The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

**SMR/1845-10**

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

***4 - 8 June 2007***

## Common large-scale movements in enzymatic superfamilies

Cristian MICHELETTI
*SISSA*
*International School for Advanced Studies*
*Via Beirut 2-4*
*34014 Trieste*
*ITALY*

# Common large-scale movements in enzymatic superfamilies

Cristian Micheletti – SISSA (Trieste)

**Outline:**

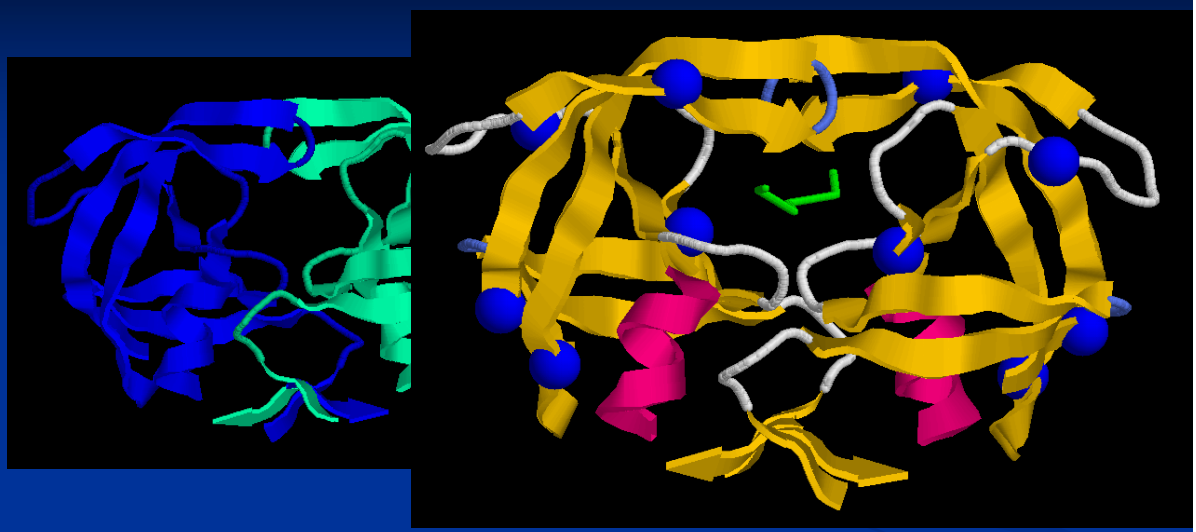
Large-scale movements in enzymes:

- coarse-grained models and all-atom MD
- long-range mechanical couplings in enzymes
- common large-scale movements in proteases and other enzymatic families

June 2007

---

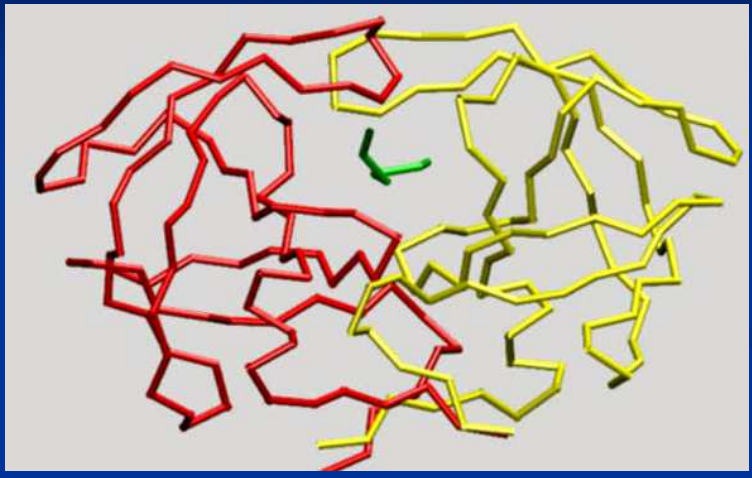
# HIV-1 Protease



Mutations causing resistance to inhibiting drugs are located at a few sites: 36, 46, 63, 71, 84.

