaux: J. Chem. Phys: 1997,1998, 1999; Phys. Rev. Lett. 2000 ; J. Chem. Phys. 2003, Structure 2004, JACS 2004, Proteins 2007

OPEP force field*: implicit solvent, pH cannot be varied, no electrostatic charges*

Potential for stereochemistry

- Harmonic terms for bond lengths, bond angles, and improper torsions near their equilibrium values: $E_L = \sum K_q (q - q_0)^2$,
- Excluded volume-potential : $w_{M,M} E_{M,M}, w_{M,SC} E_{M,SC}$

Pairwise Potential between side chains **20AA, 210 terms**

A 12-6 potential if $\epsilon_{ij}>0$ and a 6-potential if $\epsilon_{ij}<0$:

$$E_{SC,SC} = \epsilon_{ij} \left(\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{6} \right) H(\epsilon_{ij}) - \epsilon_{ij} \left(\frac{r_{ij}^0}{r_{ij}} \right)^{6} H(-\epsilon_{ij})$$

Two-body Potential for one H-bond

$$E_{hb} = \varepsilon_{hb} [5(\frac{\sigma}{r_{ij}})^{12} - 6(\frac{\sigma}{r_{ij}})^{10}] \cos^2 \alpha_{ij}$$

σ the O..H distance and α_{ij} the NHO angle.

Cooperative energy between two H-bonds ij and kl

$$E_{2hb} = \varepsilon_{2hb} \exp(-(r_{ij} - \sigma)^2/2) \exp(-(r_{kl} - \sigma)^2/2)$$

ij and kl must satisfy α-helix and β-sheet patterns

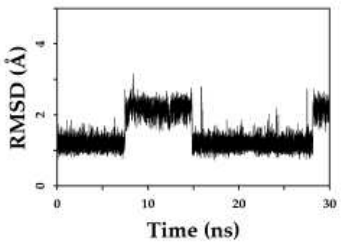
How do we optimize the parameters: r^0 and q_0 (from structure analysis), kq (from vibrational modes), the 210 Sc-Sc ε terms and the H-bond ε terms (from *decoys of 30 proteins with α , β and α/β structures)*

OPEP discriminates native from decoys

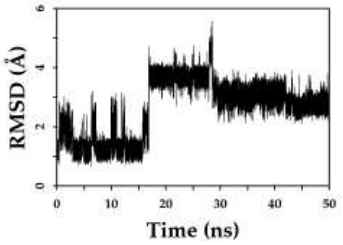
set	M	L	*Number of Decoys*				cRMSd (Å)		TM-score		α-helix	β-strand
			MD	TH	GR	TOT	*min*	*max*	*min*	*max*	%	%
Full Training Set												
1ABZ	1	38	49	281	278	608	2.6	11.9	0.14	0.58	46.9	5.82
Betanova	-	20	347	299	283	930	0.7	12.2	0.01	0.59	10.36	23.87
1DV0	1	45	118	209	275	602	1.9	13.6	0.16	0.76	35.81	4.91
1E0M	1	37	59	213	183	451	1.5	12.8	0.14	0.82	17.14	17.33
1ORC	1	64	79	89	262	430	0.9	13.6	0.22	0.92	26.69	20.00
1PGB	1	56	67	179	289	535	0.9	36.9	0.12	0.92	26.57	20.41
1PGBF	1	16	28	276	300	604	0.2	11.1	0.01	0.99	10.31	15.05
1QHK	1	47	59	189	277	525	3.3	31.1	0.13	0.78	24.22	10.03
1SHG	1	57	41	601	286	928	1.3	41.7	0.12	0.85	16.99	21.98
1SS1	1	60	90	128	212	430	2.2	13.8	0.19	0.85	39.77	4.49
1VII	1	36	45	250	270	565	1.6	10.9	0.13	0.68	37.41	4.33
2CI2	6	65	30	166	298	494	2.2	34.9	0.11	0.87	21.52	16.33
2CRO-fisa†	3	65	25	-	-	525	2.2	12.3	0.23	0.99	61.75	0.02
Full Validation Set												
1BBA-lmds†	1	36	-	-	-	500	0.9	9.2	0.41	0.82	47.33	0.00
1CTF-4state†	2	68	-	-	-	616	0.5	12.5	0.24	0.88	33.56	4.09
1CTF-lattice†	2	68	-	-	-	977	5.1	10.5	0.25	0.51	34.65	0.26
1CTF-lmds†	2	68	-	-	-	489	3.3	15.1	0.28	0.68	51.28	21.16
1CTF-semfold†	2	68	-	-	-	996	4.4	9.2	0.28	0.59	50.70	0.63
1F4I	1	45	11519•	-	-	11519	0.1	36.9	0.12	0.99	4.74	7.37
1FSD	1	28	84	319	290	693	1.4	10.7	0.13	0.80	34.49	4.68
1KHM-semfold†	1	73	-	-	-	998	3.9	7.9	0.32	0.58	44.31	2.22
1R69	1	63	72	99	281	452	0.1	27.7	0.23	0.99	46.81	3.51
1R69-4state†	1	63	-	-	-	670	0.8	11.4	0.25	0.94	36.63	0.36
1R69-rosetta†	1	61	-	-	-	998	0.2	14.9	0.27	0.99	63.13	0.00
1S04-CASP6†	1	110	27	-	-	213	0.6	56.5	0.15	0.98	28.19	15.66
1TE7-CASP6†	1	103	18	-	-	222	1.8	38.3	0.18	0.88	22.69	21.79
1UBI-lmdsv2†	1	76	61	-	-	361	2.0	15.4	0.22	0.81	15.54	30.44
2CRO-4state†	3	65	25	-	-	697	0.1	9.3	0.25	0.99	36.33	0.19
2CRO-lmds†	3	65	25	-	-	525	0.1	12.9	0.25	0.99	63.85	0.06

OPEP and DOPE (Sali et al., JMB 2006) have similar performances and outperform Rosetta force field (Maupetit, Tuffery, Derreumaux, Proteins, 2007 in press).

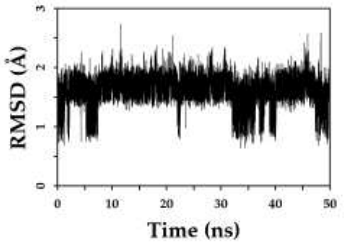
MD-OPEP provides good dynamical properties



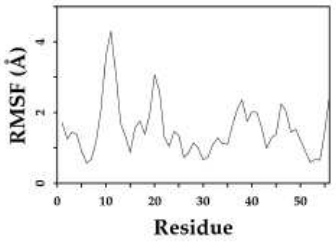
(a) protein G



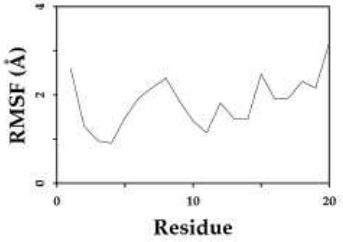
(b) betanova



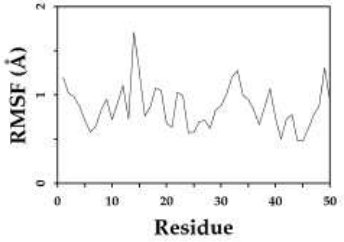
(c) protein A



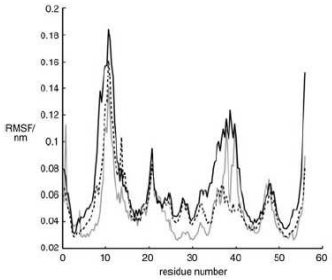
(d) protein G

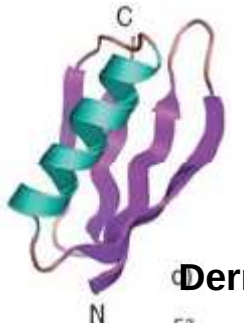


(e) betanova



(f) protein A

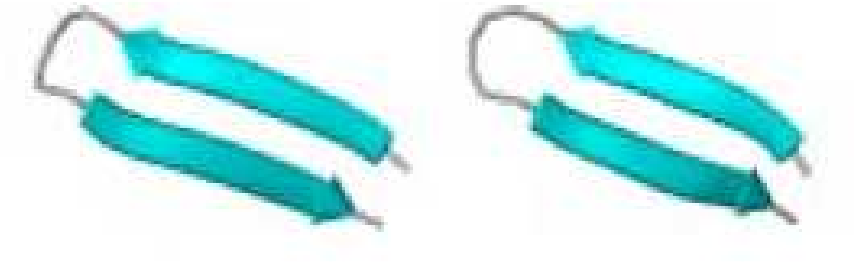




coarse-graining does not have an uniform effect on all modes, increases the clock by 1-2 orders of magnitude.

Derreumaux, Mousseau, JCP 2007.

