Facoltà di Fisica, scuola di dottorato
A.A. 2014-2015

Università di Pisa

Simulation of proteins: Coarse Graining to minimalist models

Andrea Giuntoli

Summary

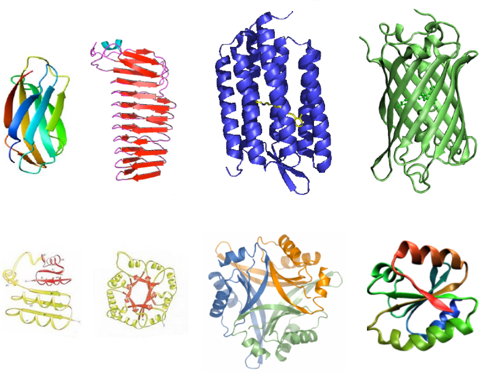
1. The structure of proteins
2. Proteins modeling
3. Minimalist models
4. Conclusions

Why should we study proteins?
A wide spectrum of different functions

Proteins: finely structured, highly specialized biomolecules.

Typical size: $1 - 10nm$
Functions:

- Catalysis (enzymes)
- Regulatory
- Structural
- Protection
- Energetics



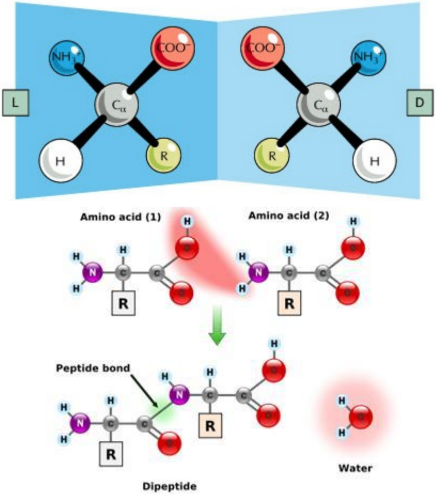
Hierarchic organization: primary, secondary, tertiary and quaternary structures

Primary structure
Amino acids

Amino acids: the basic elements of a protein.

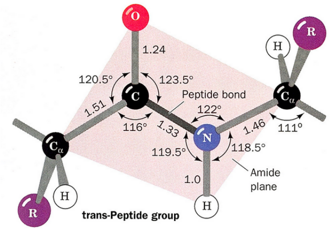
Chirality of the central carbon atom (Cα).

Connection through the peptide bond: directionality of the chain.

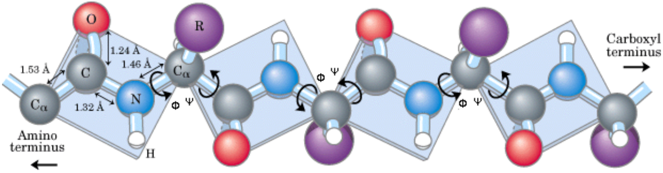


Primary structure
Peptide bond

Stiffness of the peptide bond: planar configuration.



the primary structure is given by the amino acidic sequence

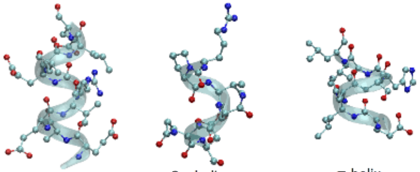


Φ and Ψ dihedrals: a complete set of degrees of freedom for the backbone

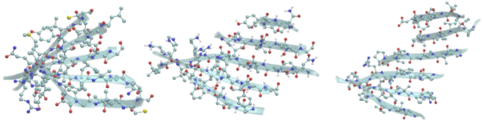
Secondary structure
uniform structures

The secondary structure is determined by the (Φ, Ψ) values along the chain.

Main "uniform structures": helices and extended structures (beta-strand and beta-sheets).



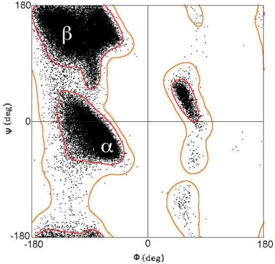
Each uniform structure is characterized by a given (Φ, Ψ) value.

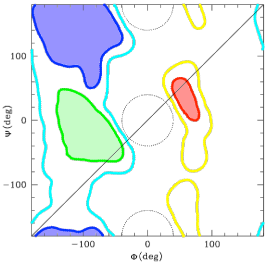


Secondary structure
Ramachandran Plot

Ramachandran Plot: mathematical description of the secondary structure in the Φ-Ψ plane.

