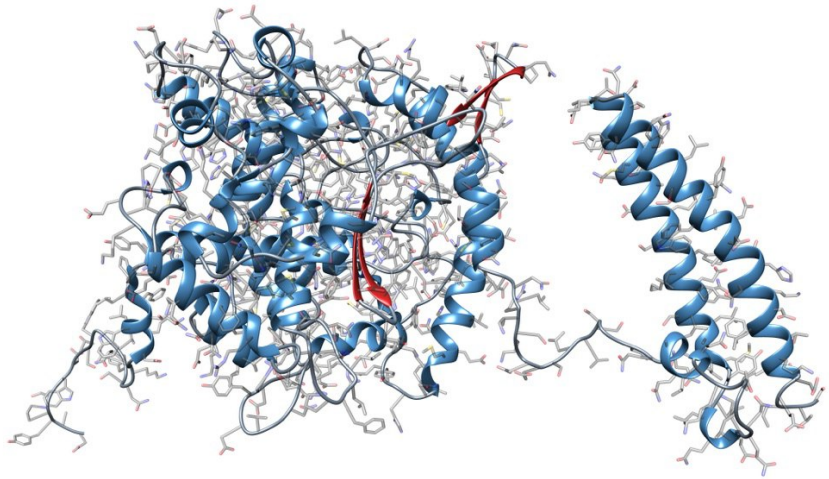
The Science and T.T. behind Sibylla Biotech SRL

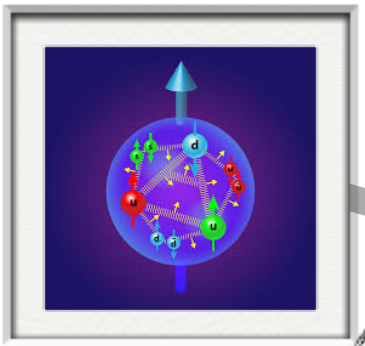
Pietro Faccioli

UNIVERSITÀ DEGLI STUDI
DI TRENTO
Dipartimento di Fisica

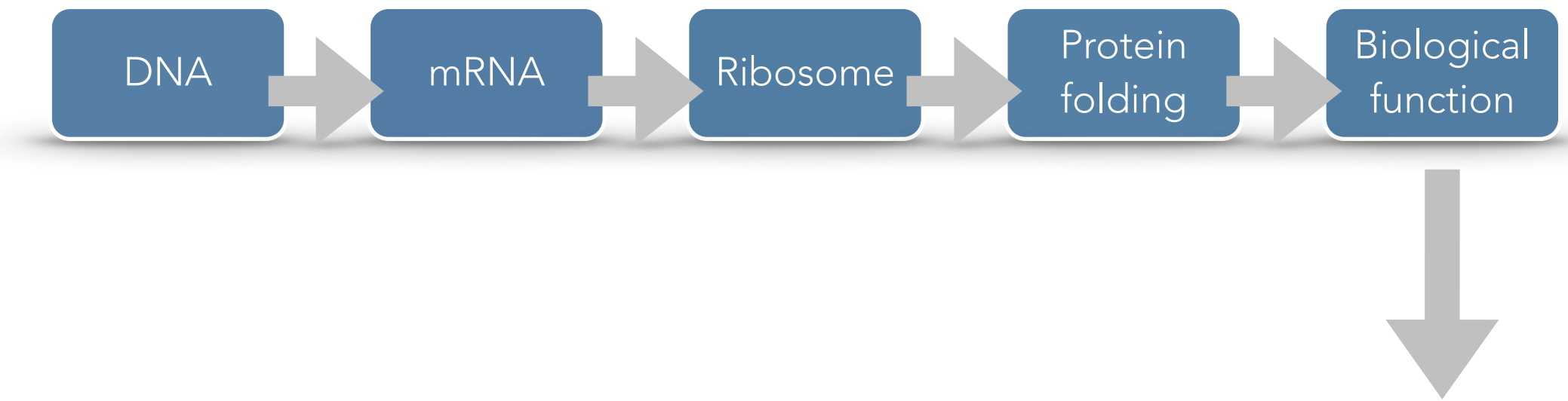
TIFPA
Trento Institute for
Fundamental Physics
and Applications



A SCIENTIFIC JOURNEY

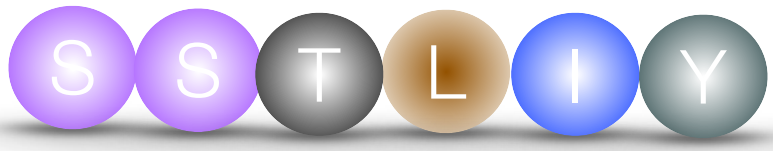


FUNDAMENTAL DOGMA OF MOLECULAR BIOLOGY

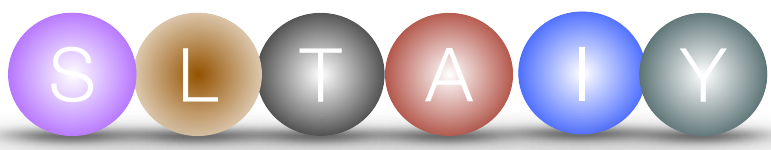


…proteins are Life's nanomachines!