Four examples of ab initio prediction



1PGB(41-56)
(Wei et al., Proteins 2004)

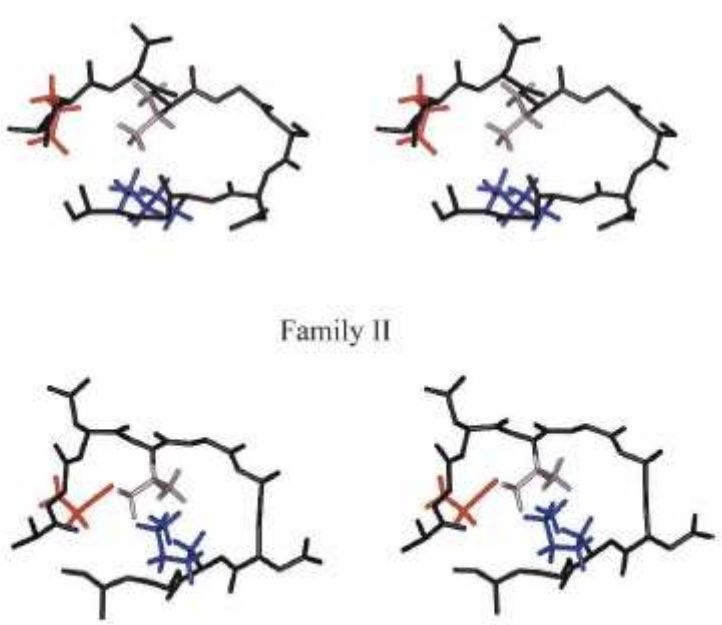
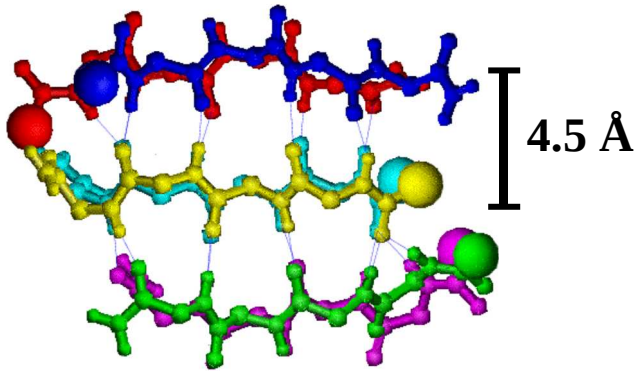


Figure 5. Final mean structures for the Aβ(21–30) conformers. The geometric average structure for each family of the Aβ(21–30) structures was determined using Analysis (InsightII) and minimized using Discover (InsightII). The entire backbone (black) and all residue side chains are shown. The side chains of residues Glu22, Val24, and Lys28, are colored red, gray, and blue, respectively.

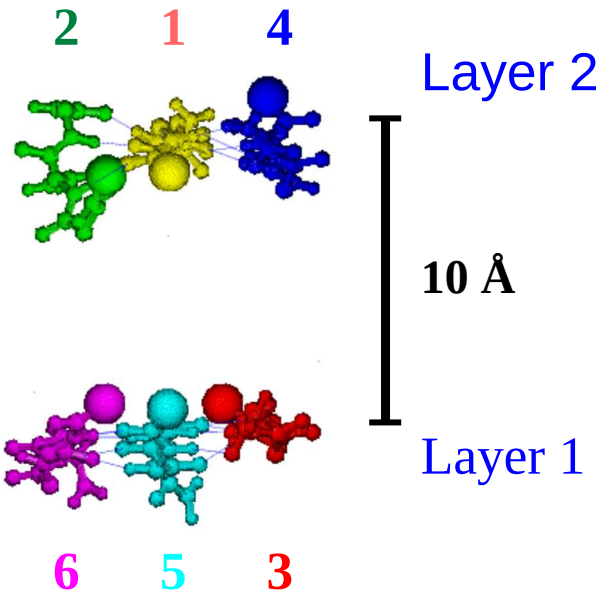
Aβ(21-30) monomer
(Chen et al., JCP 2006)

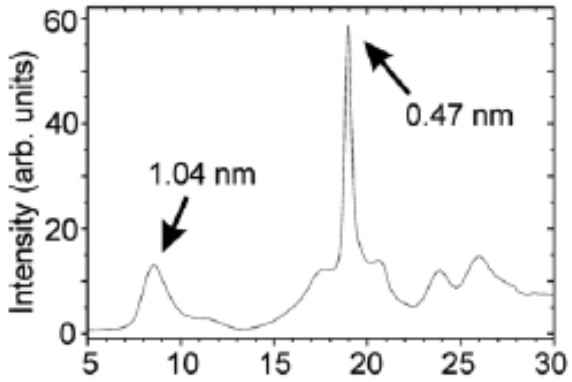
(Teplow et al., Protein Science 2005)

One local minimum for 6-chains of Aβ(16-22): two-layer β-sheets



β-strands ⊥ to fibril axis

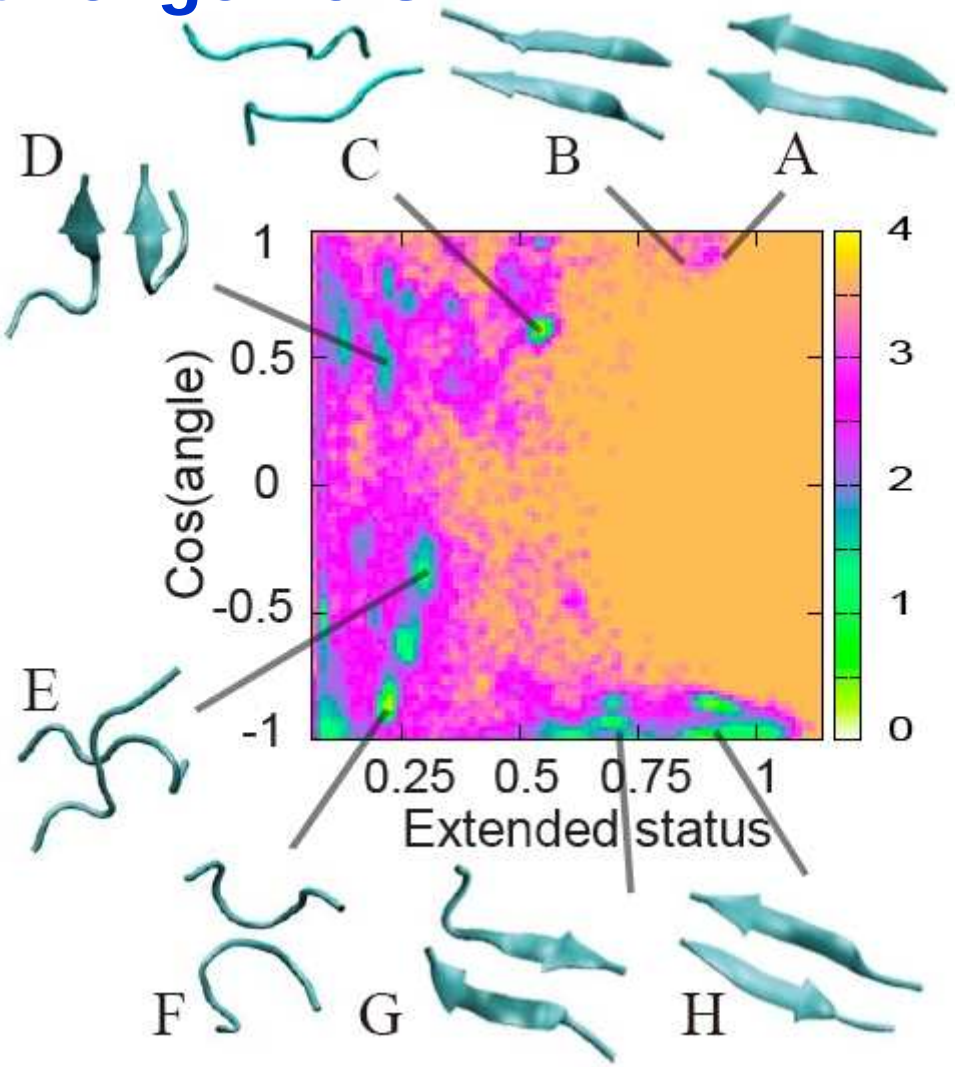




Equatorial and meridional reflections in the diffraction patterns of most amyloid fibers

Mousseau, Derreumaux, Acc. Chem. Res. 2006

Question 1: Thermodynamics and dynamics for small oligomers



dimer Aβ(16-22)

REMD-OPEP, 8 replicas, each of 50 ns

7 clusters from parallel sheets, cross conformations, amorphous and antiparallel sheets.