*Elastic properties of HIV-1 PR: Jacobs et al.* Proteins *(2001); Kurt et al.,* Proteins *(2003), Piana et al.* J Mol Biol *(2002), Micheletti et al.* Proteins *(2004)*

# MD simulations

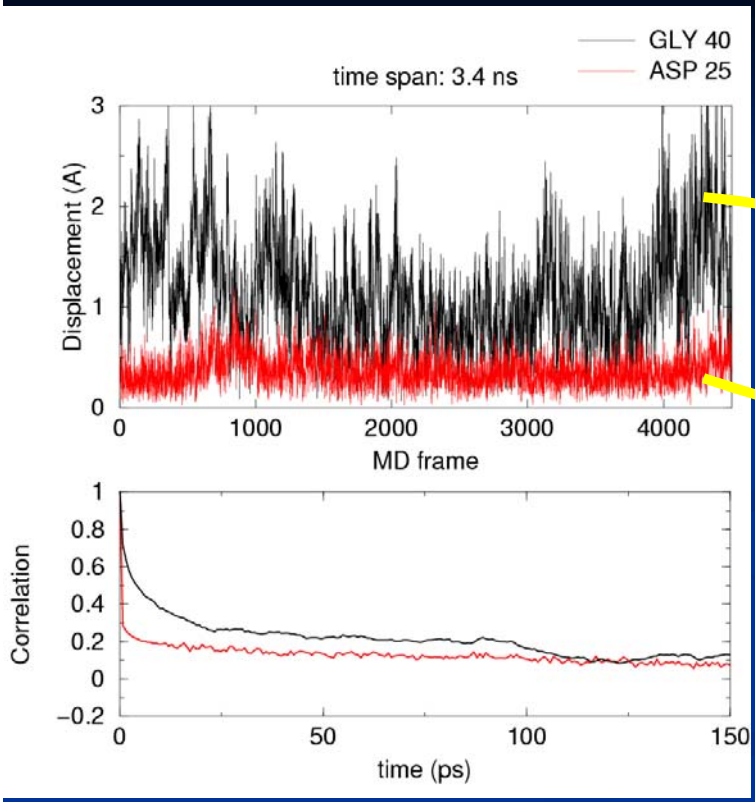


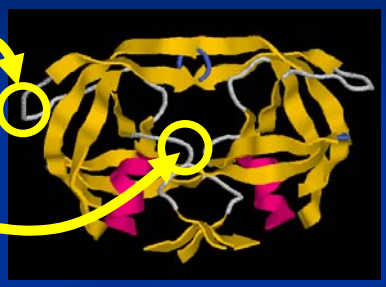
HIV-1 Protease with inhibiting peptide

Simulation at 300 K with explicit solvent

Clip time-span: 150 ps

*Piana et al., JMB, 2002*



Auto-correlation:

$$C(\tau) \propto \langle \delta\vec{x}_i(t) \cdot \delta\vec{x}_i(t+\tau) \rangle_t$$

with $\delta\vec{x}_i(t) = \vec{x}_i(t) - \langle \vec{x}_i \rangle$

# Analysis of MD trajectories

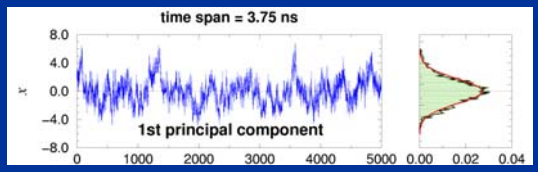
Covariance Matrix: degree of dynamical correlation of all pairs of residues

$$C_{ij} = \langle \delta\vec{x}_i(t) \cdot \delta\vec{x}_j(t) \rangle$$

*Amadei et al.*, Proteins, *1993*; *Garcia*, PRL, *1992*

The top eigenvectors of *C* identify the concerted movements that most contribute to the enzyme's fluctuations (principal components analysis).