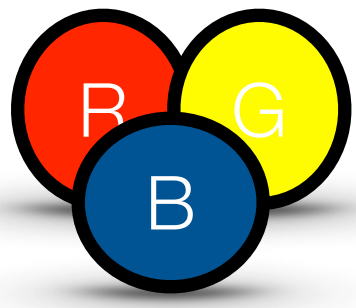
PROTEINS AND HADRONS ARE VERY SPECIAL PHYSICAL SYSTEMS



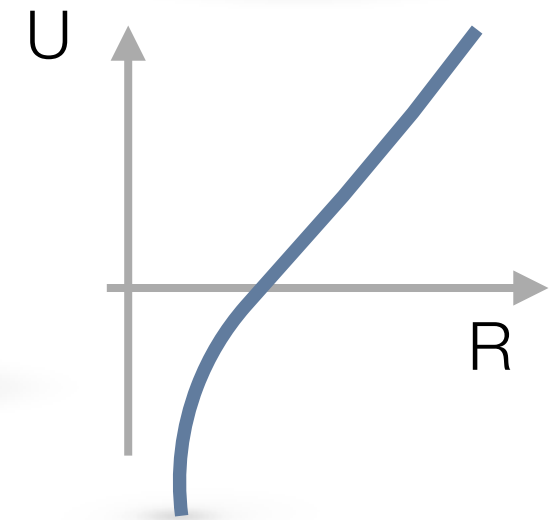
Random polypeptide



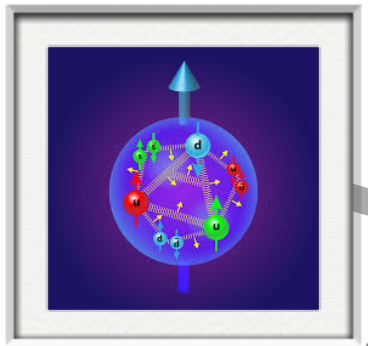
Protein

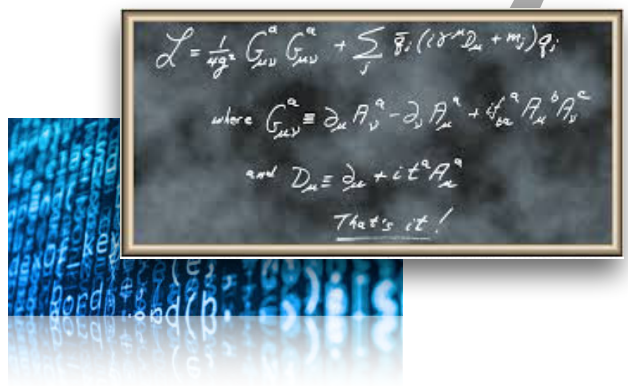


Baryon

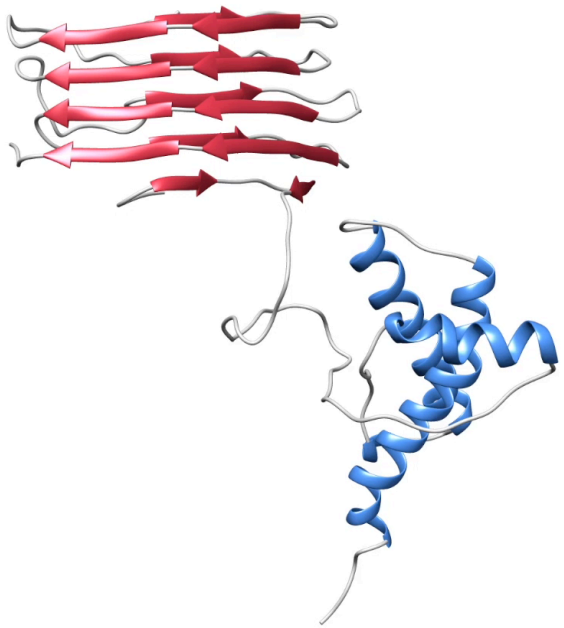


PHASE 1: MATHEMATICAL FORMALISM & HIGH PERFORMANCE COMPUTING





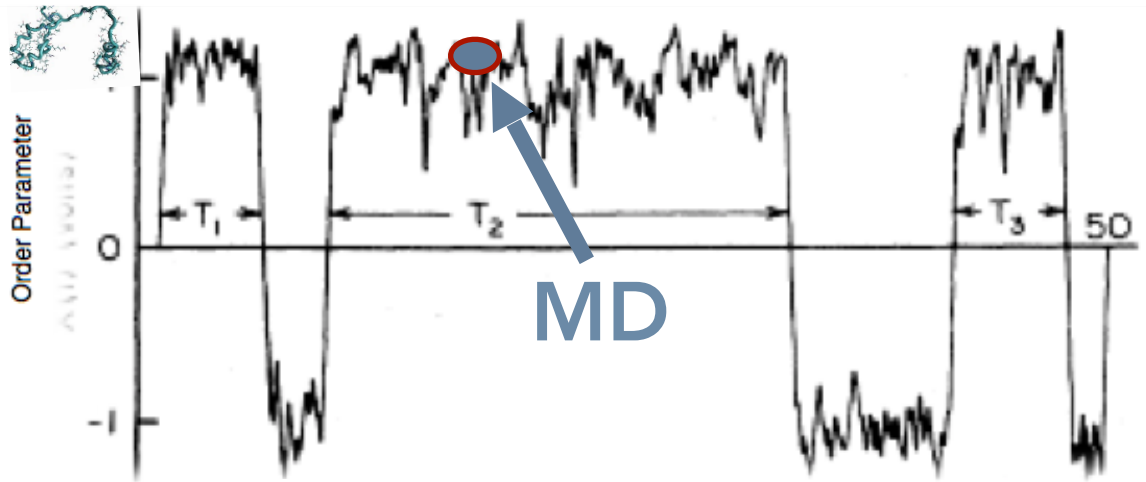
REDUCTIONIST'S APPROACH TO MOLECULAR BIOLOGY

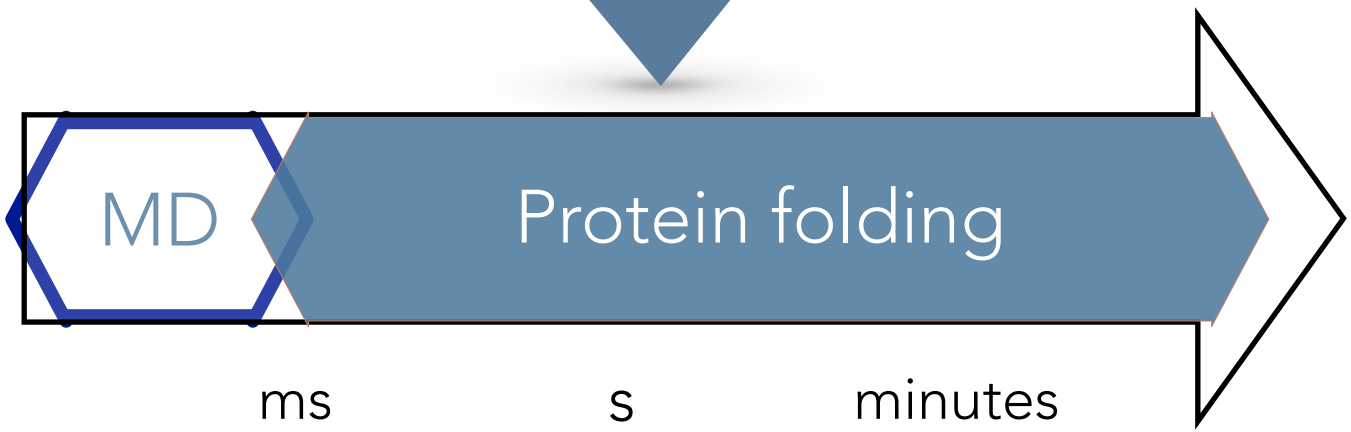


Challenge:

Integrate ~10^6 coupled Newton-type equations looking for **extremely rare events**

PROTEIN DYNAMICS IS FULL OF **RARE EVENT PROBLEMS**

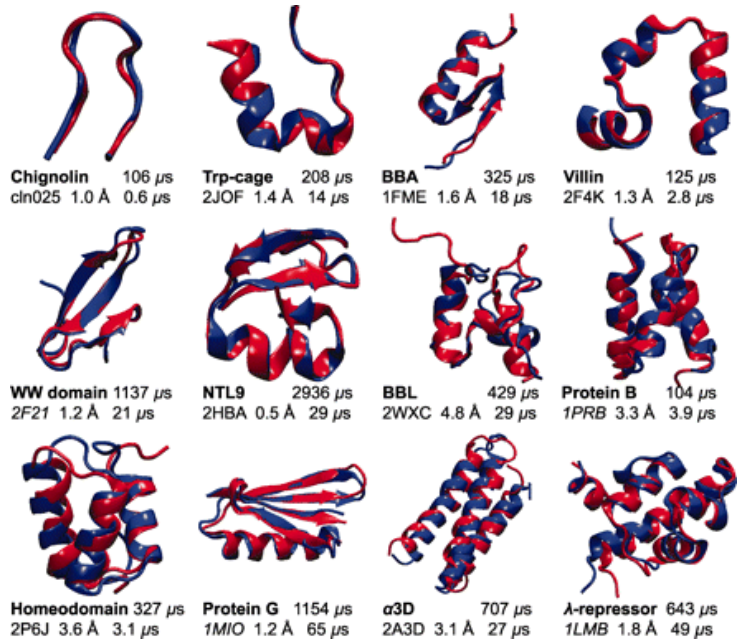




MD YIELDS CORRECT PROTEIN NATIVE STATES

Anton supercomputer
(DES Research)

MD



Science
AAAS

Atomic-Level Characterization of the Structural Dynamics of Proteins
David E. Shaw, *et al.*
Science **330**, 341 (2010);
DOI: 10.1126/science.1187409

How Fast-Folding Proteins Fold

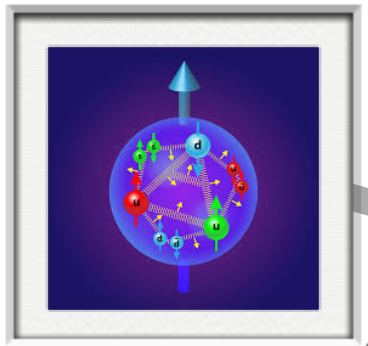
Kresten Lindorff-Larsen,[1*†] Stefano Piana,[1*†] Ron O. Dror,[1] David E. Shaw[1,2†]

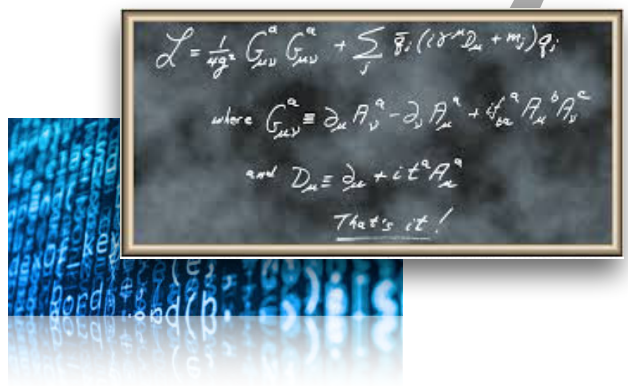
ZOOLOGY OF ENHANCED SAMPLING METHODS

Markov State Models, Milestoning, Transition Path Sampling, Transition Interface Sampling, Forward Flux Sampling, Temperature Accelerated Molecular Dynamics, Metadynamics, Umbrella Sampling, Blue Moon Sampling, String Method,Stochastic Difference, … [and counting]

They are **all too computationally demanding** for many biologically relevant problems.