Boltzmann probabilities of A-H: 59%, Pmax (state F) = 21%, P(G+H)> P(B+A),

Overall 64% RC at 310 K

Wei, Mousseau, **Derreumaux, Prion, 1:1-6 (2007)**

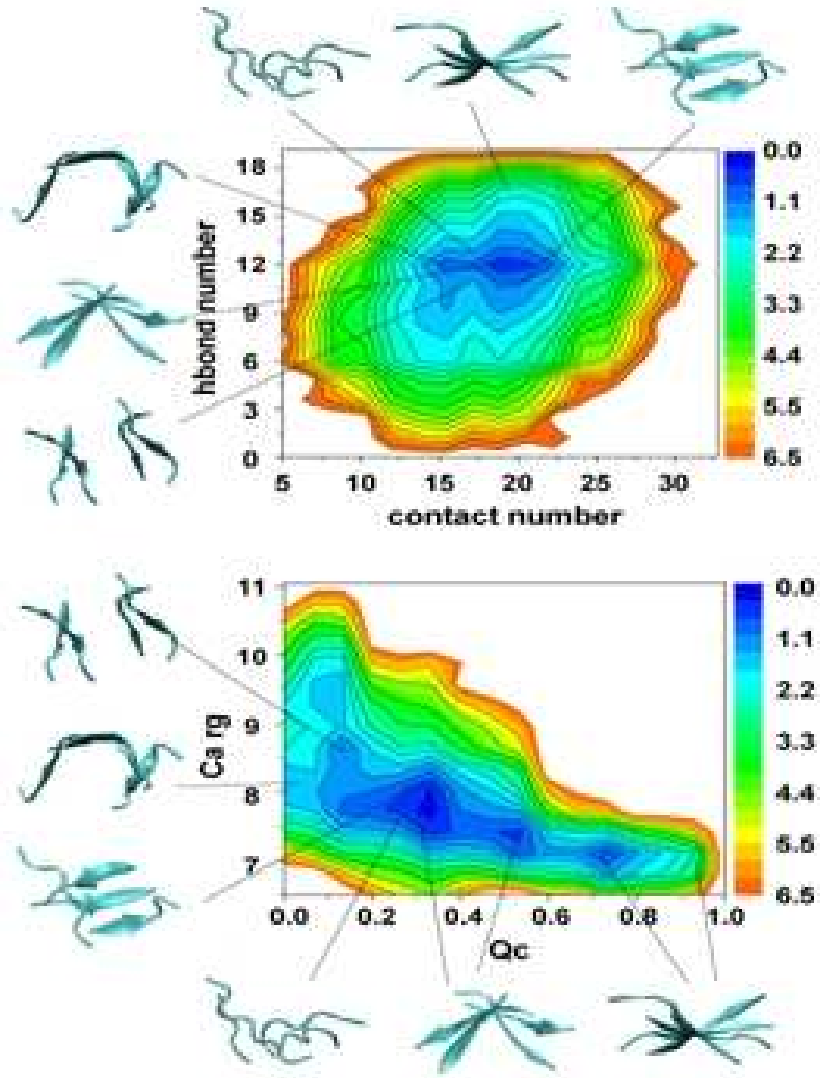


Figure 1: Free energy surfaces (in kcal/mol) of β2m (83-89) at 310 K as a function of: (a) number of inter-peptide side-chain contacts and number of inter-peptide H-bonds; (b) fraction of native inter-peptide side-chain contacts (Qc) and radius of gyration (rg) of the backbone Cα atoms.

4-mers of β2m(83-89)
20MD, each of 100 ns at 310 K + REMD

-> 6 clusters. Amorphous states share similar H-bond and Sc-Sc contacts with amyloid-competent states (Boltzmann prob: 17%)

-> Annular disordered forms occur

-> Small free energy barriers

***Song,Wei, Mousseau,* Derreumaux, J. Chem. Phys. in revision (2007)**

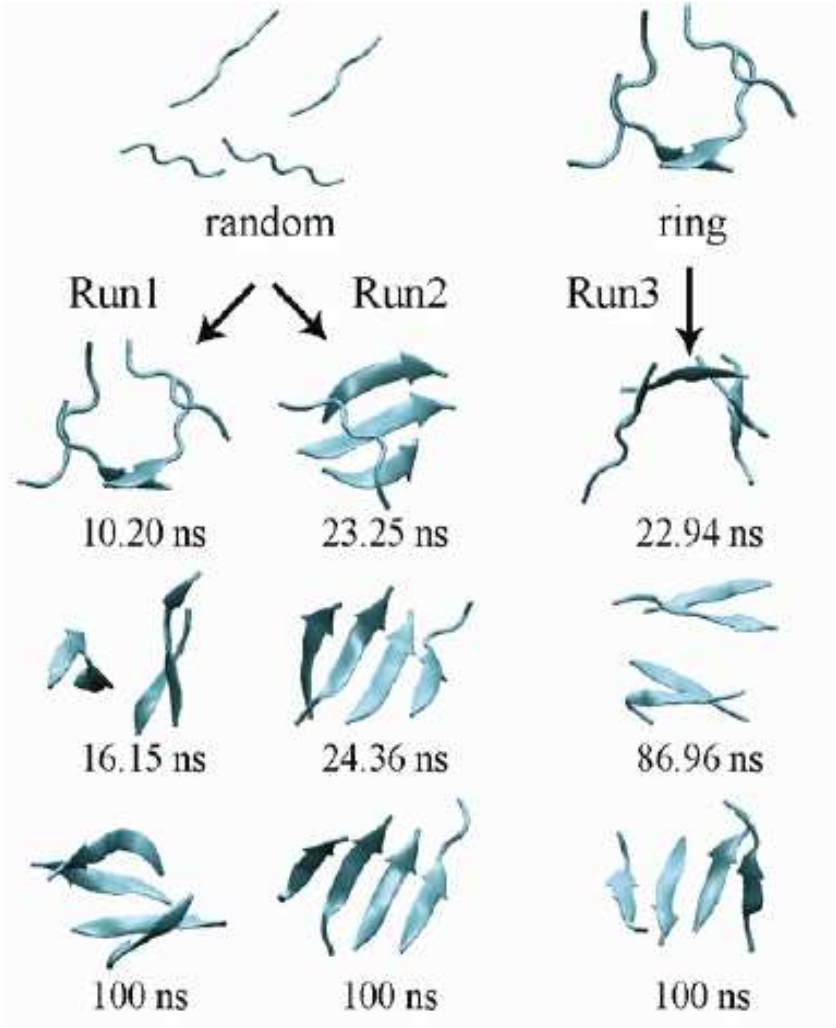


Figure 3. Three MD-OPEP simulations at 310 K starting from a randomly chosen state (Run 1 and Run 2) or the ring-like state generated by Run 1 (Run 3). Representative snapshots are shown between 0 and 100 ns.

MD analysis:
Species are in dynamic equilibrium

***Song,Wei, Mousseau,* Derreumaux,**
J. Chem. Phys. in revision (2007)

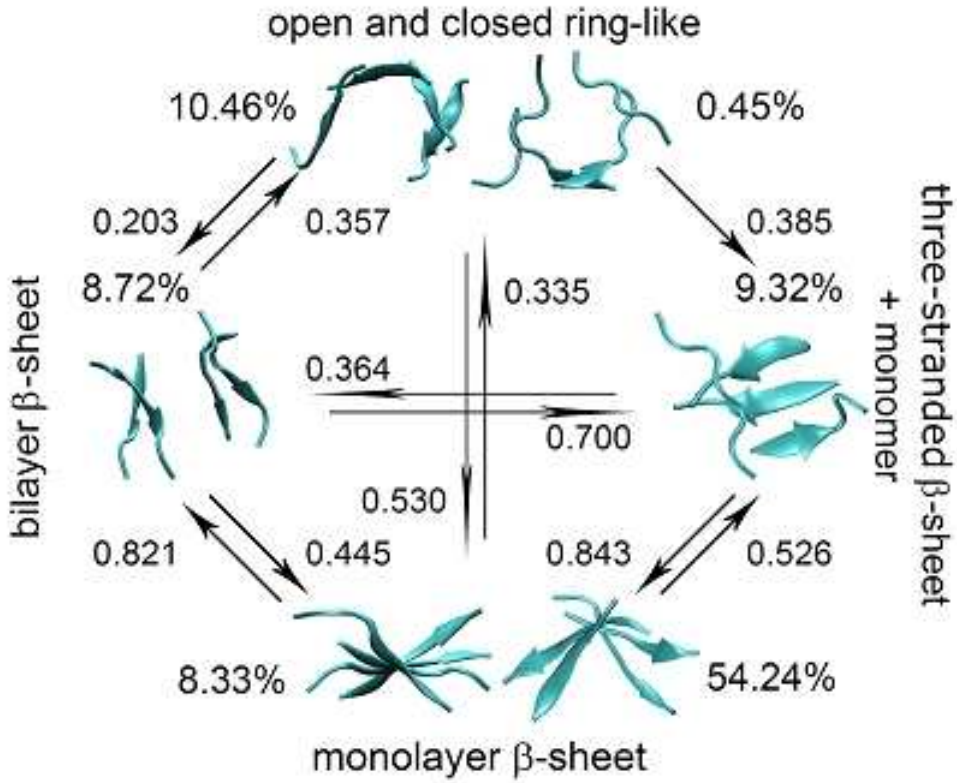


Figure2. Transitions and rates (in ns^{-1}) between the four most populated predicted topologies at 310 K. The β-sheet composition is not constant within each topology, and overall the population of β-sheet for all conformations is 42% (vs. 58% of random coil) by using the DSSP program.[26] For simplicity, we only show the parallel bilayer β-sheet (but orthogonal β-sheets exist), and the monolayer β-sheet with parallel chains (but various H-bond registries and mixed parallel/antiparallel organizations exist).