MD trajectory projected along the first principal component:




Gaussian distribution for *P(x)* suggests that the collective motion occurs in an effective quadratic potential, *V(x)*:

$$P(x) \propto e^{-\beta V(x)} \qquad V(x) = \frac{k}{2}x^2$$

**Quasi-harmonic free energy approx.** : Pontiggia et al., *PRL* 2007

---

VOLUME 77, NUMBER 9 PHYSICAL REVIEW LETTERS 26 AUGUST 1996

**Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis**

Monique M. Tirion*
*Department of Membrane Research and Biophysics, Weizmann Institute of Science, Rehovot 76100, Israel*

Detailed potential…

$$E_p = \frac{1}{2}\sum_{\text{bonds}} K_b(b - b_0)^2 + \frac{1}{2}\sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \frac{1}{2}\sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)] + \sum_{\text{non bonded pairs}}\left[\frac{A}{r^{12}} - \frac{B}{r^6} + \frac{q_1 q_2}{Dr}\right]. \quad (1)$$

replaced by a quadratic one:

$$E(\mathbf{r}_a, \mathbf{r}_b) = \frac{C}{2}(|\mathbf{r}_{a,b}| - |\mathbf{r}^0_{a,b}|)^2$$

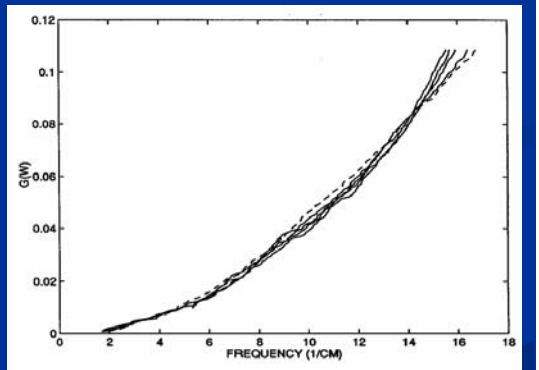
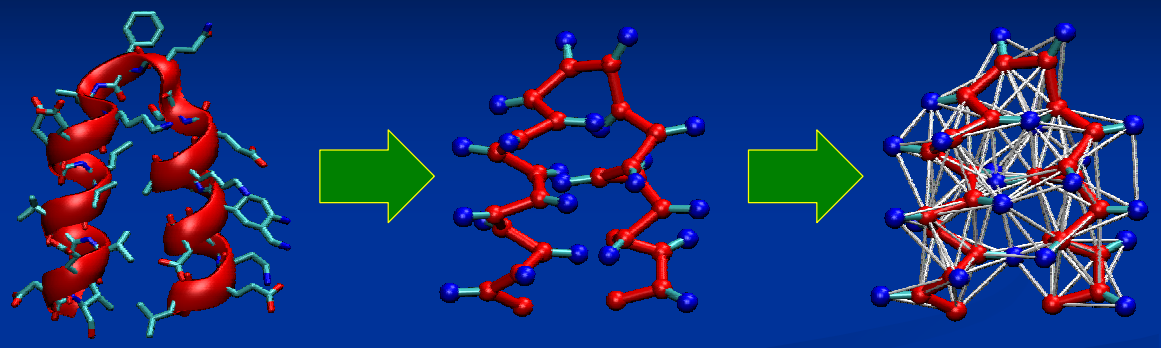



FIG. 1. The fraction of the total number of modes up to frequency $\omega$ ($cm^{-1}$) for the slowest 150 modes of the G-actin:ADP:$Ca^{++}$ system. The dashed line pertains to data obtained using the L79 potential, while the four solid curves are obtained using $R_C$ values of 1.1, 1.5, 2.0, and 2.5 Å. The $R_c$ = 1.1 Å curve is nearest the dashed line at higher frequencies, with the fit progressively worsening for the higher cutoff values.

# Structure-based approach



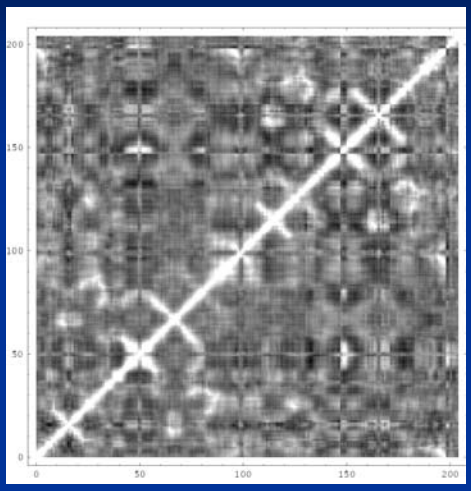
**Gaussian network approximation:**
Penalize quadratically deviations of centroids from native pair distances.