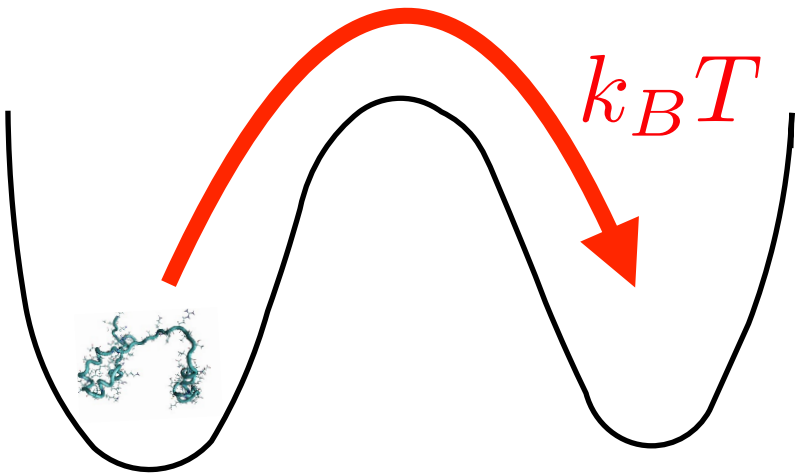
PHASE 1: MATHEMATICAL FORMALISM & HIGH PERFORMANCE COMPUTING





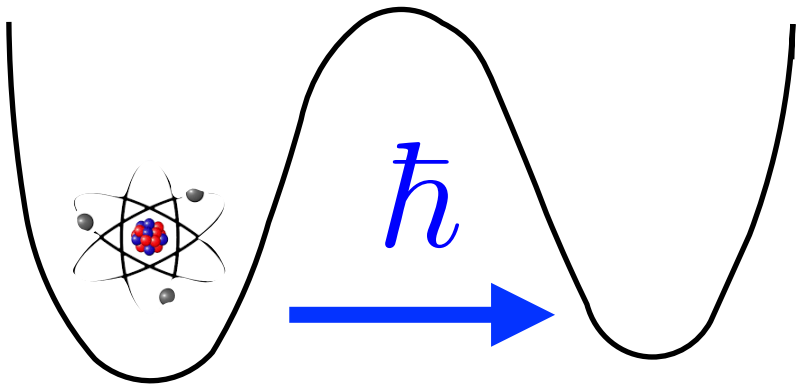
A USEFUL ANALOGY

Thermal activation



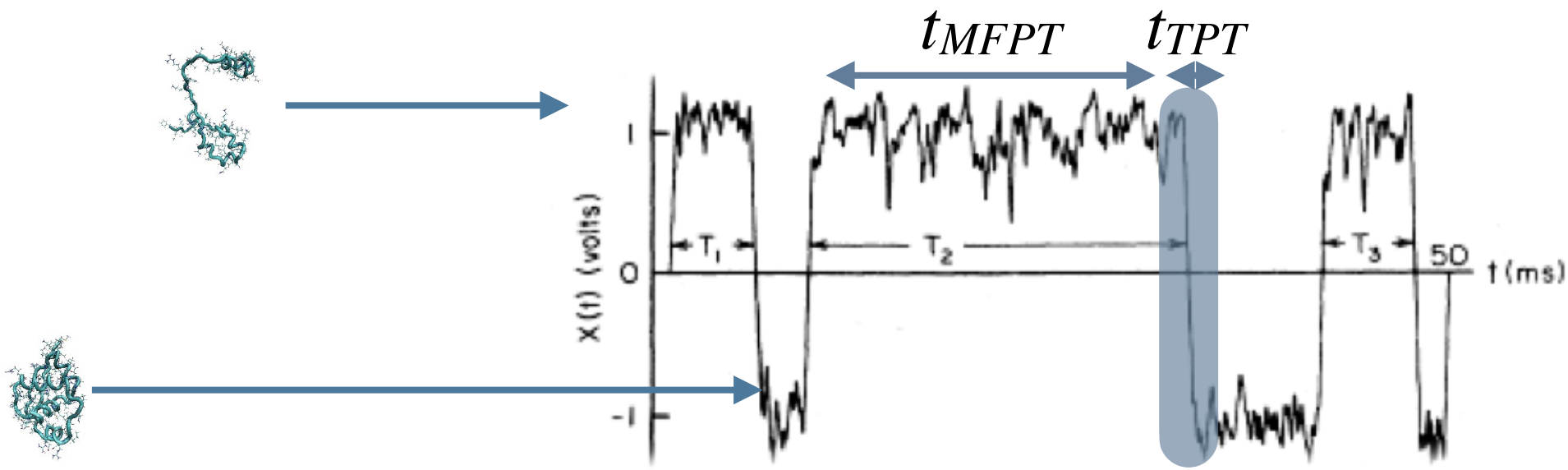
$$P(x_f, t|x_i) = \int_{x_i}^{x_f} \mathcal{D}q \; e^{-\frac{\beta}{4M\gamma} \int_0^t d\tau (M\ddot{q} + M\gamma\dot{q} + \nabla U(q))^2}$$

Quantum tunneling



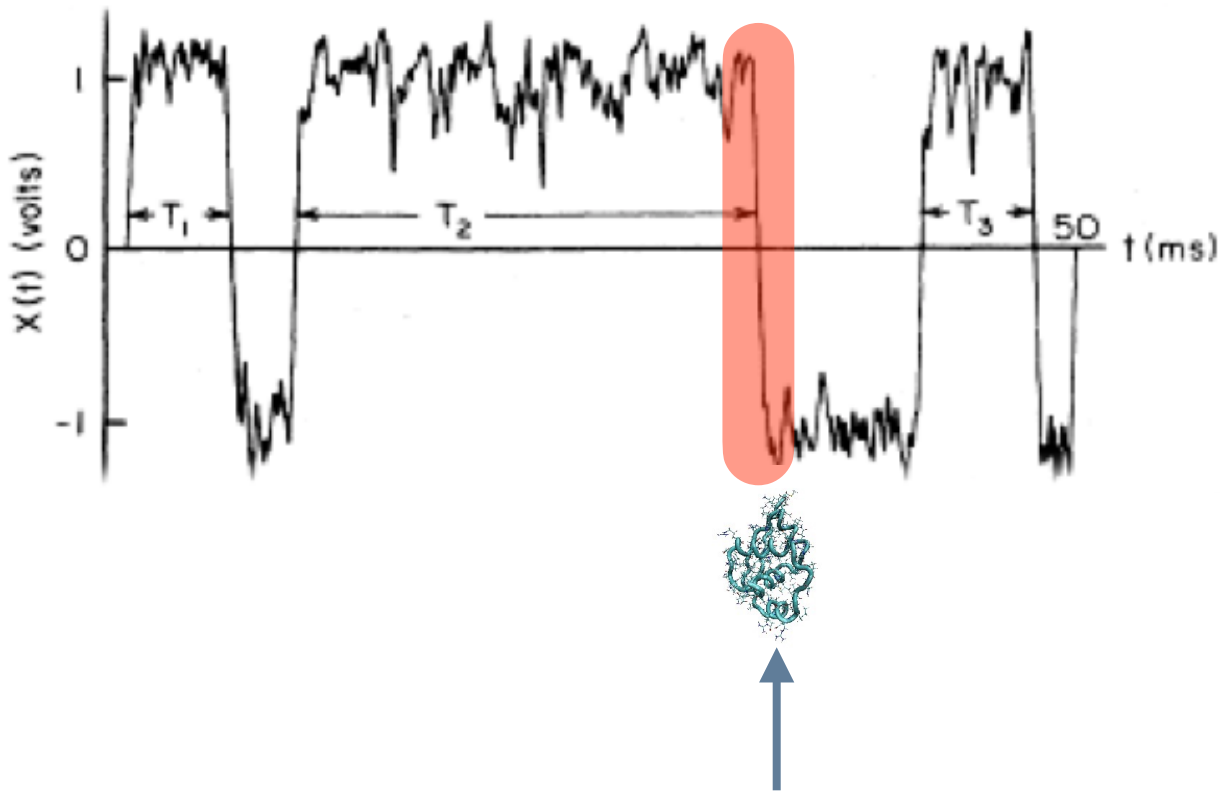
$$K_E(x_f, t|x_i) = \int_{x_i}^{x_f} \mathcal{D}q \; e^{-\frac{1}{\hbar} \int_0^t d\tau \left(\frac{M}{2}\dot{q}^2 + U(q) \right)}$$

ADVANTAGES OF PATH INTEGRALS



$$t_{TPT} \sim \tau_0 \log\left[\log\left(\frac{t_{MFPT}}{\tau_0}\right)\right]$$

IS THIS A "FREE LUNCH"?