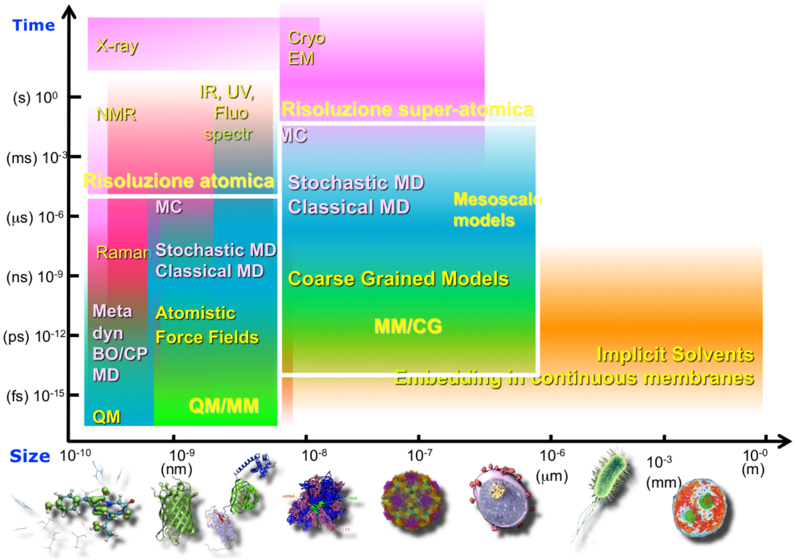
densely-packed areas: main uniform structures.
Sparse areas: ”forbidden regions” due to sterical hindrances.





Why different models?

space and time scales of interest

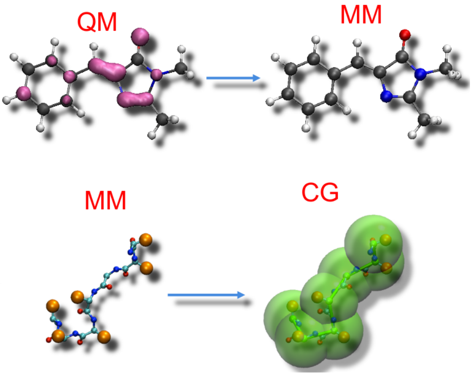


How much accuracy do we need?
resolution of the models

How to build a model:

- definition of the degrees of freedom
- description of the interactions
- dynamics of the system

Simplification is needed in order to reach longer simulation times: atomistic and Coarse Grained models.

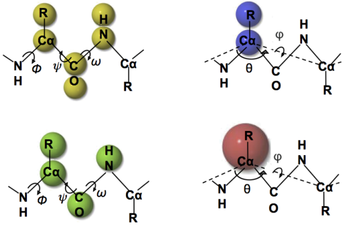


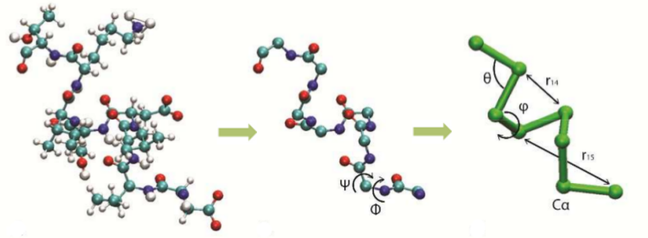
Coarse Grained models
Definition of the degrees of freedom

New variables $Q_I = Q_I(\{R_i\} \in B_I)$

Classification based on the number and position of the beads per amino acid.

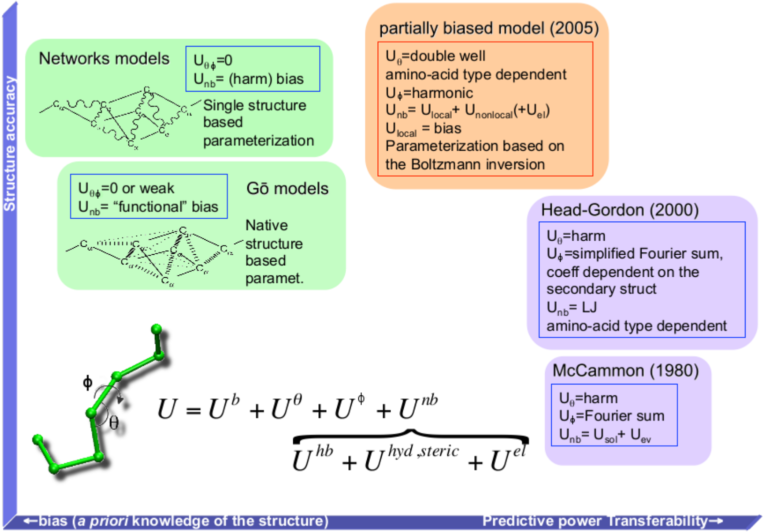
Minimalist models: 1 bead placed on the Cα.