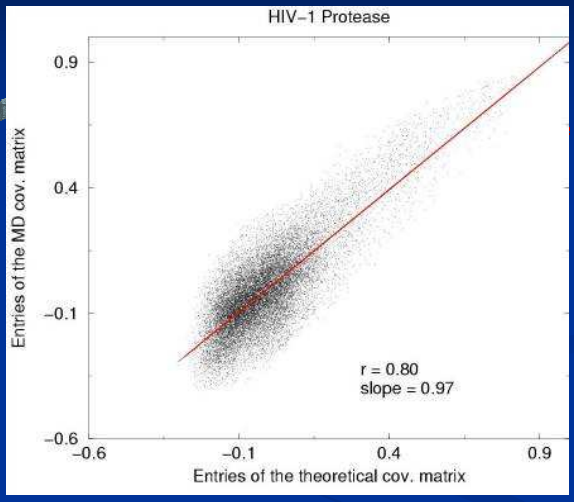
$$\begin{aligned}\mathcal{H} &= K\sum_i \delta\vec{r}_{i,i+1}\, M_1\, \delta\vec{r}_{i,i+1} + \sum_{i,j}\Delta_{i,j}\delta\vec{r}_{i,j}\, M_2\, \delta\vec{r}_{i,j} \\ &= \frac{1}{2}\sum_{i,j}\delta\vec{r}_i\, M_{ij}\, \delta\vec{r}_j\end{aligned}$$

*See e.g. Atilgan et al.* Biophys. J. *(2001); Delarue and Sanejouand,* JMB *(2002); CM, Banavar and Maritan.* Phys. Rev. Lett. *(2001).*
*CM, Carloni and Maritan* Proteins (2004) - BetaGM code available upon request

Thermodynamic characterization by inversion of M: $\langle \delta r_i \delta r_j \rangle = M_{ij}^{-1}$

# HIV-1 PR Covariance Matrix





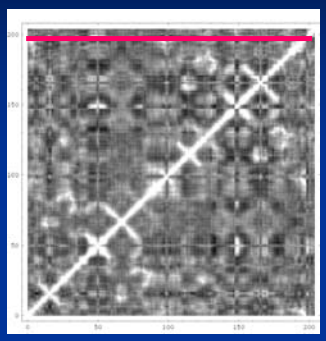

**Normalised covariance matrix:**

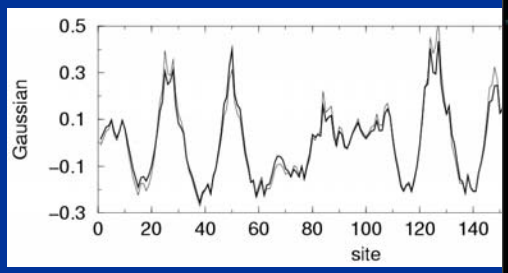
$$\begin{aligned}C_{ij} &= \langle \delta\vec{x}_i \cdot \delta\vec{x}_j \rangle / |\delta\vec{x}_i|\,|\delta\vec{x}_j| \\ &= M_{ij}^{-1} / \sqrt{M_{ii}^{-1}\, M_{jj}^{-1}}\end{aligned}$$

See poster by F. Pontiggia

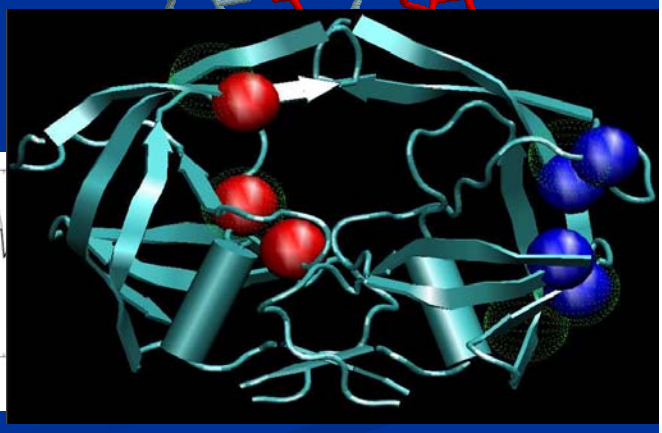
*Pontiggia, Orland, CM. Phys. Rev. Lett. 2007*
*CM, Carloni and Maritan, Proteins 2004*