Timescales within 100-500 ns by all-atom MD

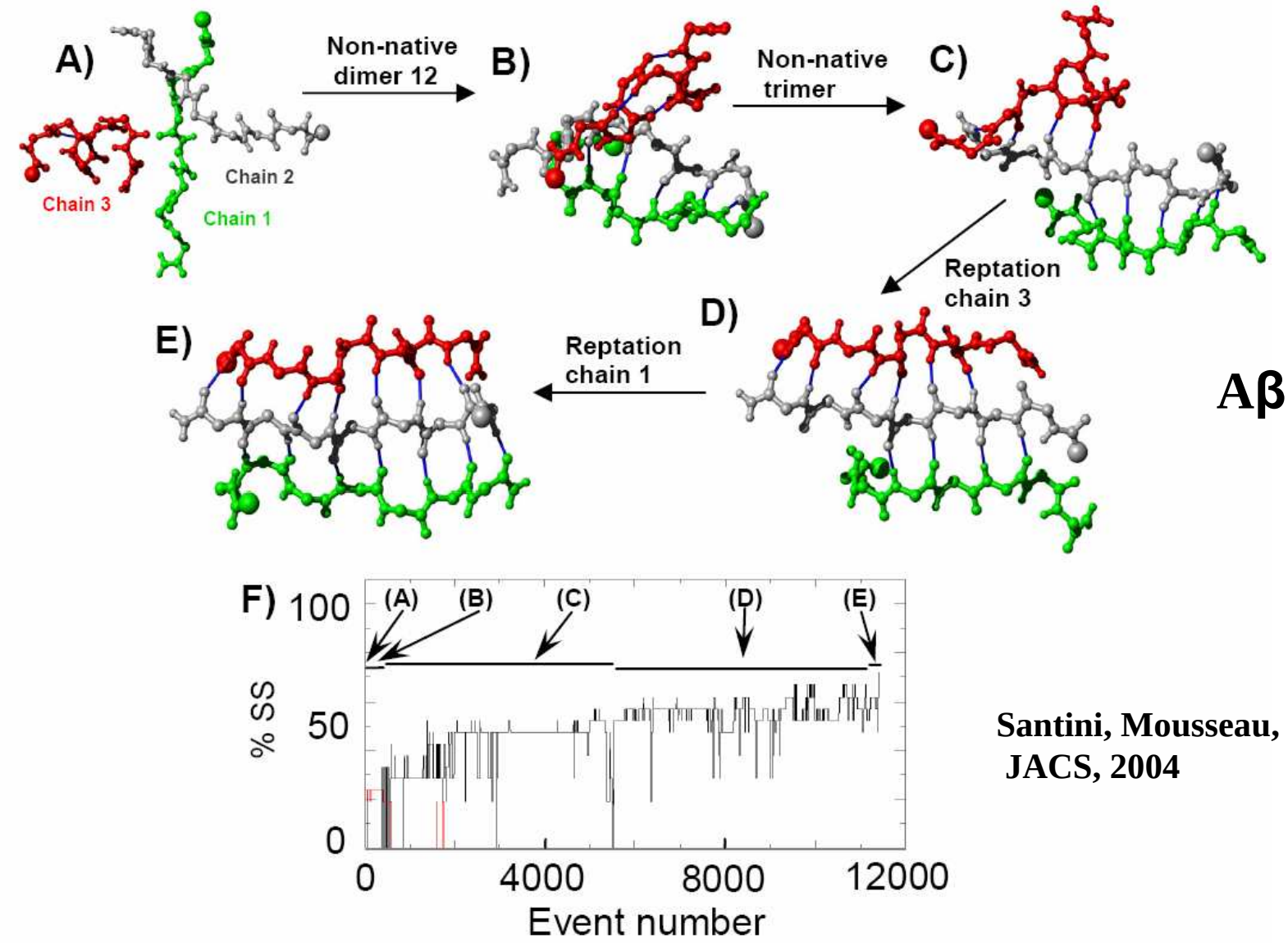
Question 2: Paths and mechanisms from random to fibrillar-like states (3-mers to 12-mers).

Two elementary oligomer growth mechanisms:

1. **Reptation move.**

2. **Dock and lock.**

1. Reptation mechanism

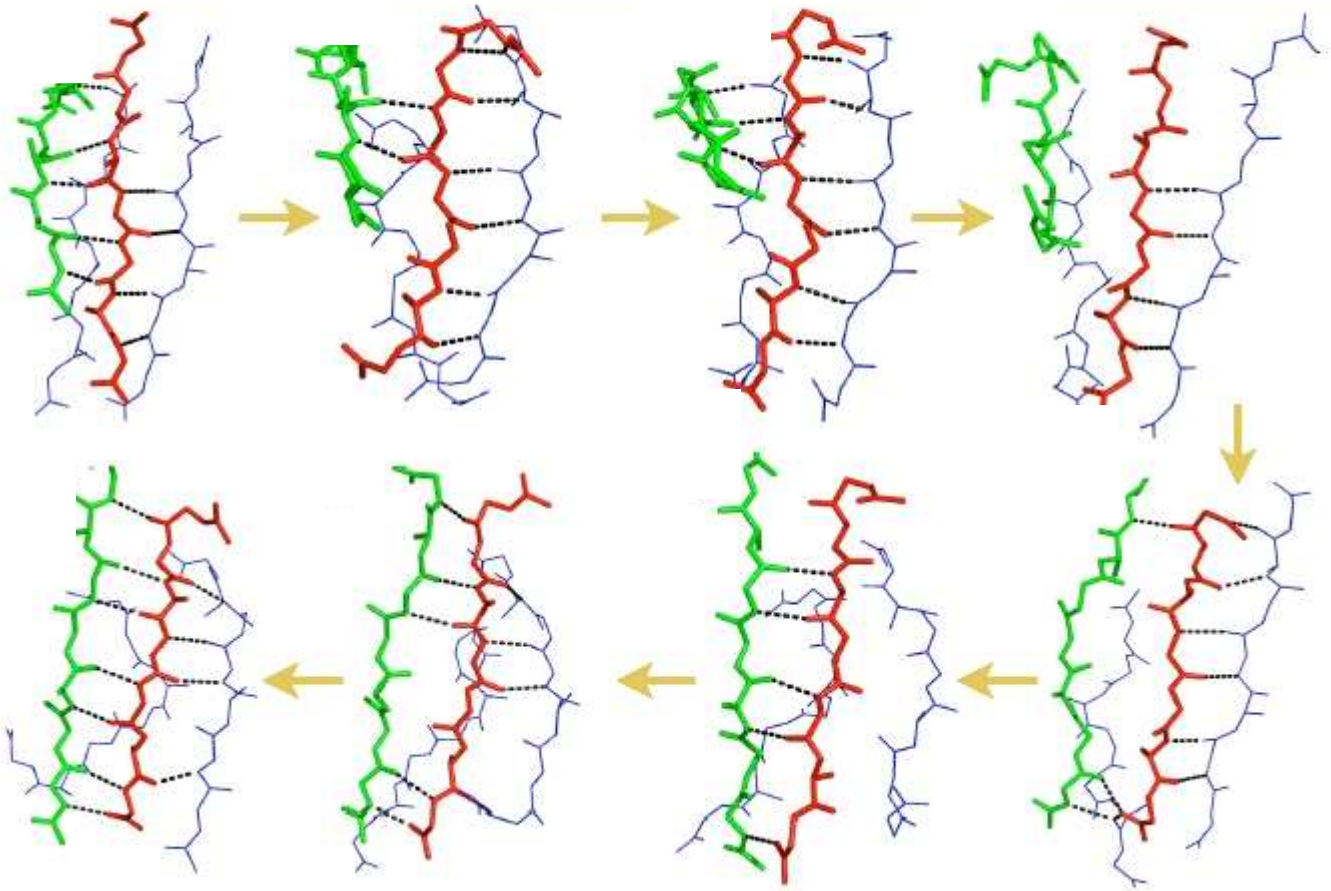


Aβ(16-22)

Santini, Mousseau, Derreumaux JACS, 2004

Reptation move observed by ART-OPEP consistent with isotope-edited IR spectroscopy study on Aβ(16-22) (Petty and Decatur, PNAS 2005).

2. Dock and lock process



4Aβ16-22, MD-OPEP, timescale: ~100 ns

Derreumaux, Mousseau, JCP 2007.