All atom 3D structure of the native state **are given in input,** not predicted

FULLY EXPLOITING THEORETICAL PHYSICS TOOLS

between the Gibbs distribution and the SCR estimate forward- and backward-committors, as in Eq. (A3). Introducing the distribution

$$P^{(P)}(x,t) \equiv \int dx_i\, P^{(P)}(x,t|x_i,0)\, \rho_0(x_i), \qquad (23)$$

the density in Eq. (22) reads

$$m_{SCR}(x) = \frac{1}{t_f - \tau_0} \int_{\tau_0}^{t_f} dt\; Q^{(R)}(x, t_f - t) P^{(P)}(x,t).$$

Using the detailed balance condition, we find $P^{(P)} = e^{-\beta U(x)} \frac{1}{Z_R} Q^{(P)}(x,t)$. Then, inserting this result into Eq. we find

$$m_{SCR}(x) = \frac{e^{-\beta U(x)}}{Z_R\,(t_f - \tau_0)} \int_{\tau_0}^{t_f} dt Q^{(R)}(x, t_f - t) Q^{(P)}(x,t).$$

Finally, recalling that $Q^{(R)}(x,t)$ and $Q^{(P)}(x,t)$ are n time-independent in the SCR and using Eqs. (17) and we recover a fundamental result of TPT [cf. Eq. (A Appendix A],

$$m_{SCR}(x) \propto e^{-\beta U(x)}\, q^{+}_{SCR}(x)\,(1 - q^{+}_{SCR}(x)\,).$$

Within the same framework, it is possible to d the reactive current in the SCR in complete analogy Eq. (22),

$$J^{i}_{SCR}(x) = \frac{-D}{t_f - \tau_0} \int_{\tau_0}^{t_f} dt Q^{(R)}(x, t_f - t) \times (\overrightarrow{\nabla} - \overleftarrow{\nabla} + \beta \nabla U(x))\, P^{(P)}(x,t).$$

$$\mathcal{P}_{\rm bias}[X] = \int \mathcal{D}Y e^{-S_{\rm bias}[X,Y] - U(X_i,Y_i)/k_BT}, \qquad (3)$$

where the functional $S_{\rm bias}[X,Y]$ is defined as

$$S_{\rm bias} \equiv \frac{1}{4k_BT} \int_0^t d\tau \Big[\sum_{i=1}^{N} \frac{1}{\gamma_i m_i} (m_i \ddot{\mathbf{x}}_i + m_i\gamma_i \dot{\mathbf{x}}_i + \nabla_i U - \mathbf{F}_i^{\rm bias})^2 + \sum_{j=1}^{N'} \frac{1}{\gamma_j m_j} (m_j \ddot{\mathbf{y}}_i + m_j\gamma_j \dot{\mathbf{y}}_j + \nabla_j U)^2 \Big]. \qquad (4)$$

The Onsager-Machlup functional $S_{\rm OM}[X,Y]$ entering Eq. (2) is recovered, setting $\mathbf{F}_i^{\rm bias} = 0$ in Eq. (4).

Let us now return to the problem of computing the reaction pathways in the *unbiased* Langevin dynamics [Eq. (1)]. Using the standard reweighting trick we can write the variational condition $(\delta/\delta X)\mathcal{P}[X] = 0$ as

$$\frac{\delta}{\delta X} [\mathcal{P}_{\rm bias}[X] \langle e^{-(S_{\rm OM}[X,Y] - S_{\rm bias}[X,Y;t])} \rangle_{\rm bias}] = 0. \qquad (5)$$

We now introduce our main approximation, by restricting the search for the optimum path $X(\tau)$ within an ensemble of trajectories generated by integrating the *biased* Langevin equation. By definition, these paths have a large statistical weight in the biased dynamics; i.e., they lie in the functional vicinity of some path $\bar{X}(\tau)$ which satisfies $(\delta/\delta\bar{X})\mathcal{P}[\bar{X}] = 0$. Thus, the typical biased paths approx-

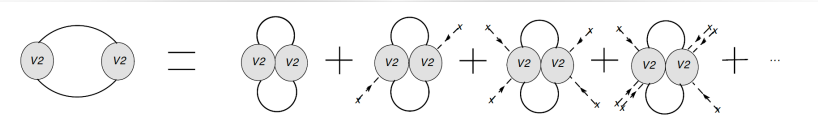


FIG. 3: Diagrammatic representation of the local time-derivative expansion of a non-local diagram —Eq. (49)—. Solid lines are fast-mode propagators, while dashed lines represent a single time derivative acting on the corresponding vertex function.

Notice that each term in the perturbative expansion (35) generates a new vertex, with an increasing power of the $x_>(\tau)$ field. The couplings to the fast modes depend implicitly on the time τ, through the slow modes $x_<(\tau)$.

By Wick theorem, each term in the series (34) can be related to a Feynman graph with vertexes given by (36) and propagators given by —see appendix A —:

$$\langle x^i_>(\tau_1)\; x^j_>(\tau_2)\rangle_0 = \sum_{|\omega_m|,|\omega_n| \in S_b} G^{0\;ij}_>(\omega_n,\omega_m)\; e^{i(\omega_m\tau_1 + \omega_n\tau_2)} = \sum_{|\omega_n| \in S_b} \delta_{ij}\, \frac{2}{\beta\;\gamma\;t\;\omega_n^2}\; e^{i\omega_m(\tau_2 - \tau_1)}. \qquad (37)$$

The expansion (34) can be re-organized as the exponent of the sum performed over only connected diagrams:

$$e^{-\beta\; S_>[x<(\tau)]} = e^{\sum \text{(all connected diagrams)}}. \qquad (38)$$

Hence, the path integral (26) for the slow modes can be given the following exact diagrammatic representation

$$Z(t) \equiv \oint \mathcal{D}x_<\; e^{-\beta\; S_{eff}[x_<(t)]] + \sum \text{(all connected diagrams)}}. \qquad (39)$$

Below we give a classification of all the connected diagrams that may give a contribution to the expansion above

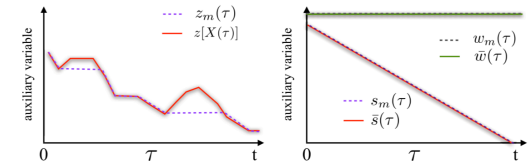


FIG. 1. Illustrative representation of the dynamics of the auxiliary variables introduced in the path integral representation of rMD (left panel) and in the derivation self-consistent path sampling algorithm (right panel).

of such a variable is frozen any time z_m becomes smaller than $z(X)$ and any time the collective coordinate $z(X)$ is increasing. Its time derivative is otherwise set equal to $\dot{z}(X)$. Therefore, by choosing the initial conditions $z_m(0) = z(X(0))$, $z_m(\tau)$ is identically set equal to the minimum value attained by the collective coordinate z until time τ (see left panel of Fig. 1).

The functional $S_{rMD}[X, z_m]$ in the exponent of Eq. (8) coincides with an OM action with the addition of the unphysical biasing force $\mathbf{F}_i$,

$$S_{rMD} = \sum_{i=1}^{N} \Gamma_i \int_0^t d\tau \,[m_i\ddot{\mathbf{x}}_i + m_i\gamma_i\dot{\mathbf{x}}_i + \nabla_i U - \mathbf{F}_i]^2. \qquad (9)$$

In Eq. (8), $\Phi[z_m, X]$ denotes a Jacobian factor that needs to be introduced in order to ensure that the statistical weight of the paths is not affected by the measure of the $\int \mathcal{D}z_m$ integral, i.e.,

$$\int \mathcal{D}z_m \Phi[z_m, X]\; \delta\left[z_m(\tau) - \int_0^\tau d\tau' \dot{z}[X(\tau')]\theta(-\dot{z}[X(\tau')]\right)$$

III. SELF-CONSISTENT PATH SAMPLING

Let us now introduce our new algorithm, which provides major improvement with respect to the rMD and BF schemes discussed in Sec. II A. Indeed, it follows directly from the unbiased Langevin equation and allows us to remove the systematic errors associated to the choice of biasing coordinate.

Our starting point is path integral representation of the *unbiased* Langevin dynamics (2). We introduce two dumb auxiliary variables $w_m(\tau)$ and $s_m(\tau)$ into this path integral by means of appropriate functional Dirac deltas,

$$p(X_N, t|X_U) = \int_{X_U}^{X_N} \mathcal{D}X \cdot e^{-S[X]} \int_{\bar{s}(0)} \mathcal{D}s_m \int_{\bar{w}(0)} \mathcal{D}w_m \cdot \delta\left[w_m(\tau) - \int_0^\tau d\tau' \dot{\bar{w}}(\tau')\,\theta(-\dot{\bar{w}}(\tau'))\,\theta(w_m(\tau') - \bar{w}(\tau'))\right] \cdot \delta\left[s_m(\tau) - \int_0^\tau d\tau' \dot{\bar{s}}(\tau')\,\theta(-\dot{\bar{s}}(\tau'))\,\theta(s_m(\tau') - \bar{s}(\tau'))\right], \qquad (12)$$

where $\bar{s}(\tau)$ and $\bar{w}(\tau)$ are two external time-dependent functions to be defined below. In analogy with the path integral repre-

$$V^R_{eff}(\mathbf{X}) \simeq \frac{D_0\,(1-b)}{\pi\, b\Omega} \nabla^2 V_{eff}(\mathbf{X}) + \frac{1}{2}\left(\frac{D_0(1-b)}{\pi\, b\Omega}\right)^2 \nabla^4 V_{eff}(\mathbf{X}) + \frac{1}{6}\left(\frac{D_0\,(1-b)}{\pi b\Omega}\right)^3 \nabla^6 V_{eff}(\mathbf{X}) - \frac{D_0^2(1-b^3)}{3\pi\,(b\Omega)^3}\left(\partial_i\partial_j V_{eff}(\mathbf{X})\right)^2. \qquad (24)$$

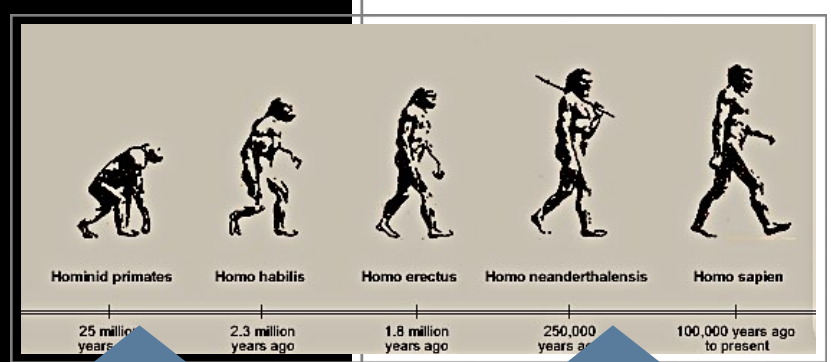
Note that the first line is the leading order term (i.e. $L = 1$) , while the second and third lines display the order $L = 2$ and $L = 3$ corrections, respectively.