Coarse Grained models

Force Field in the minimalist models



Parametrization methods
Boltzmann inversion

Parametrization based on the distribution of internal variables, assuming thermal equilibrium.

$$P(Q_1, ...Q_n) = P_0(Q_1, ..., Q_n)\exp[-(U(Q_1, ..., Q_n))/kT]$$

$$\Downarrow$$

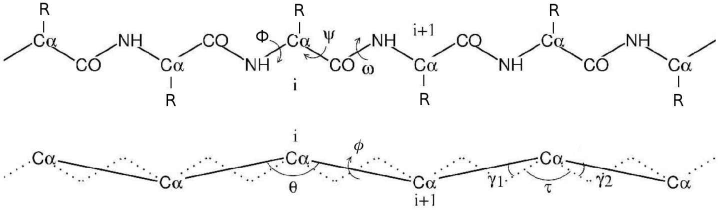
$$U(Q_1, ..., Q_n) = -kT\ln\left(\frac{P(Q_1, ...Q_n)}{P_0(Q_1, ...Q_n)}\right)$$

P_0 distribution of the non-interacting system.

In case of independent variables:

$$P(Q_i) = P_0(Q_i)\exp[-(U(Q_i))/kT] \Longrightarrow U(Q_i) = -kT\ln\left(\frac{P(Q_i)}{P_0(Q_i)}\right)$$

Mapping to the minimalist models
how to prevent loss of information

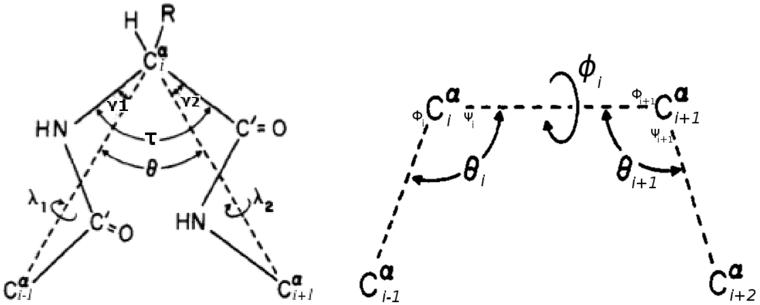


Analytical mapping: $(\Phi, \Psi) \leftrightarrow (\theta, \phi)$.

- mapping information from the atomistic resolution level
- helpful tool to approach the PBRP (Protein Backbone Reconstruction Problem)
- bottom-up approach to the parametrization problem

Mapping to the minimalist model

the analytical mapping



$$\theta_i(\Phi_i,\Psi_i) = \arccos[\cos\tau(\cos\gamma 1\cos\gamma 2 - \sin\gamma 1\sin\gamma 2\cos\Phi_i\cos\Psi_i) - \sin\gamma 1\sin\gamma 2\sin\Phi_i\sin\Psi_i + \sin\tau(\cos\Psi_i\sin\gamma 2\cos\gamma 1 + \cos\Phi_i\cos\gamma 2\sin\gamma 1)]$$

$$\phi_i(\Phi_i,\Psi_i,\Phi_{i+1},\Psi_{i+1}) = 180^\circ + \lambda_2(\Phi_i,\Psi_i) + \lambda_1(\Phi_{i+1},\Psi_{i+1}) \simeq 180^\circ + \Psi_i + \Phi_{i+1} + \gamma 1\sin\Phi_i + \gamma 2\sin\Psi_{i+1}$$

Mapping the Ramachandran Plot

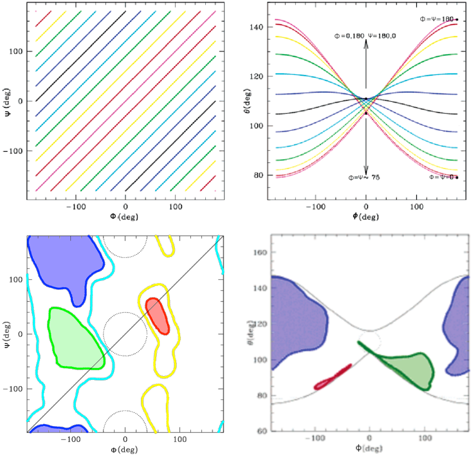
Uniform structures:

$\Phi_i = \Phi_{i+1},\ \Psi_i = \Psi_{i+1}$

Mapping 2→2:

$(\Phi, \Psi) \rightarrow (\theta, \phi)$

- chirality:
  $\theta(\Phi, \Psi) = \theta(-\Phi, -\Psi)$
  $\phi(\Phi, \Psi) = -\phi(-\Phi, -\Psi)$
- directionality:
  $\theta(\Phi, \Psi) = \theta(\Psi, \Phi)$
  $\phi(\Phi, \Psi) = \phi(\Psi, \Phi)$