A two-step aggregation mechanism

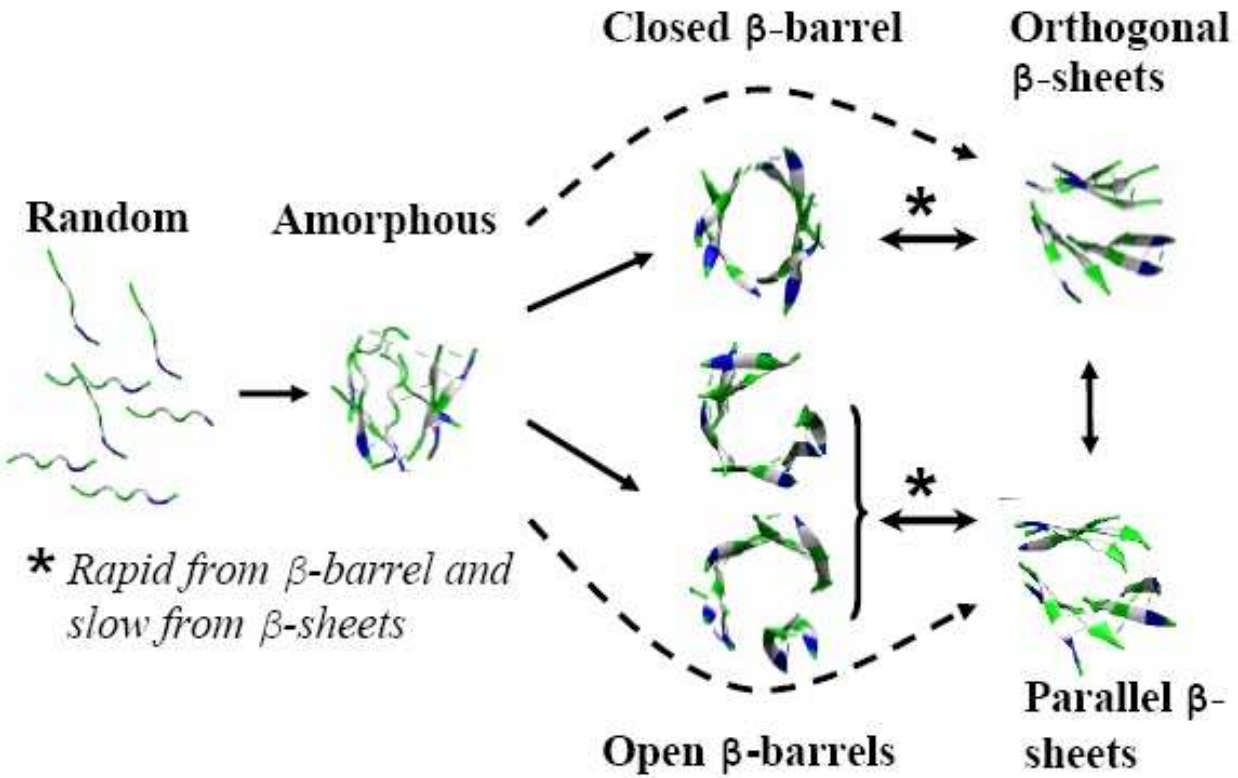
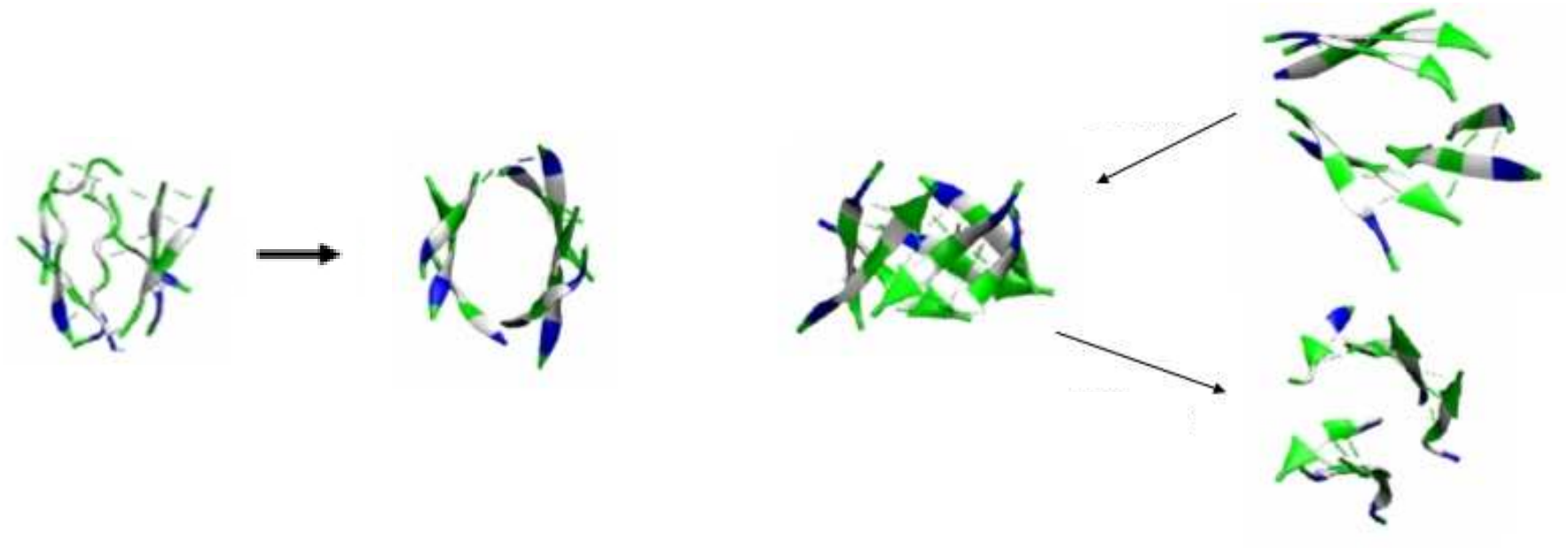


Fig.2 A generic aggregation picture derived from ART- and MD-OPEP simulations. Starting from a randomly chosen state, the peptides form amorphous aggregates. From there, the outcome changes with the oligomer size (OS) and chain length (L). For OS <9 and L < 8, rapid aggregation proceeds directly to ordered β-sheets or indirectly through β-barrels. The double arrows indicate reversibility. For larger OS or L, aggregation into β-barrels and ordered β-sheets is very slow and rare.

Models: KFFE, A(16-22), β2m(83-89), Aβ(11-25), NFGAIL from 4-mers to 12-mers.

β-barrel: a non-obligatory intermediate



7 chains of β2-microglobulin (83-89):
MD-OPEP, 310 K, **5 µs, timescale~500 ns**

This barrel intermediate has never been detected experimentally. Question: impact of all-atom force field and explicit solvent? Next: GROMOS-SPC, REMD, 16 replicas of 60 ns.

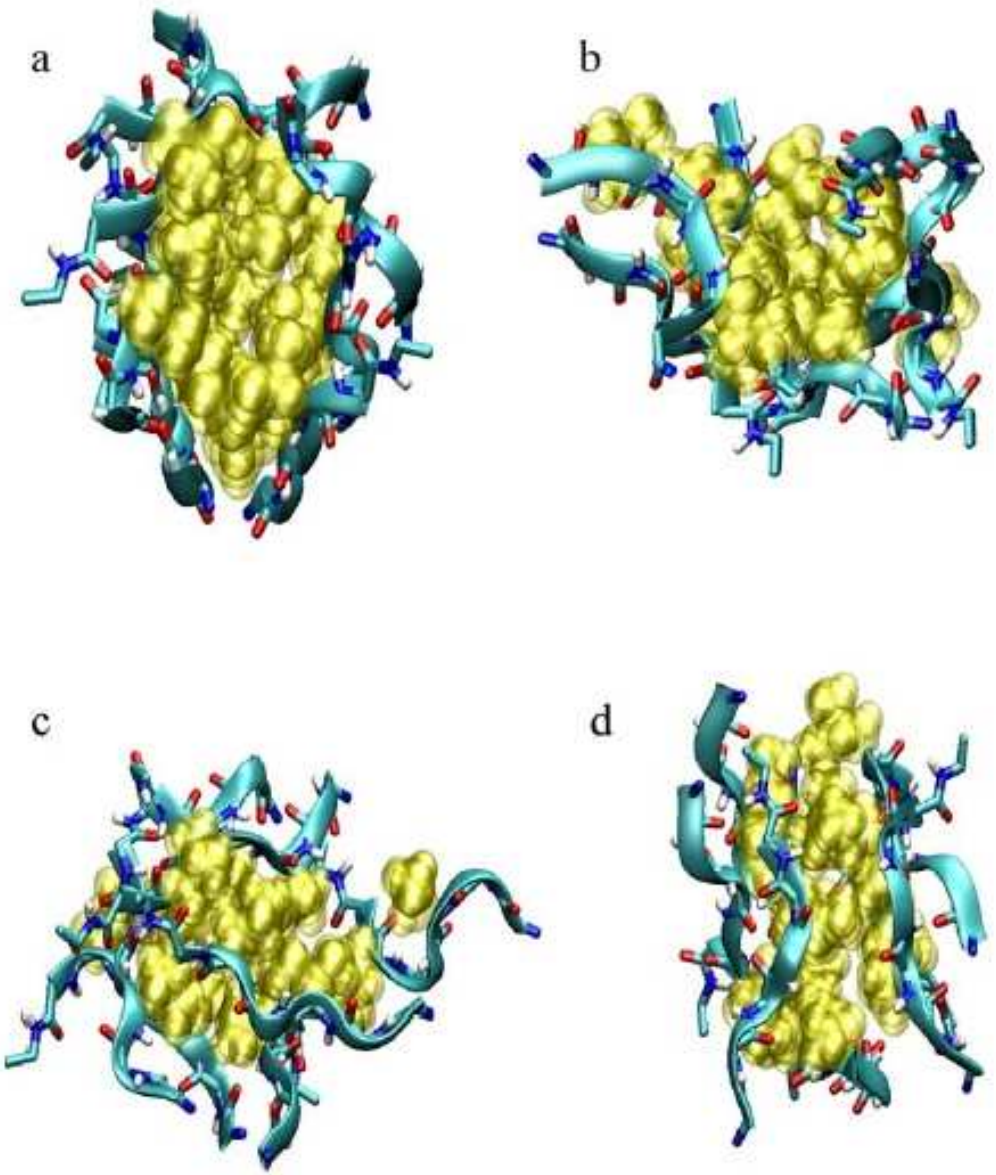


Figure 7. Representative structures of the sampling. a) elliptic barrels. b) open barrels. c) amorphous aggregates. d) 6-chains barrels. The main chain is shown by sticks and cyan ribbons. Yellow Van der Waals spheres represents the hydrophobic residues.

The β-barrel is also found in equilibrium with the steric zipper!!