# Protease – substrate coupling



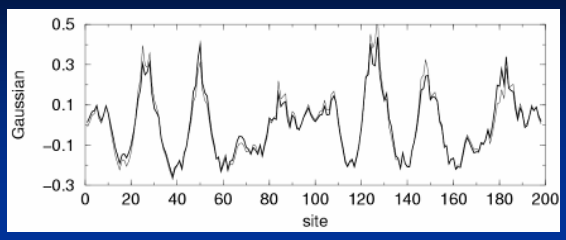


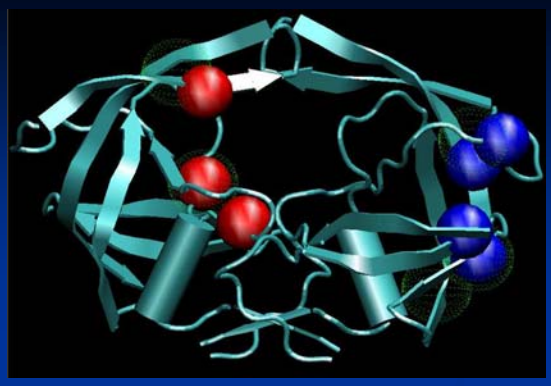

Positive correlation: 25, 47, 84
Negative correlation: 15, 37, 59, 63

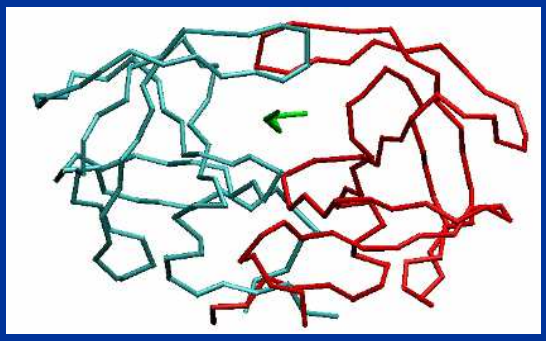


Mutations: 36, 46, 63, 71, 84

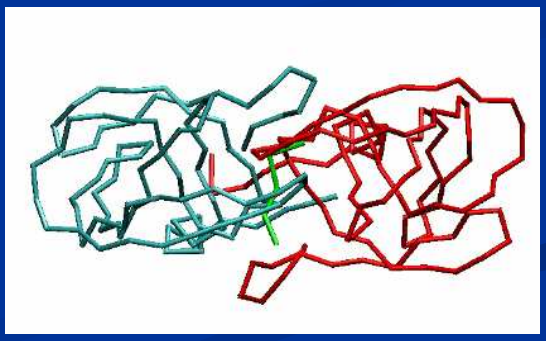
---





Front view



Top view

# Proteases

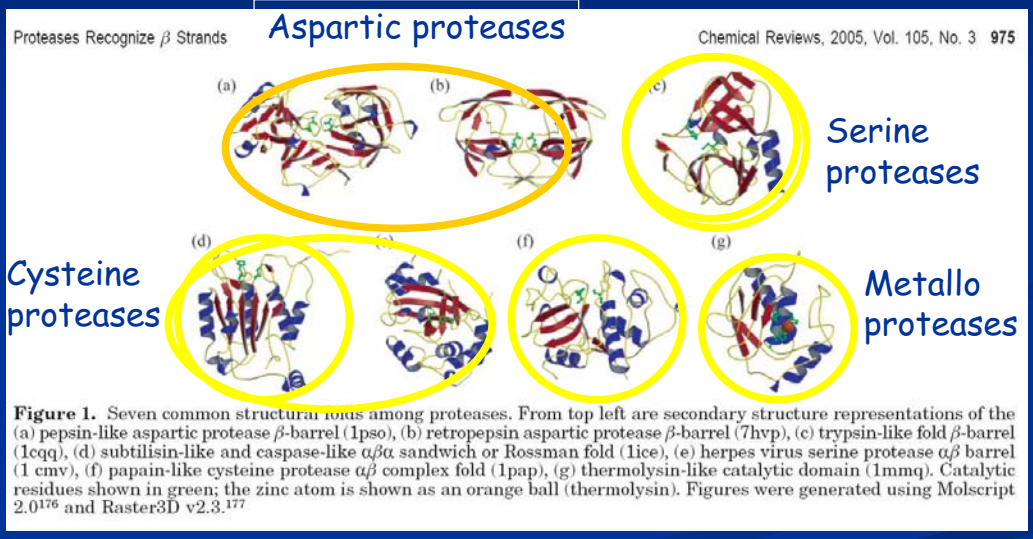
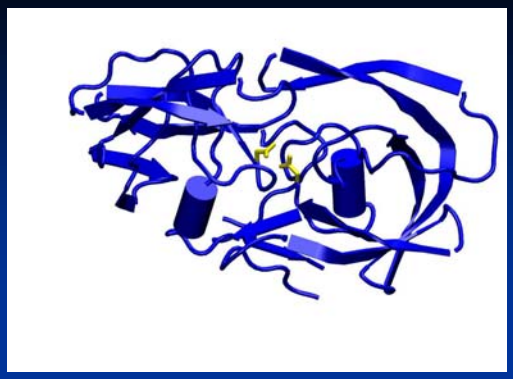



**Figure 1.** Seven common structural folds among proteases. From top left are secondary structure representations of the (a) pepsin-like aspartic protease β-barrel (1pso), (b) retropepsin aspartic protease β-barrel (7hvp), (c) trypsin-like fold β-barrel (1cqq), (d) subtilisin-like and caspase-like αβα sandwich or Rossman fold (1ice), (e) herpes virus serine protease αβ barrel (1 cmv), (f) papain-like cysteine protease αβ complex fold (1pap), (g) thermolysin-like