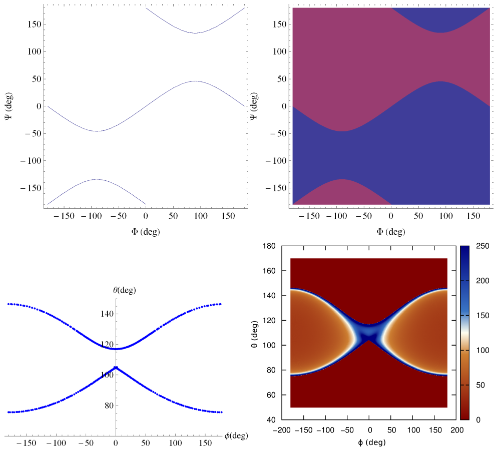


The Jacobian Matrix
mapping the density distributions

$$J(\Phi,\Psi) = \begin{bmatrix} \frac{\partial\theta(\Phi,\Psi)}{\partial\Phi} & \frac{\partial\theta(\Phi,\Psi)}{\partial\Psi} \\ \frac{\partial\phi(\Phi,\Psi)}{\partial\Phi} & \frac{\partial\phi(\Phi,\Psi)}{\partial\Psi} \end{bmatrix}$$

Formal description of:

- bijective regions of the mapping
- density distributions in the (θ,ϕ) plane



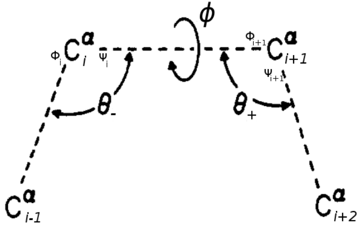
$$\rho(\theta,\phi) = \mid \det J^{-1}(\theta,\phi) \mid \rightarrow \bar{f}(\theta,\phi) = f(\Phi(\theta,\phi),\Psi(\theta,\phi))\rho(\theta,\phi)$$

Unstructured case

Symmetries of the system and variables distribution

Generalization of the mapping: $(\Phi_i, \Psi_i, \Phi_{i+1}, \Psi_{i+1}) \to (\theta_-, \theta_+, \phi)$
Symmetries:

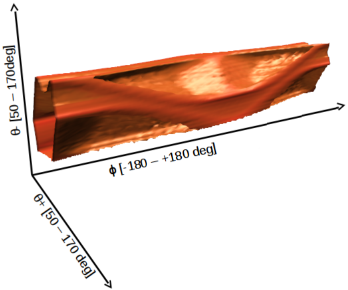
- $(\Phi_i, \Psi_i, \Phi_{i+1}, \Psi_{i+1}) \leftrightarrow$
  $(-\Phi_i, -\Psi_i, -\Phi_{i+1}, -\Psi_{i+1})$
  $\Rightarrow (\theta_-, \theta_+, \phi) \leftrightarrow (\theta_-, \theta_+, -\phi)$
- $(\Phi_i \leftrightarrow \Psi_{i+1})$, $(\Psi_i \leftrightarrow \Phi_{i+1})$
  $\Rightarrow (\theta_-, \theta_+, \phi) \leftrightarrow (\theta_+, \theta_-, \phi)$



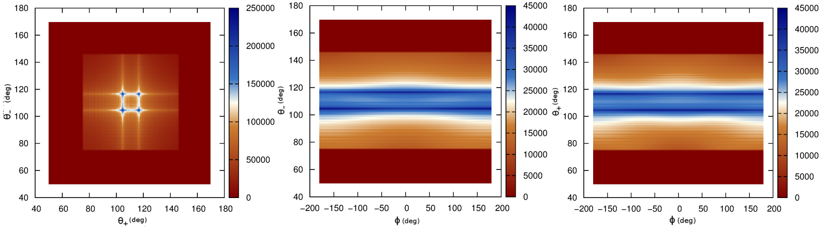
The Jacobian of the mapping is not defined.
$\rho(\theta_-, \theta_+, \phi)$ derived numerically.