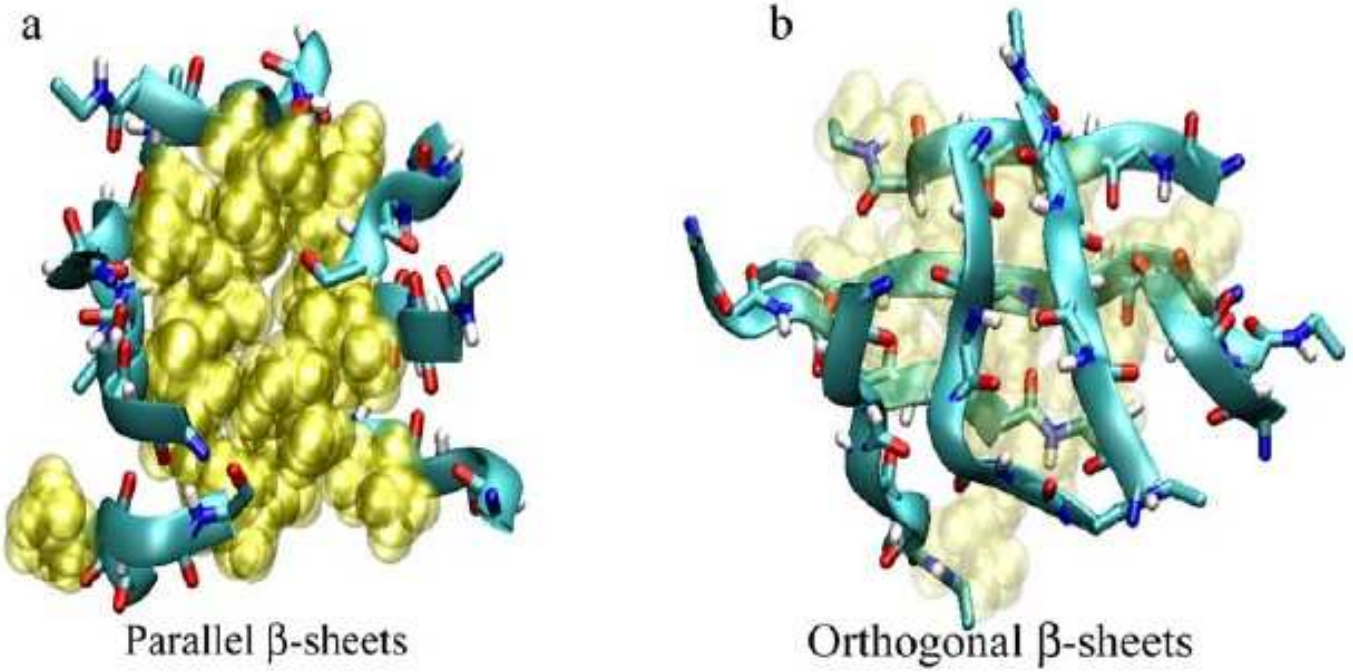


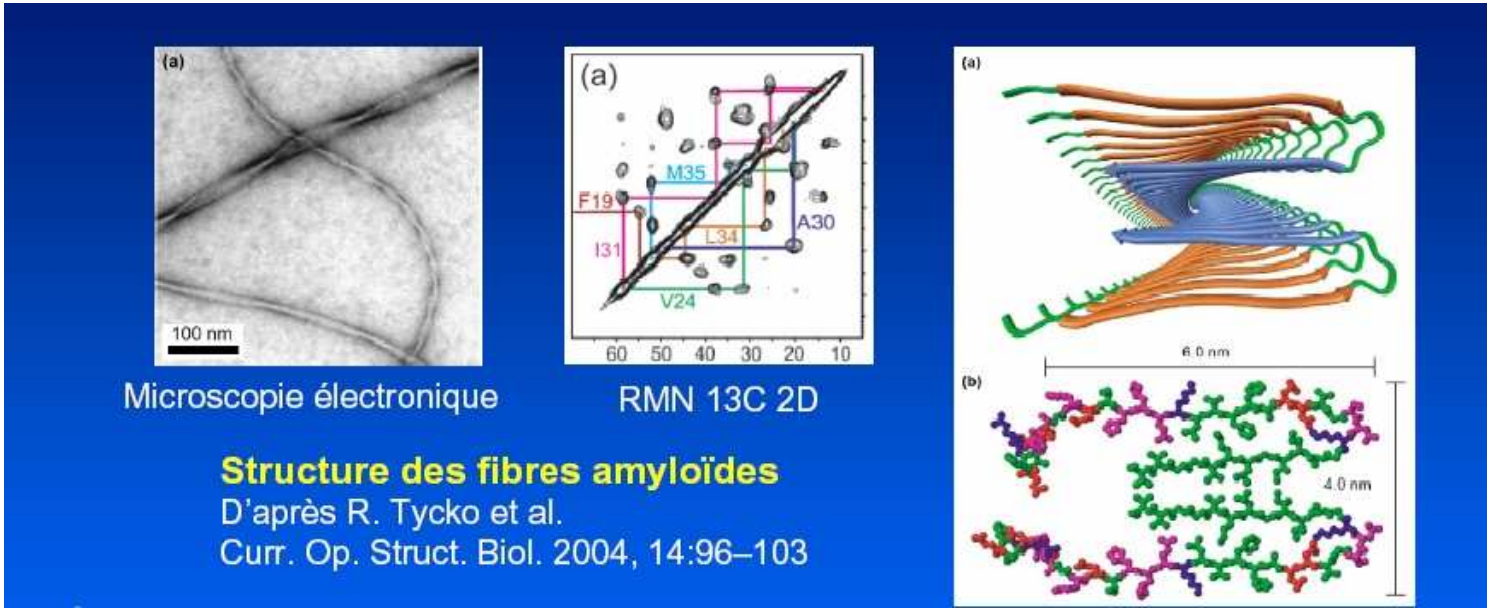
Figure 8. Steric zippers. a) Parallel beta-sheets. b) Orthogonal beta-sheets.

A. De Simone, P. Derreumaux, in preparation

Question 3: What are the equilibrium structures of full-length Aβ dimers?

. FRET and gel filtration chromatography suggests stable dimers at low concentration (Glabe et al., JBC, 1997)

. Formation of the loop 23-28 might nucleate folding of Aβ monomer (Teplow et al., Prot. Sci., 2005) and be the rate-limiting step in Aβ fibrillogenesis (Meredith et al., Biochemistry, 2005)



REMD-OPEP Aβ(1-42)
32 replicas between 300 and 700 K, 120 ns/replica

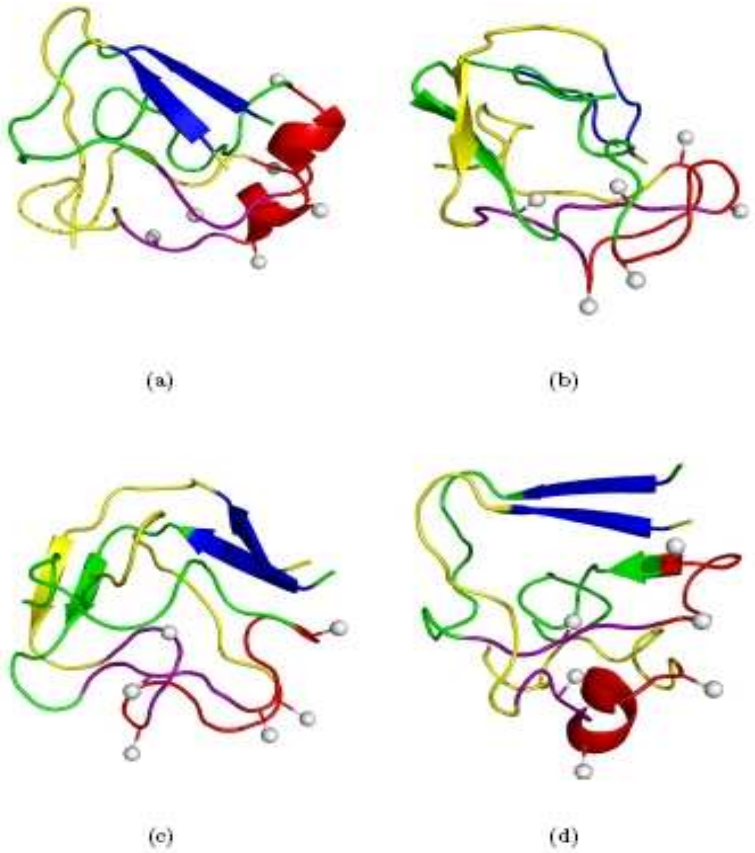


FIG. 2: Dimers of Aβ42. The four most populated REMD-OPEP structures are essentially random coil in character, but display local structures: (a) short α-helices at positions Glu23-Asn27 and Asp22-Ser26 and a β-sheet at Ala2-Arg5 in both chains; (b) two parallel β-strands at positions 11-14; (c) two short parallel β-sheets in the terminal residues and (d) a short α-helix at positions Gly21-Ser26 in chain 1 and a three-stranded parallel β-sheet. Chain 1 is in green, chain 2 in yellow. Residues 2-5 are colored in blue, residues 17-21 in purple and residues 22-28 in red. The positions of the side chains of Phe19, Glu22 and Lys28 are indicated by balls.

- 20 clusters with Prob between 0.10 and 0.02.

- Averaged β-sheet: 0.2, α-helix: 0.05 vs. exptl: 10-20%, <10%

- Probability of Glu22-Lys28 and Asp23-Lys 28 contacts:
intra: 0.36 and 0.42,
inter: 0.14 and 0.24

→These salt bridges form in a minority of structures and can act as a seed in higher-order species.

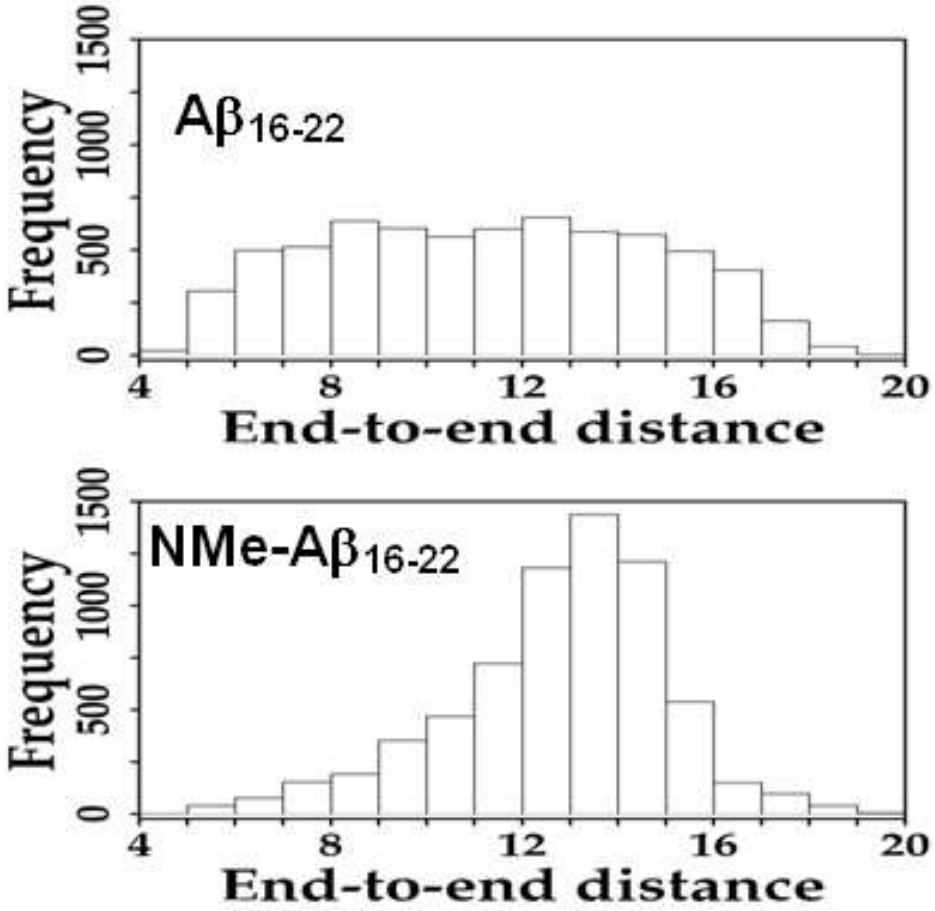
- But fibril formation requires Lys28 tobe buried. Tarus et al. estimated a free energy cost of 7 kcal/mol.