catalytic domain (1mmq). Catalytic residues shown in green; the zinc atom is shown as an orange ball (thermolysin). Figures were generated using Molscript 2.0[176] and Raster3D v2.3.[177]

---

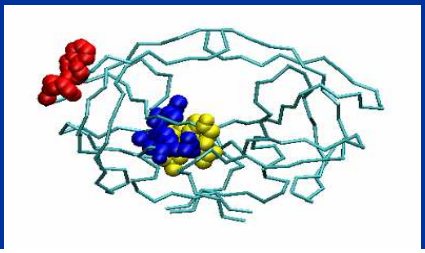


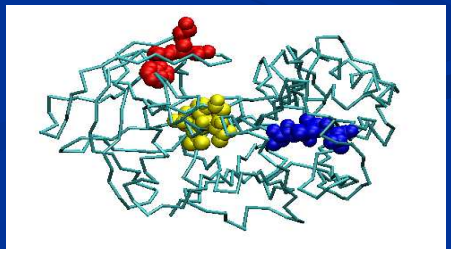
Beta-secretase and HIV-1 protease:

-Conserved residues in the loops containing ASP dyad

-Similar large scale movements in "visually-matching" regions

Rawlings et. al. Methods Enzymol. (1995), Blundell et al. PNAS (1996), Cascella et. al. JACS (2005)

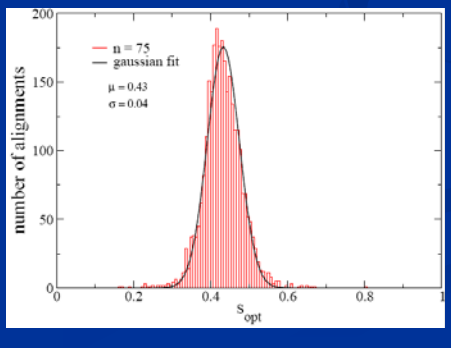




# Different structure similar motion?

- Pick residues in putative correspondence
- calculate their effective large scale movements
- calculate the hybrid structural/dynamical score

Use a stochastic optimization technique to identify the alignment providing the best score (and measure its statistical significance).