Unstructured case
Three and two variable correlations

Three variables distribution
$\rho(\theta_-, \theta_+, \phi) \rightarrow U^T = -kT\ln\rho$



Two variable correlations vanish in the general case



What have we learned?

an eye turned to the future

Results of the new approach

- More insight on the minimalist models
- Universality features
- Proposal of a new normalization potential U^T

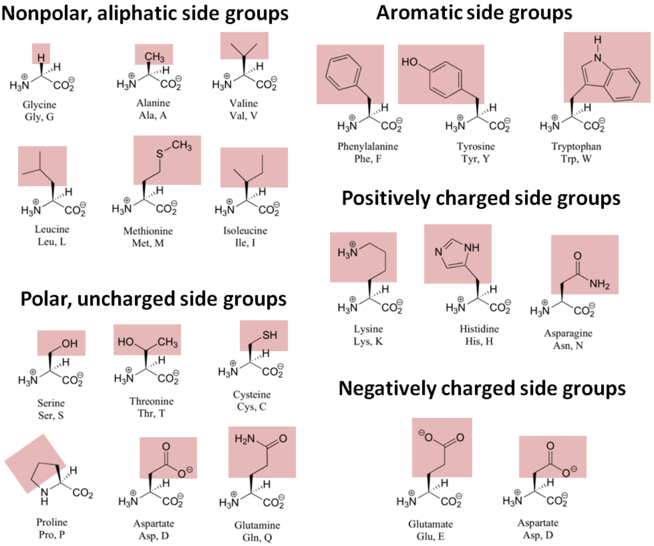
What now?

- Extension to non-uniform Ramachandran Plots
- Implementation of U^T in simulations
- Comparison with experimental results

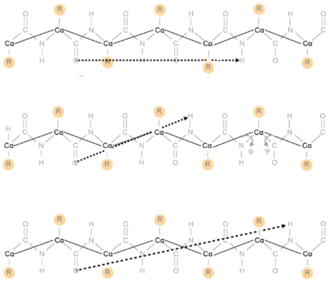
Thanks for your attention, any question is welcomed!

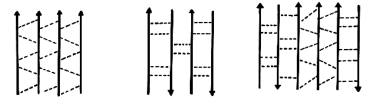
"So Long, and Thanks for All the Fish"

Side Chains

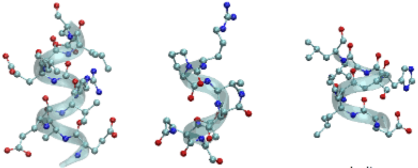

Nonpolar, aliphatic side groups
Glycine
Gly, G
Alanine
Ala, A
Valine
Val, V
Leucine
Leu, L
Methionine
Met, M
Isoleucine
Ile, I
Polar, uncharged side groups
Serine
Ser, S
Threonine
Thr, T
Cysteine
Cys, C
Proline
Pro, P
Aspartate
Asp, D
Glutamine
Gln, Q
Aromatic side groups
Phenylalanine
Phe, F
Tyrosine
Tyr, Y
Tryptophan
Trp, W
Positively charged side groups
Lysine
Lys, K
Histidine
His, H
Asparagine
Asn, N
Negatively charged side groups
Glutamate
Glu, E
Aspartate
Asp, D

Hydrogen bonds in secondary structures

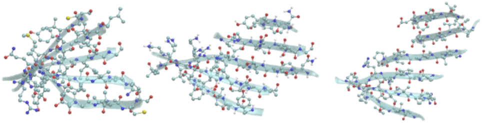




Parallel beta-sheet Antiparallel beta-sheet Mixed beta-sheet



α-helix 3_{10}-helix π-helix



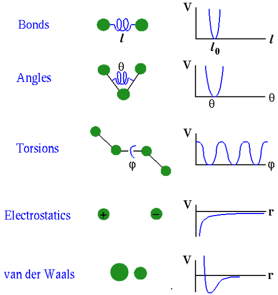
parallel anti-parallel both

Atomistic models

Degrees of freedom and interactions

"Bonded" and "non bonded" interactions.

Empiric functional form depending on many parameters (up to $\sim 10^3$) $\rightarrow$ parametrization problem.

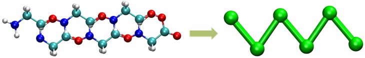


$$E = \sum_{bonds} k_b(d-d_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi(1+\cos(n\phi+\delta)) +$$

$$+ \sum_{\substack{non-bonded \\ pairs}} \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}}$$

Planar configurations

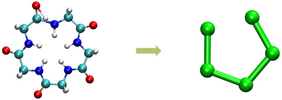
- $(\Phi, \Psi) = (0, 0)$
  $(\theta, \phi) = (75.6, 180)$



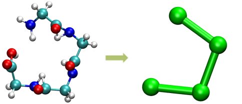
- $(\Phi, \Psi) = (180, 180)$
  $(\theta, \phi) = (105, 0)$