→ We propose that the formation of a multimeric beta-sheet spanning amino acids 17-21 is an contributes to the rate-limiting step

Melquiond, Dong, Mousseau, Derreumaux, Current Alzheimer Disease, 2007

Question 4: Inhibiting mechanisms of NME-Aβ16-22

Ac-K-(NmeL)V(NmeF)F(NmeA)E-NH_2



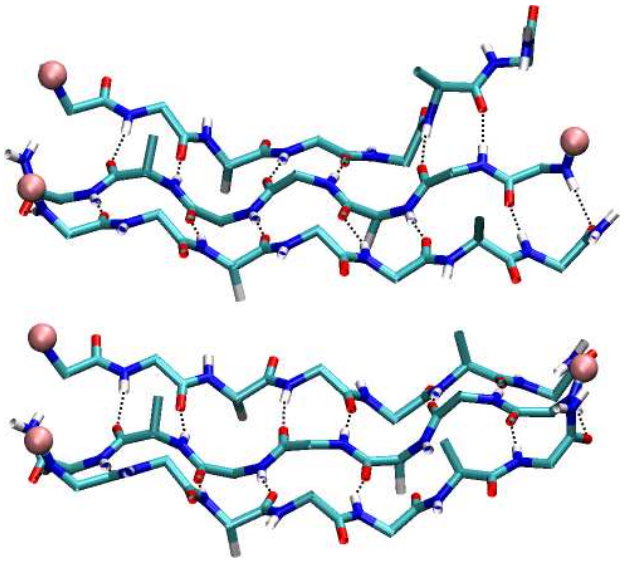
Ultracentrifugation and CD suggest NMe-Aβ(16-22) is monomeric and a β-strand (Meredith, 2001).

MD-OPEP : (RC, β-strand) % = (75 , 25) in wt and (10, 90) in inh

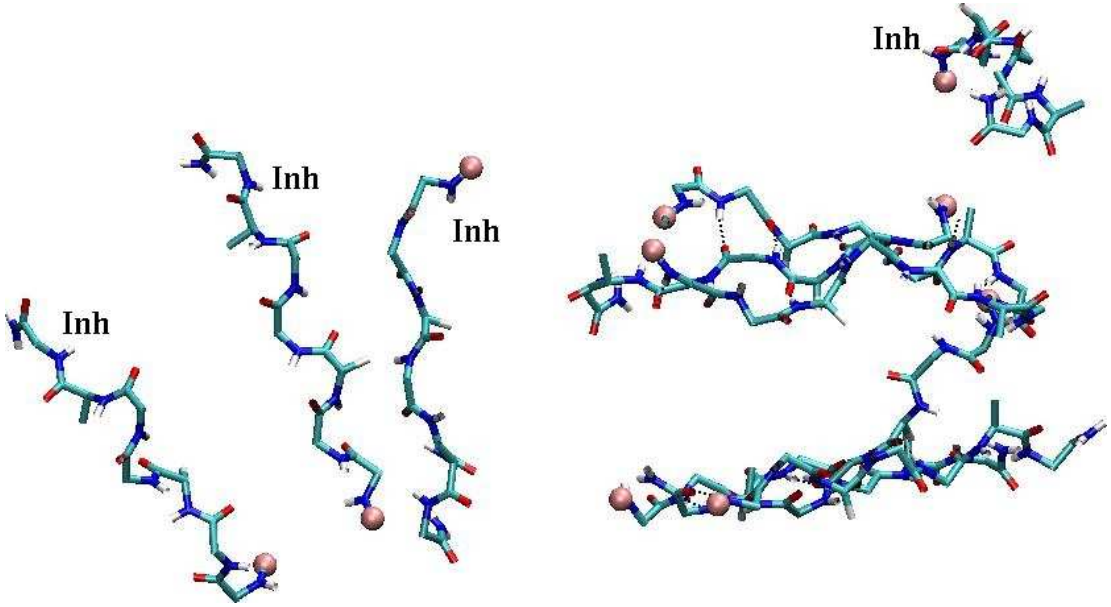
Figure 2. Histograms of the end-to-end distances of $A\beta_{16-22}$ and NMe-$A\beta_{16-22}$ at 300 K.

As a first step towards studying the dynamics of the interactions between fibrils and inhibitors:

6 chains of $A\beta_{16-22}$ (wt)

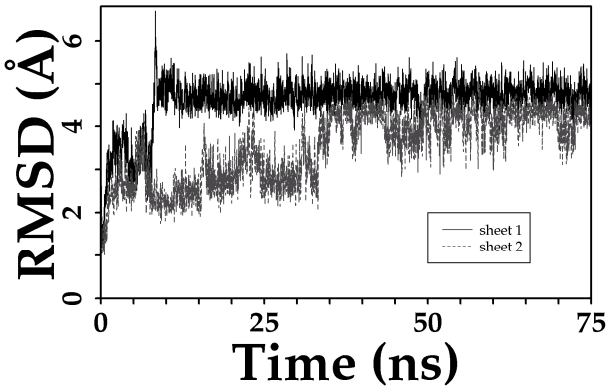


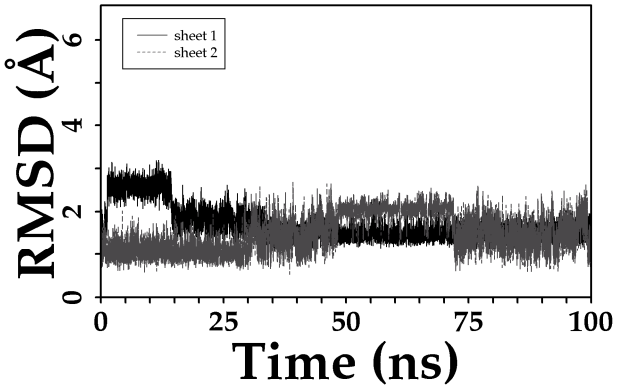
6 chains of $A\beta_{16-22}$ (wt) + 4 chains of N-methylated $A\beta_{16-22}$ (inh)