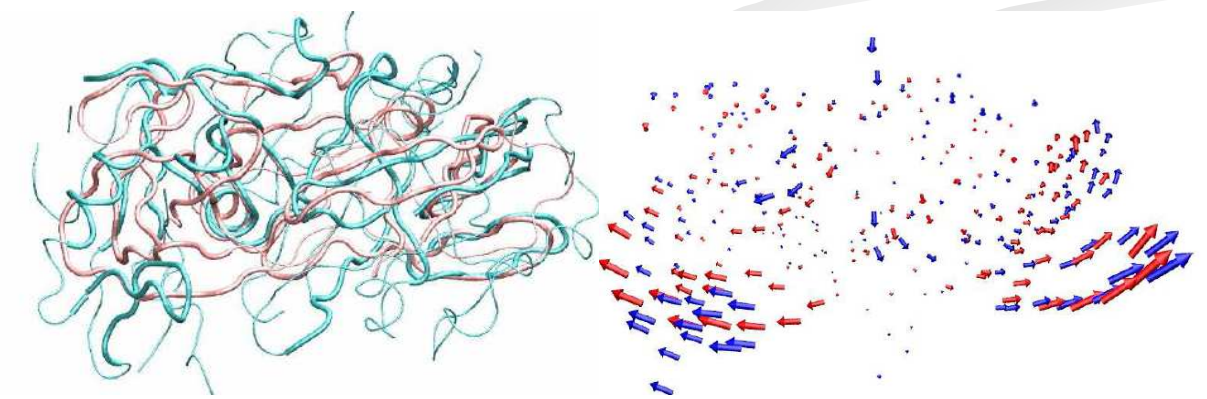



---

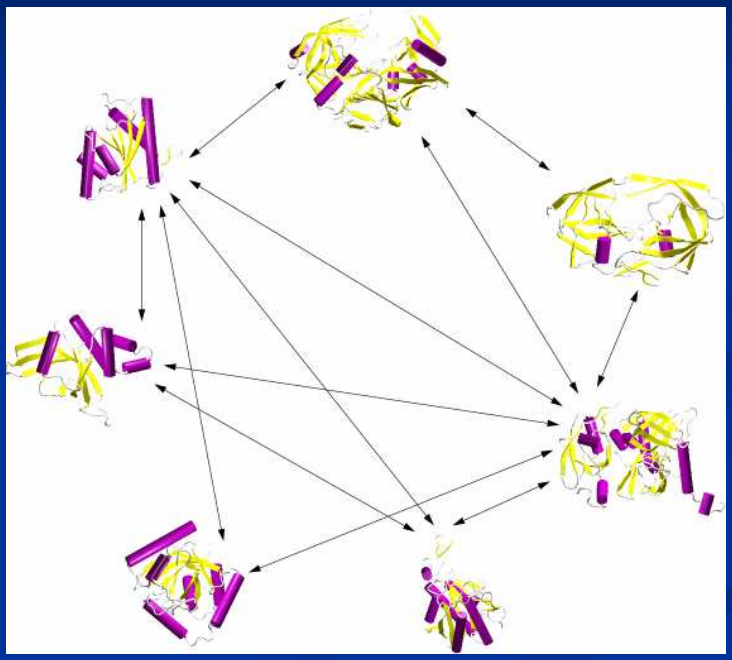
# BACE - HIV dynamical alignment

BACE (pdb code: 1er8) - 330 residues
HIV-1 PR (pdb code: 1nh0 ) - 198 residues

150 aligned residues
RMSD = 5.5 A ; RMSIP = 0.72



# Proteases with similar near-native large-scale movements



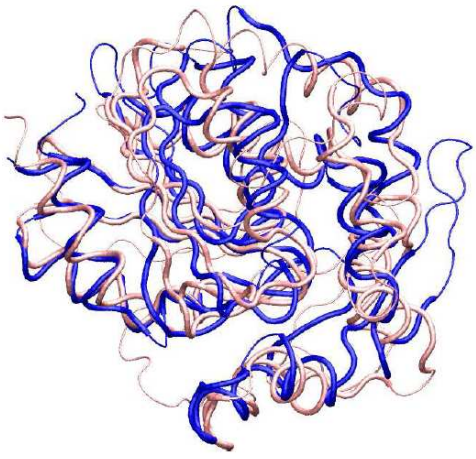
*Carnevale et al. JACS 2006*

---

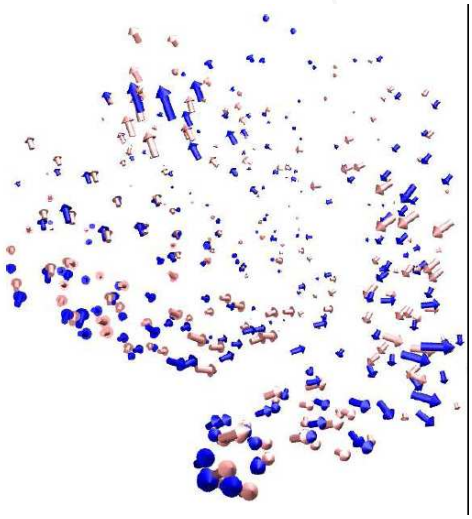
# Further examples

Hydroxynitrile lyase (PDB: 1yb7, length 256, EC: 4.1.2.39, CATH: 3.40.50.1820)
Haloalkane dehalogenase (PDB: 2had, length 310, EC: 3.8.1.5, CATH: 3.40.50.1820)

200 aligned residues, RMSIP = 0.81 and RMSD = 4.5 A