- $(\Phi, \Psi) = (0, 180)$
  $(\theta, \phi) = (117, 0)$

- $(\Phi, \Psi) = (180, 0)$
  $(\theta, \phi) = (146.4, 180)$



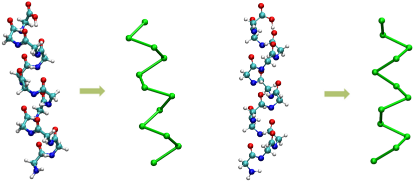
- $(\Phi, \Psi) \simeq (75, 75)$
  $(\theta, \phi) = (103, 0)$



Symmetries
Chirality and directionality

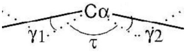
Physical symmetries:

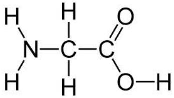
- chirality of the structure
- chain directionality



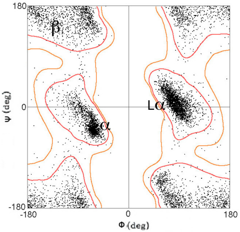
Related mathematical symmetries:

- $\theta(\Phi, \Psi) = \theta(-\Phi, -\Psi)$
  $\phi(\Phi, \Psi) = -\phi(-\Phi, -\Psi)$
- $\theta(\Phi, \Psi) = \theta(\Psi, \Phi)$
  $\phi(\Phi, \Psi) = \phi(\Psi, \Phi)$





Glicina

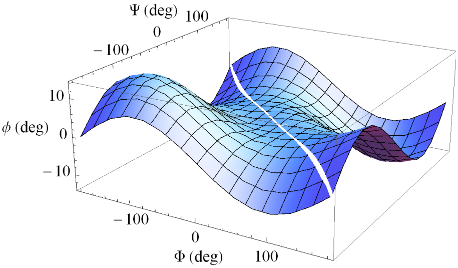


λ functions

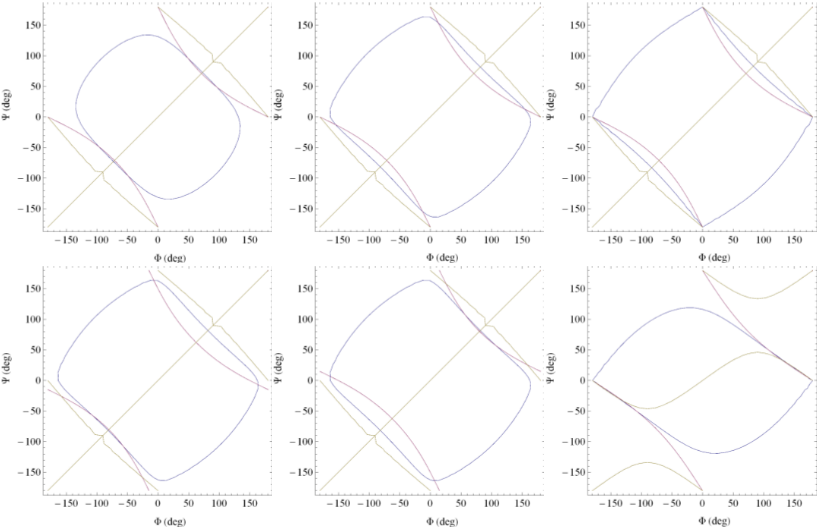
$$\begin{aligned}\tan\lambda_1(\Phi,\Psi) =&[(-\sin\tau\cos\gamma 2\sin\Phi+\cos\tau\sin\gamma 2\cos\Psi\sin\Phi-\sin\gamma 2\cos\Phi\sin\Psi)/\\&(\cos\tau\cos\gamma 2\sin\gamma 1+\sin\tau\sin\gamma 2\sin\gamma 1\cos\Psi-\\&\sin\tau\cos\gamma 2\cos\gamma 1\cos\Phi+\cos\tau\sin\gamma 2\cos\gamma 1\cos\Phi\cos\Psi+\\&\sin\gamma 2\cos\gamma 1\sin\Phi\sin\Psi)]\end{aligned}$$

$$\begin{aligned}\tan\lambda_2(\Phi,\Psi) =&[(-\sin\tau\cos\gamma 1\sin\Psi+\cos\tau\sin\gamma 1\cos\Phi\sin\Psi-\sin\gamma 1\cos\Psi\sin\Phi)/\\&(\cos\tau\cos\gamma 1\sin\gamma 2+\sin\tau\sin\gamma 1\sin\gamma 2\cos\Phi-\\&\sin\tau\cos\gamma 1\cos\gamma 2\cos\Psi+\cos\tau\sin\gamma 1\cos\gamma 2\cos\Psi\cos\Phi+\\&\sin\gamma 1\cos\gamma 2\sin\Psi\sin\Phi)]\end{aligned}$$