We emphasize that the result of the EST construction is a new expression for the *same* path integral (15), in which the UV cutoff been lowered from Ω to $b\Omega$. Equivalently, the path integral is discretized according to a larger elementary time step, $\Delta t \to \Delta t/b$:

$$Z^{\Delta t}(t) \equiv \oint_{\Delta t} \mathcal{D}\mathbf{X}\, e^{-S_{eff}[\mathbf{X}]} \propto \oint_{\Delta t/b} \mathcal{D}\mathbf{X}\, e^{-S_{eff}[\mathbf{X}] - \int_0^t d\tau\, V^R_{eff}[\mathbf{X}(\tau)]} \equiv Z^{\Delta t/b}_{EST}(t) \qquad (25)$$

In these expressions, the symbol $\oint_{\Delta t}$ denotes the fact that the path integral is discretized according to an elementary time step Δt and we have suppressed the subscript "<", in the paths. It can be shown that the proportionality factor between $Z^{\Delta t}(t)$ and

HUGE COMPUTATIONAL GAIN

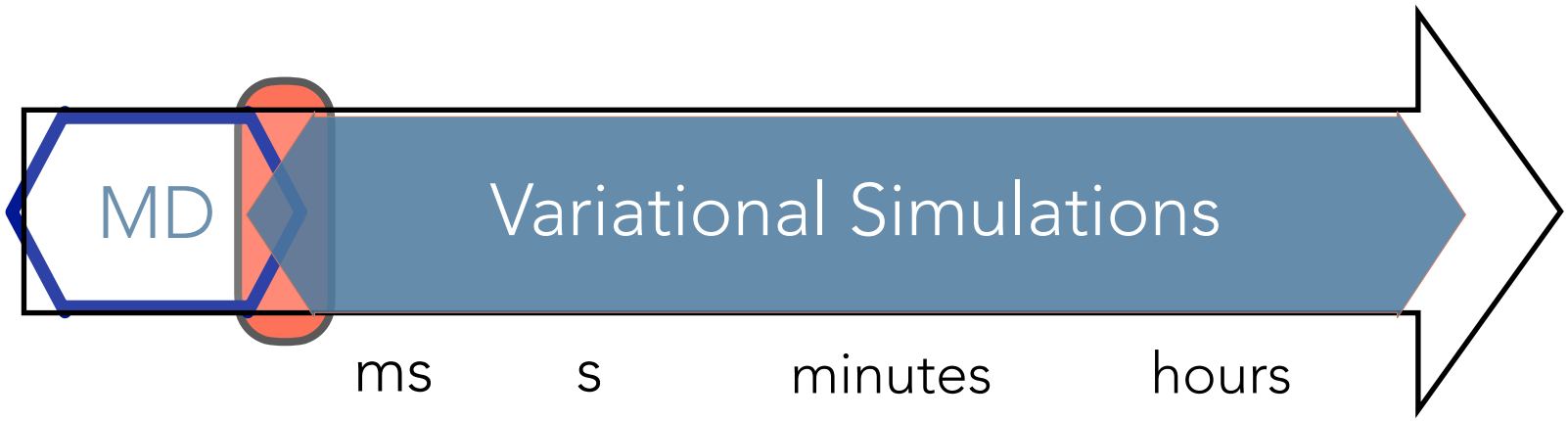


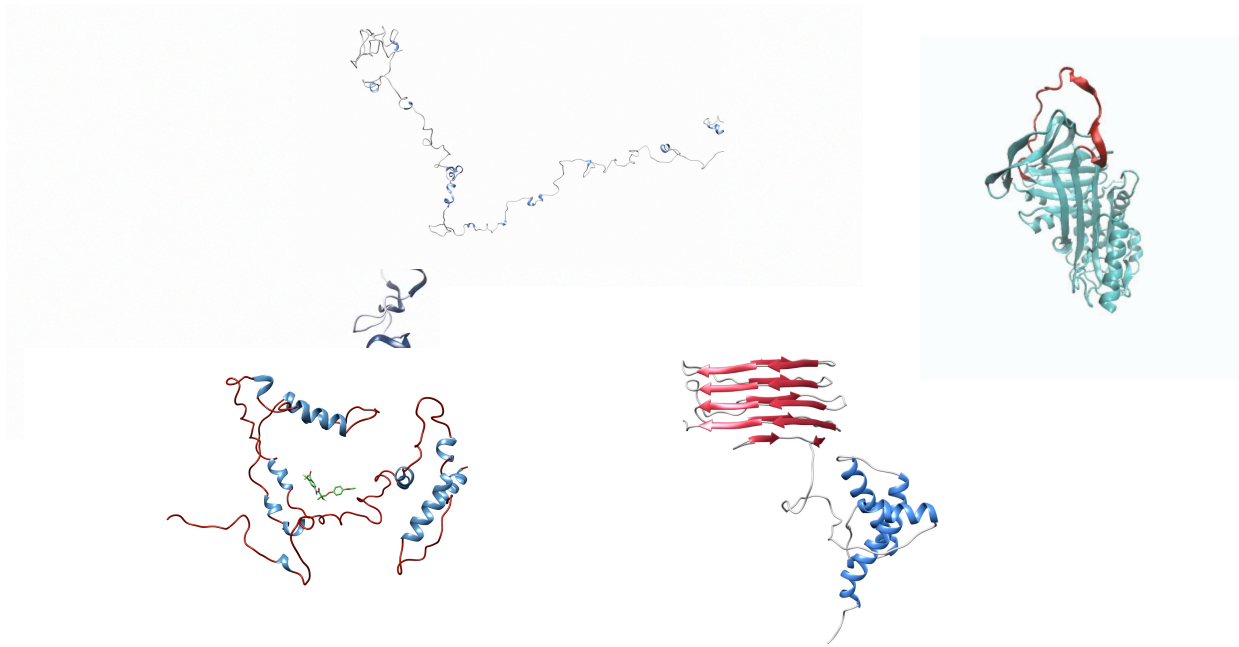
Sybilla & INFN joint project on ACE2

Using top all-purpose supercomputers

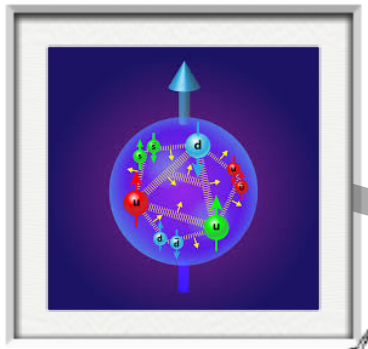
Using top special-purpose supercomputer

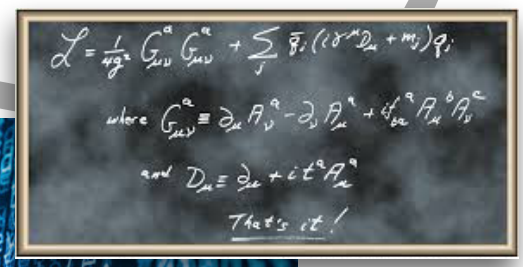
VENTURING INTO THE BIO-ZONE

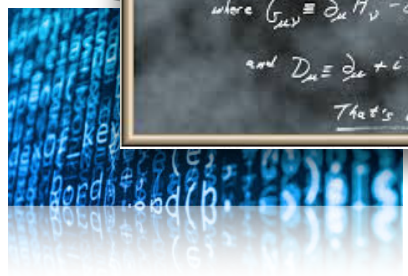




THE WAY TO PHARMACOLOGICAL RESEARCH

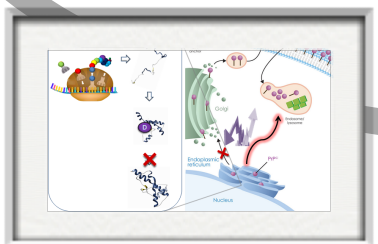






Experiment

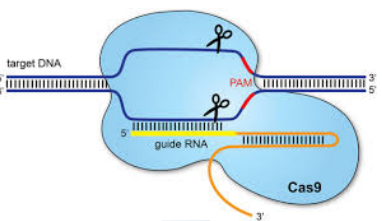
Theory



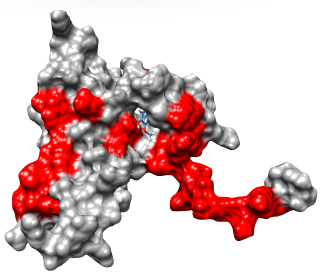
PHARMACOLOGICAL PROTEIN INACTIVATION BY FOLDING INTERMEDIATE TARGETING

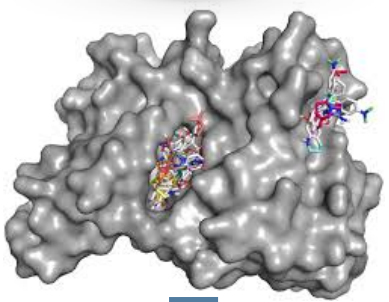
patent file # 102018000007535 (with E. Biasini)