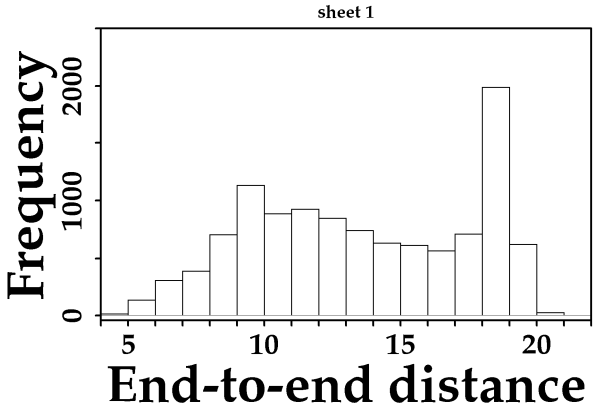


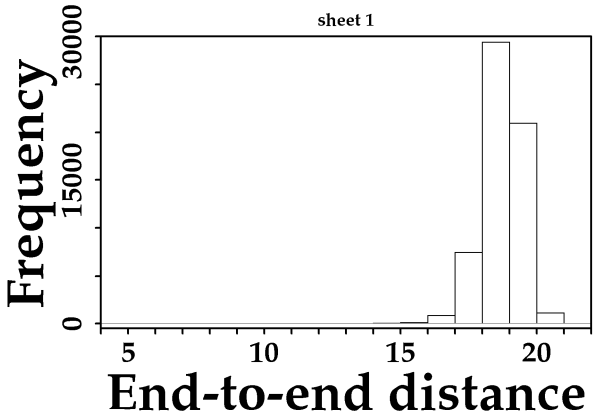
wt

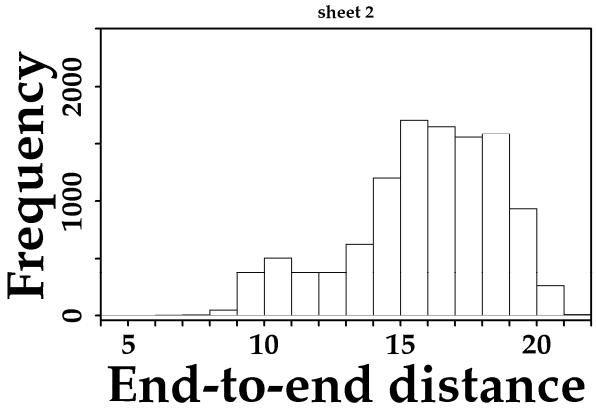
wt + inh

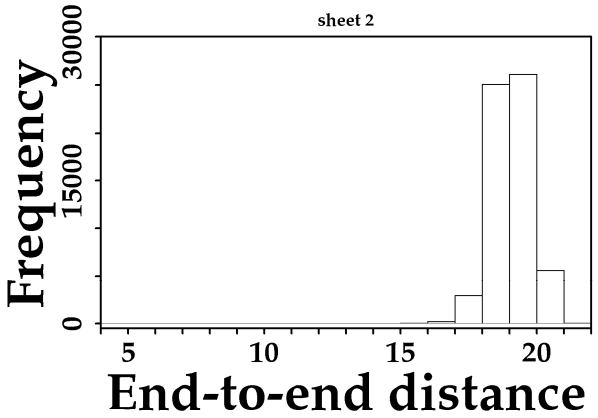






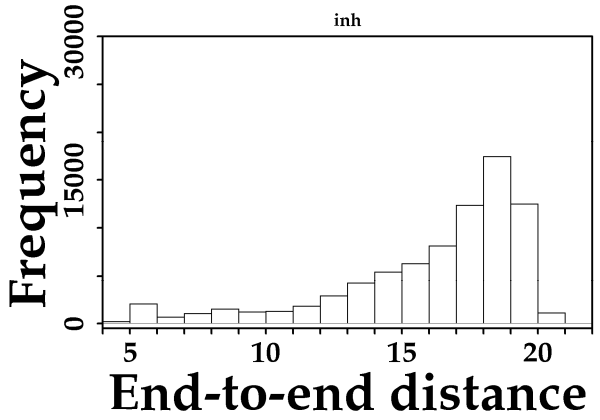


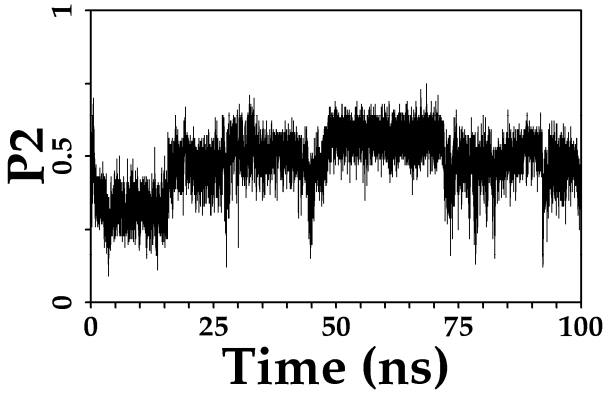


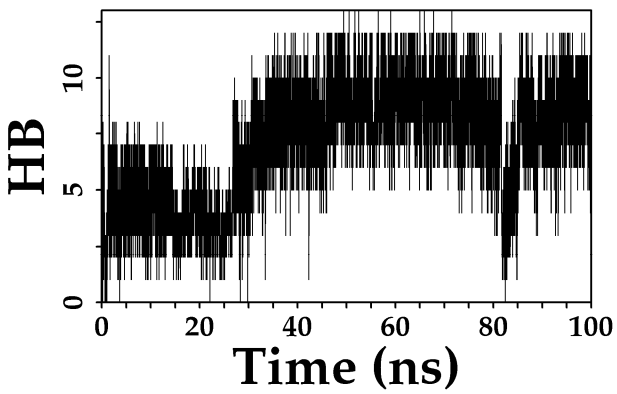


NMe inh stabilizes Aβ(16-22) hexamers

Properties of the N-methylated chains







P2 (nematic order parameter) describes the orientational order of the system and discriminates between ordered and disordered conformations.

$$\overline{P_2} = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}\left(\hat{z}_i \cdot \hat{d}\right)^2 - \frac{1}{2}$$

$$Q_{\alpha\beta} = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}e_{i\alpha}e_{i\beta} - \frac{1}{2}\delta_{\alpha\beta}$$

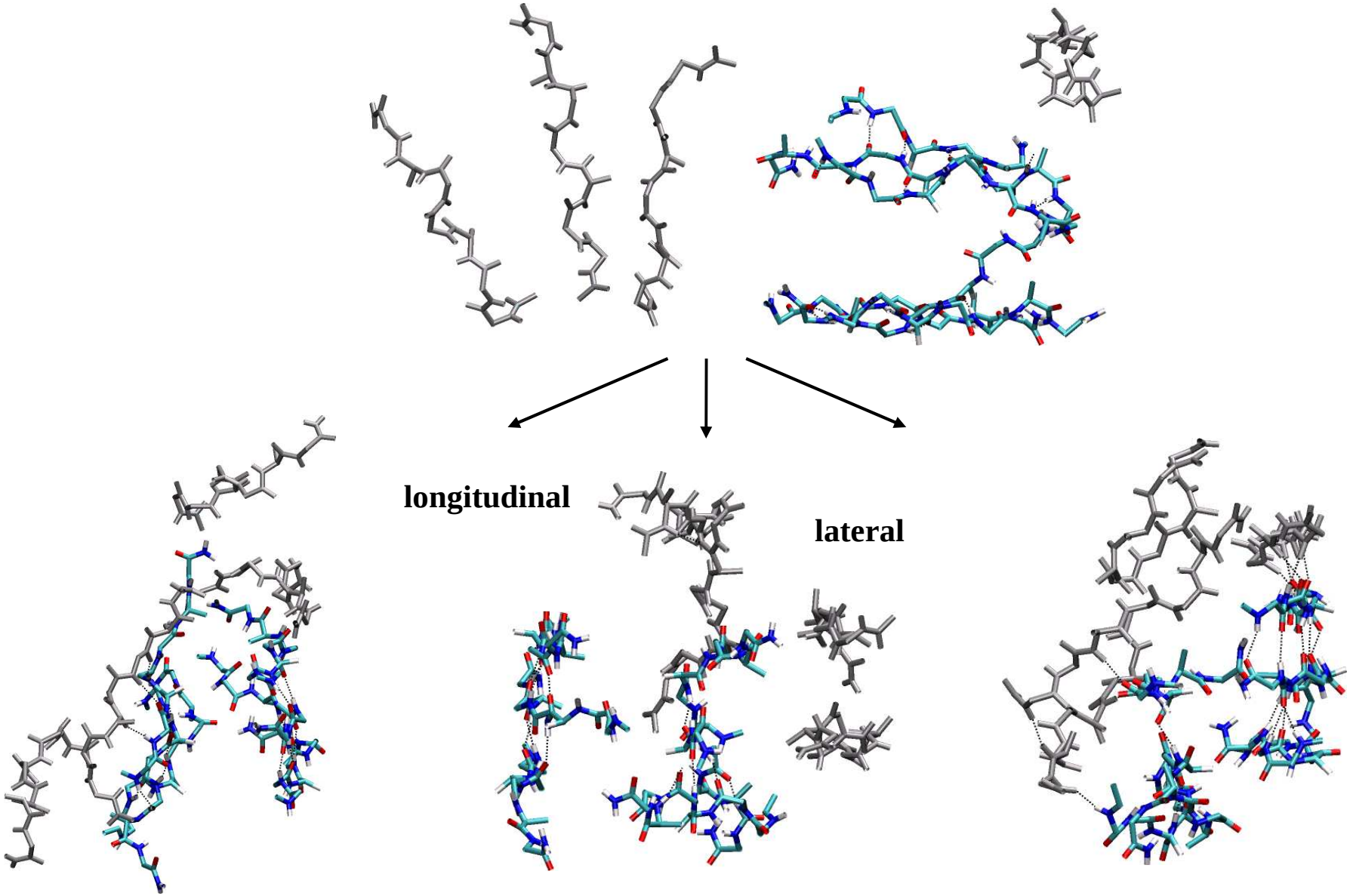
$(\alpha, \beta = x, y, z)$

$\hat{z}_i$ = end-to-end molecular vector

N = number of molecules in the simulation box

$\hat{d}$ = unit vector defining the preferred direction of alignement (eigenvector associated to the largest eigenvalue of the ordering matrix Q)

Interaction sites



Conclusions

. Free energy surfaces of 4(-7) mers reveal all the transient oligomers reported experimentally for large proteins. The low population of fibril-like states indicates the nucleus associated with fibril growth is larger than 4-(7)-mers for peptides. Sequence-dependent must be however considered

. Fibril formation is a two-step mechanism dominated by both H-bonds and hydrophobic interactions in the early steps (amorphous states) and then by H-bonds in the late steps (barrel or fibrillar-like).

. The region 23-28 does not drive folding of Aβ42 dimers, formation of a multimeric sheet at positions 17-21 is a very important factor contributing to the rate-limiting step.

. Simulations indicate longitudinal and lateral associations between oligomers and inhibitors blocking fibril extension, but destabilization of fibrils is a slow process.

We are currently studying tetramers and dodecamers of PrP(125-231) Aβ40 and Aβ42.

N. Mousseau, J-F. Saint Pierre (Univ. Montreal, Canada)
W. Chen, X. Dong (IBPC, Univ Montreal)
S. Santini (Univ. Bruxelles)
G. Wei, W. Song (Univ. Fudan, Shanghai, China)
A. Melquiond, J-C. Gelly (IBPC, Paris)
P. Aucouturier (Hopital Saint-Antoine, Paris)
A. De Simone (CNR, Napoli, Italy)

Alzheimer Society of Canada, CNRS,

Université Paris VII, 6th European PCRD