ϕ linear-complete differences

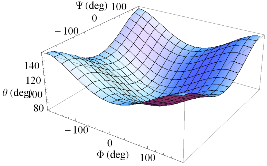


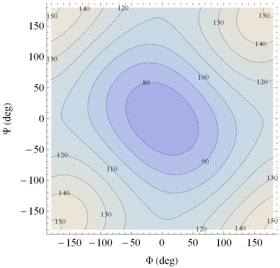
4:1 detailed analysis



3D plots, uniform case

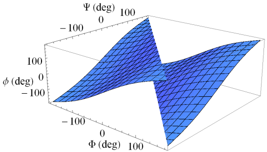
$\gamma 1 = \gamma 2$

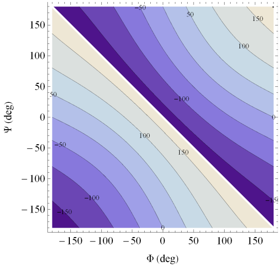




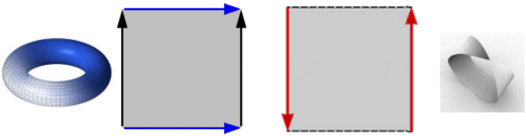
Hypotheses:

- $\Phi_i = \Phi_{i+1}$,
  $\Psi_i = \Psi_{i+1}$
- $\gamma 1 = \gamma 2 = 20°$

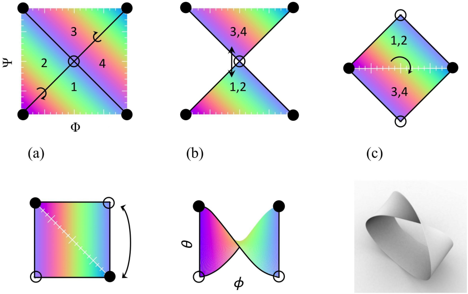




$\gamma 1 = \gamma 2$: topology of the mapping



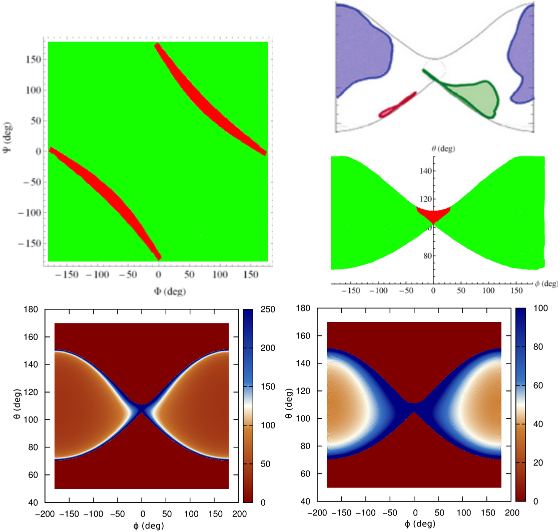
Φ and Ψ cyclic $\rightarrow$ toroidal topology of the (Φ, Ψ) plane.
Adding the symmetry $\Phi \leftrightarrow \Psi$: Möbius strip in the (θ, ϕ).



(a) (b) (c)

(d) (e) (f)

$\gamma 1 = \gamma 2$: backmapping and tranformation potential



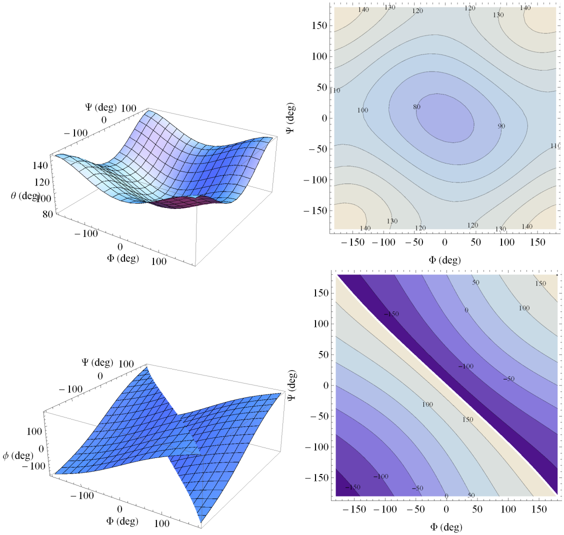
3D plot $\gamma 1 \neq \gamma 2$

- $\Phi_i = \Phi_{i+1}$,
  $\Psi_i = \Psi_{i+1}$
- $\gamma 1 = 14.7°$,
  $\gamma 2 = 20.7°$