DNA → mRNA → Ribosome → Protein folding → Protein function



RNA silencing





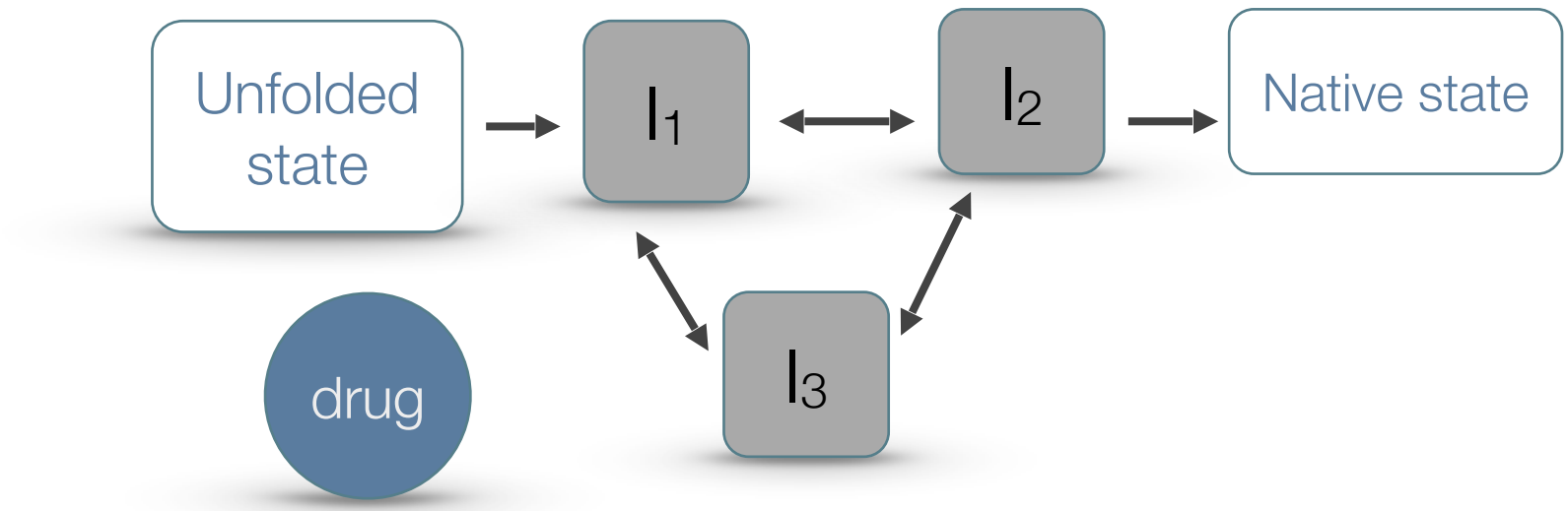
Eliminating protein from genome

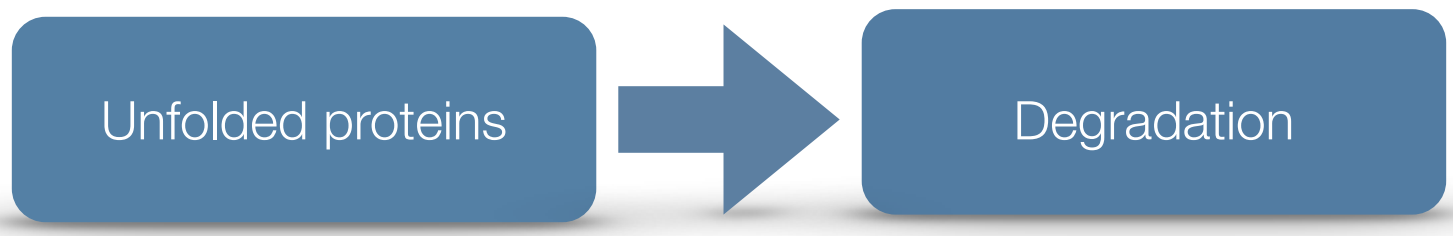
Suppressing transcription

Suppressing folding

Imparing function

PHARMACOLOGICAL PROTEIN INACTIVATION BY FOLDING INTERMEDIATE TARGETING



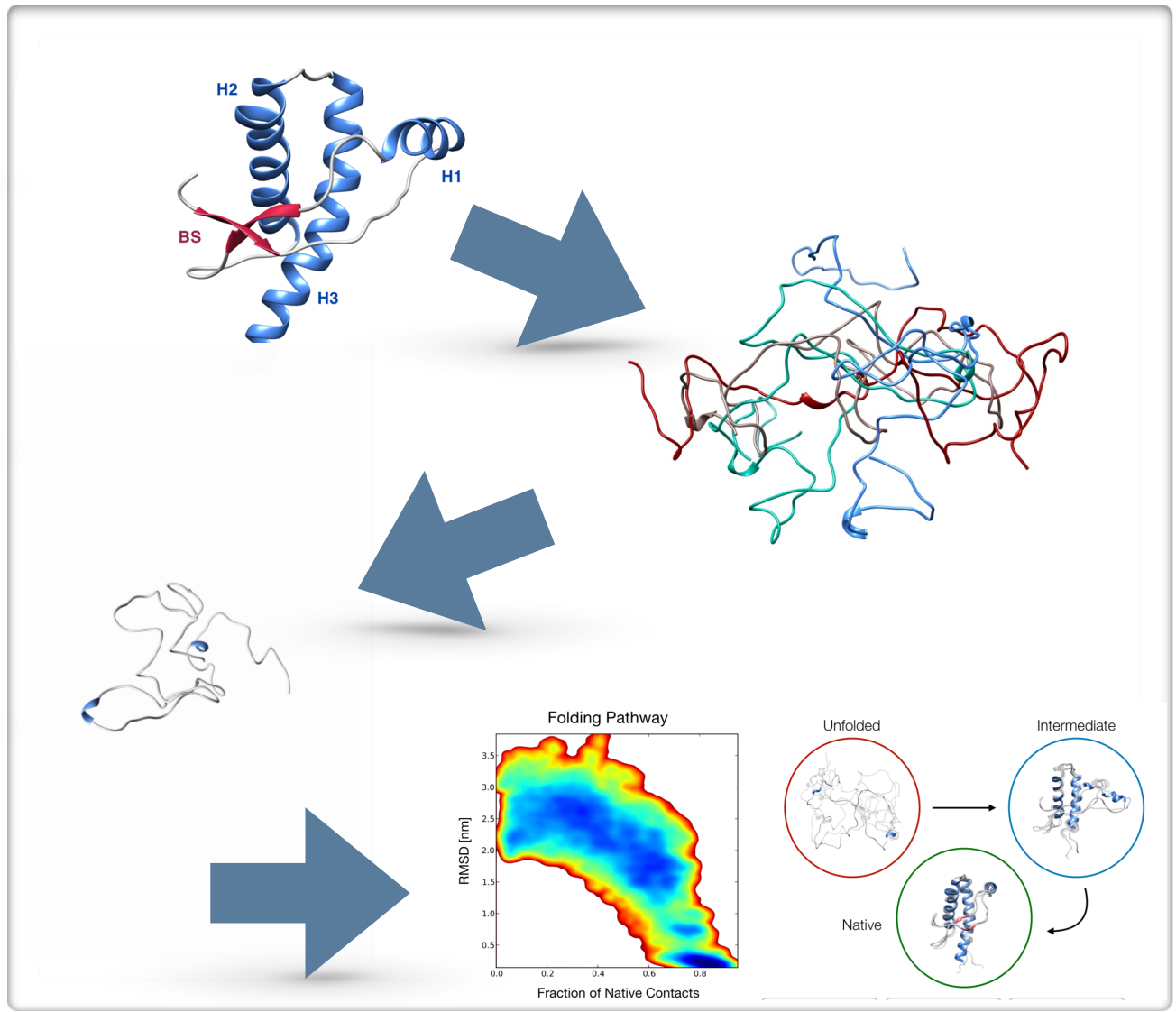


PPI-FIT PIPELINE

Folding pathway characterization

Virtual screening on folding Intermediate

Biochemical Validation

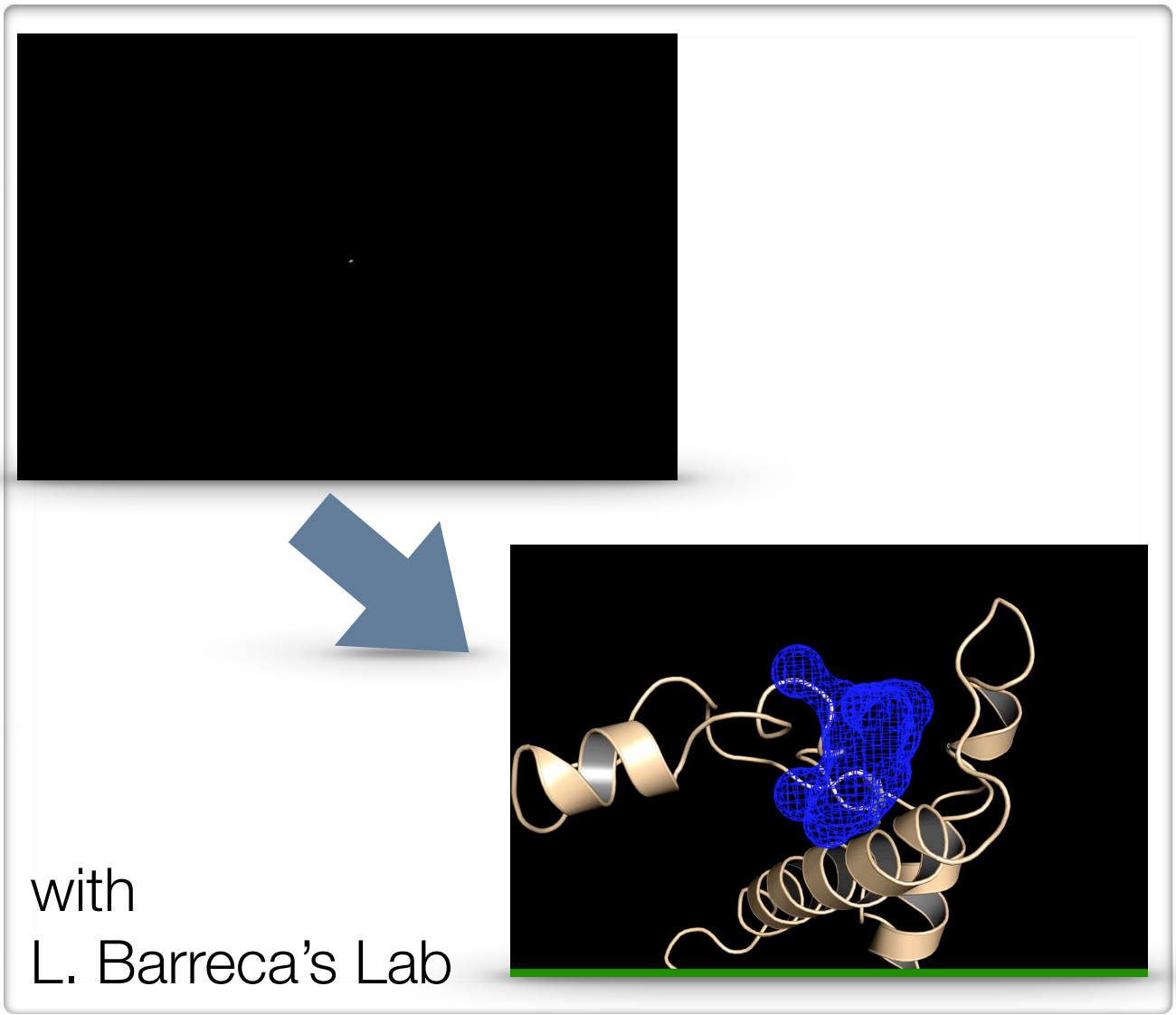


PPI-FIT PIPELINE

Folding pathway characterization

Virtual screening on folding Intermediate

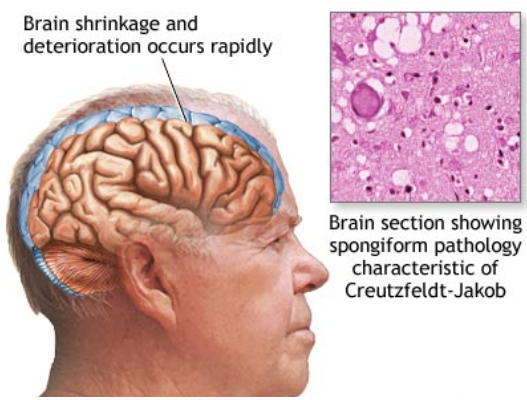
Biochemical Validation



with
L. Barreca's Lab

A FIRST VALIDATION