a1

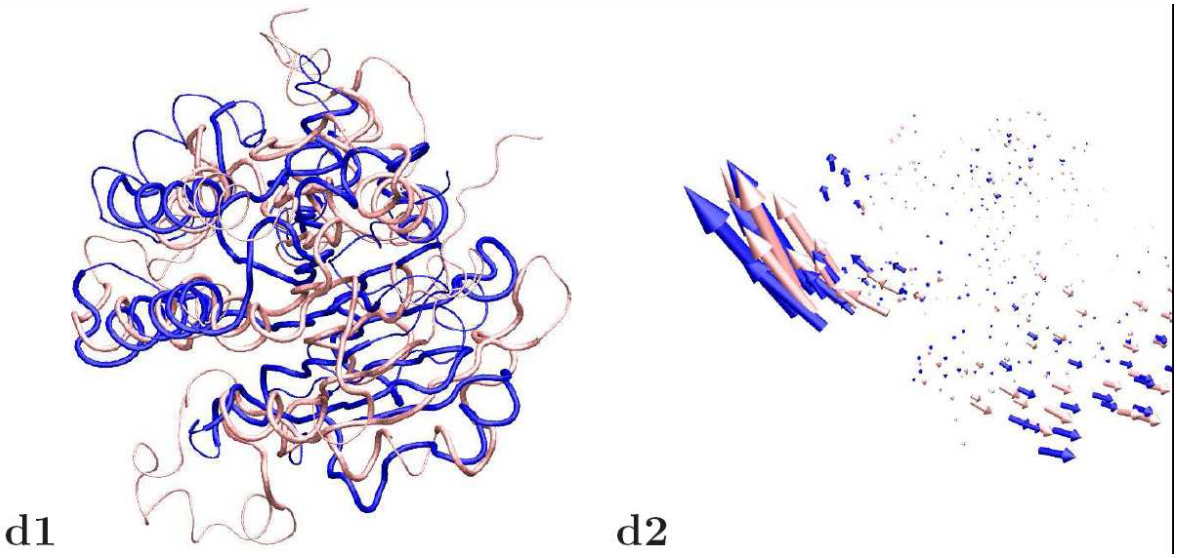


a2

# Further examples

exonuclease III (PDB: 1ako, length 268, EC: 3.1.11.2, CATH: 3.60.10.10)
enoyl reductase (PDB: 1d7o, length 297, EC: 1.3.1.9, CATH: 3.40.50.720)

175 aligned residues, RMSIP = 0.70, RMSD = 8.4




# Acknowledgements


- Paolo Carloni (Trieste)
- Enzo Carnevale (Trieste)
- Michele Cascella (Lausanne)
- Giorgio Colombo (Trieste)
- Arthur Lesk (Penn State)
- Henri Orland (Paris)
- Francesco Pontiggia (Trieste)
- Andrea Zen (Trieste)
- Ivet Bahar
- Martino Bolognesi
- Burak Erman
- Alessandro Laio
- Doriano Lamba
- Ed Lattmann
- Amos Maritan
- Henriette Molinari
- Stefano Piana