$$\Downarrow$$

Broken symmetry

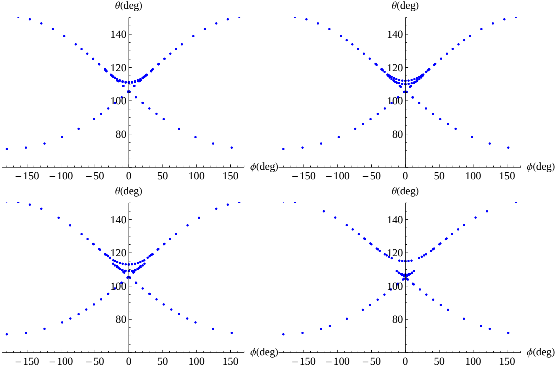
$\Phi \leftrightarrow \Psi$



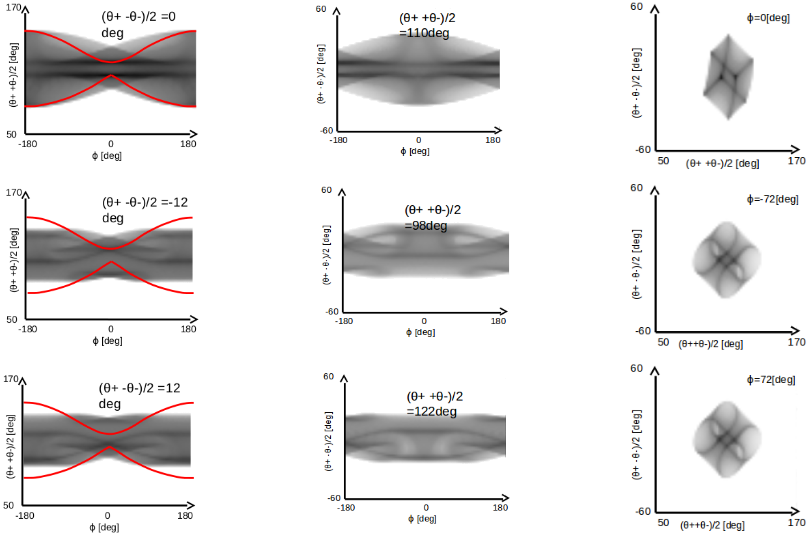
detJ zeroes in the γ parameters space

$\det J = 0$ in the (θ, ϕ) plane varying γ parameters.

Reduction of the 4:1 mapping region when $\delta\gamma = \mid \gamma 1 - \gamma 2 \mid$ grows



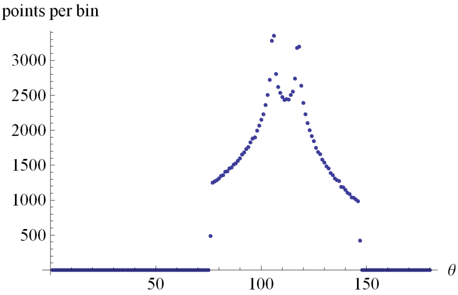
Non uniform case: two variable slices

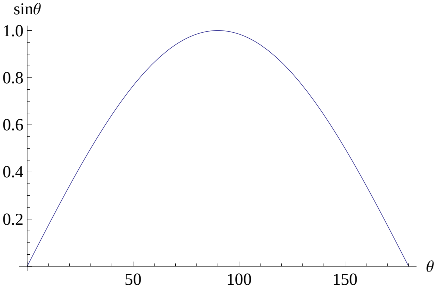


Non uniform case: single variable

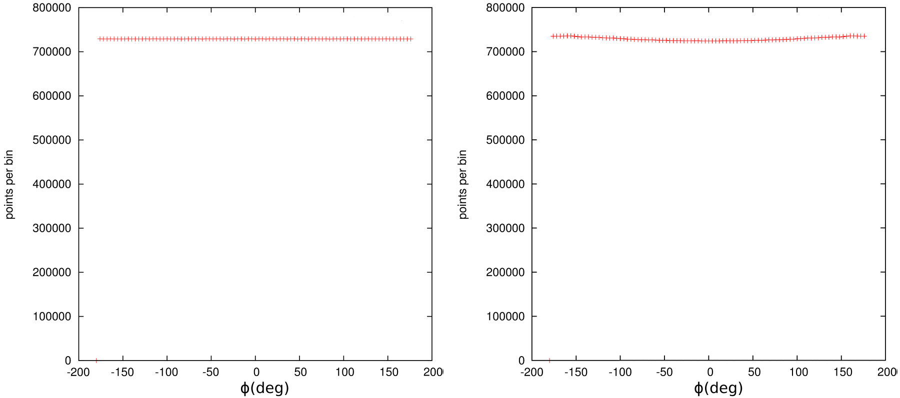
Almost uniform ϕ distribution.

Non trivial θ.





ϕ linear-complete differences
unstructured case



ϕ linear-complete differences
zoom