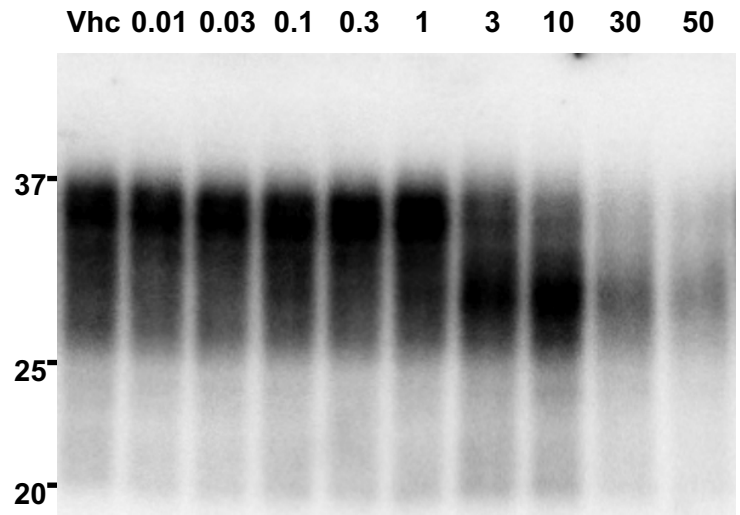
Inactivation of Cellular Prion protein

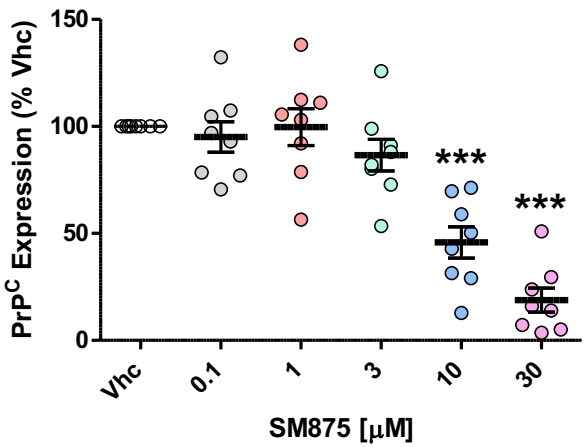


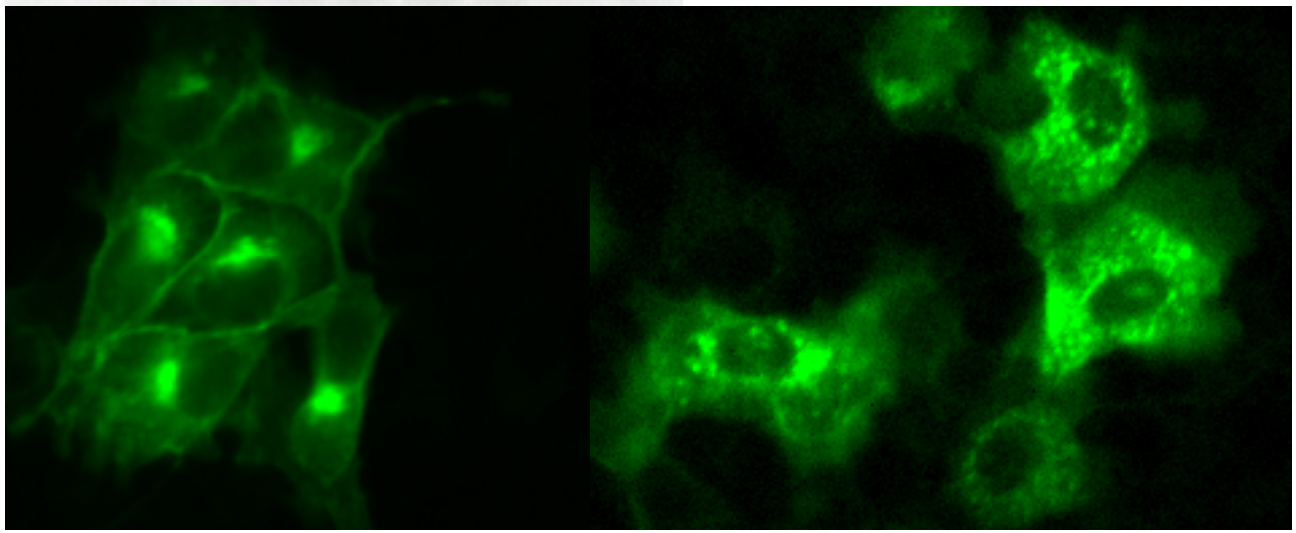
PHARMACOLOGICAL PROTEIN INACTIVATION BY TARGETING FOLDING INTERMEDIATES

Giovanni Spagnolli, Tania Massignan, Andrea Astolfi, Silvia Biggi, Paolo Brunelli, Michela Libergoli, Alan Ianeselli, Simone Orioli, Alberto Boldrini, Luca Terruzzi, Giulia Maietta, Marta Rigoli, Nuria Lopez Lorenzo, Leticia C. Fernandez, Laura Tosatto, Luise Linsenmeier, Beatrice Vignoli, Gianluca Petris, Dino Gasparotto, Maria Pennuto, Graziano Guella, Marco Canossa, Hermann Clemens Altmeppen, Graziano Lolli, Stefano Biressi, Manuel Martin Pastor, Jesús R. Requena, Ines Mancini, Maria Letizia Barreca, Pietro Faccioli, Emiliano Biasini

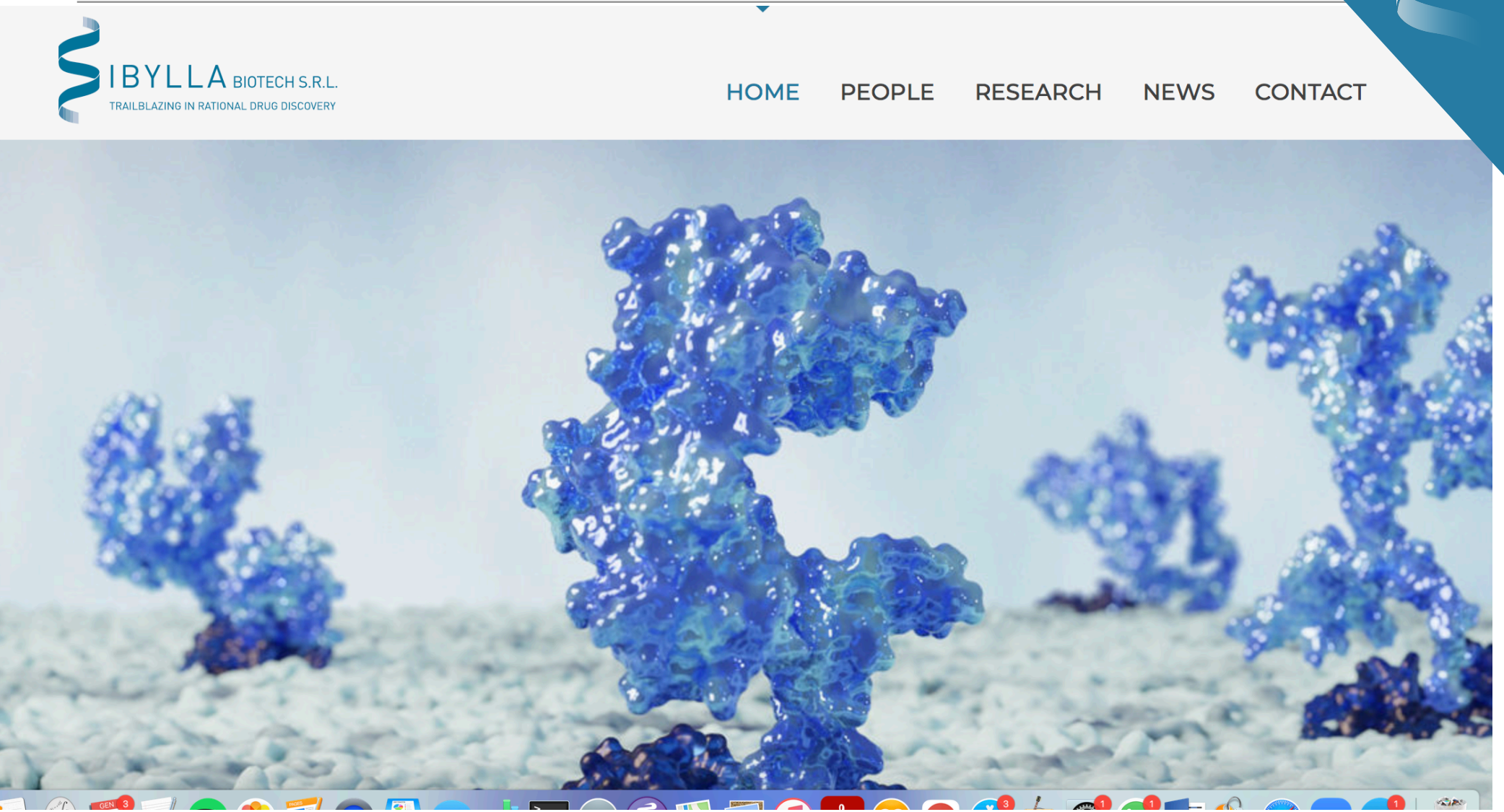
doi: https://doi.org/10.1101/2020.03.31.018069





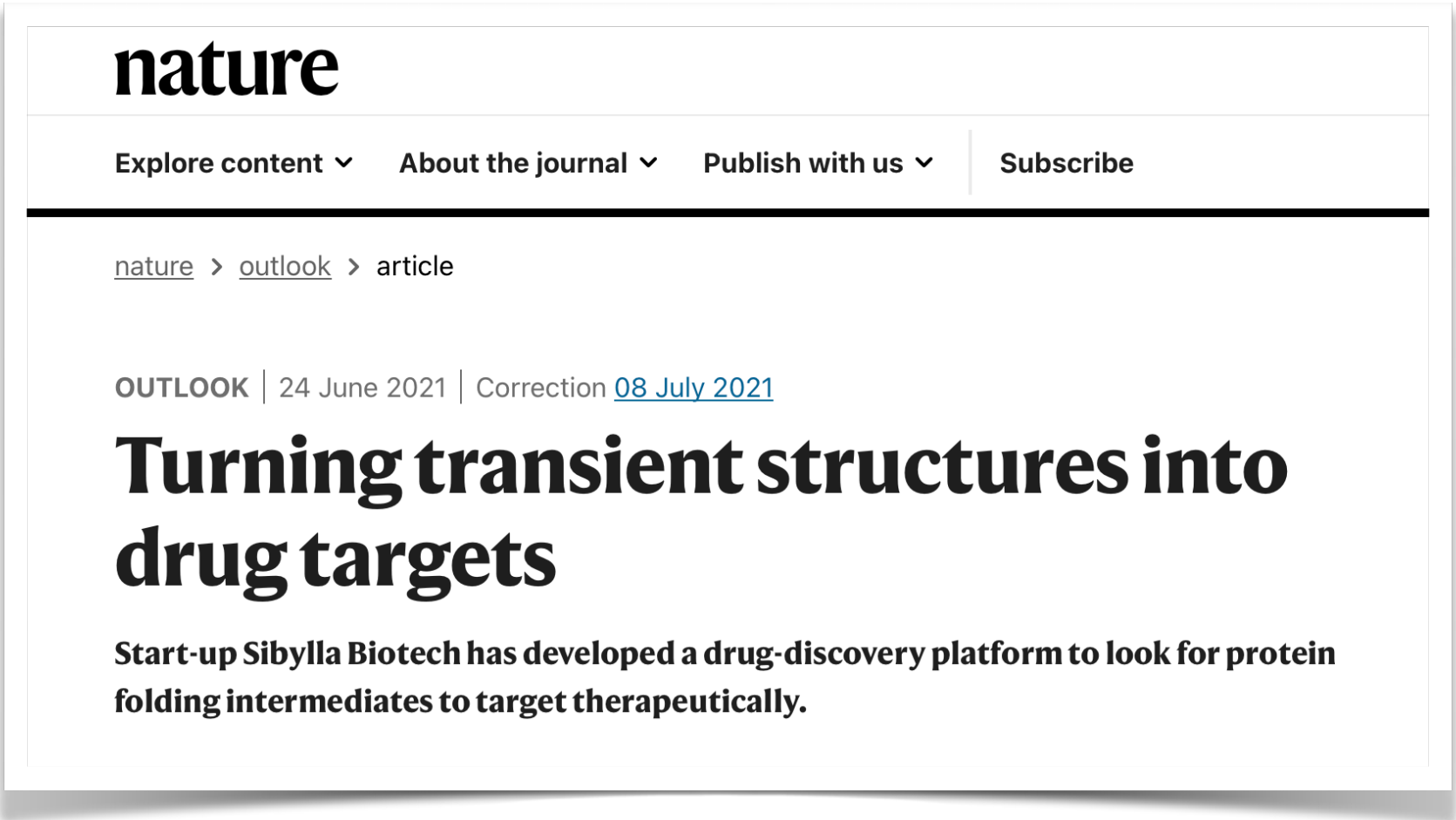


Technology Transfer Initiative



www.sibyllabiotech.it

A few facts about Siybilla Biotech



nature

Explore content | About the journal | Publish with us | Subscribe

nature > outlook > article

OUTLOOK | 24 June 2021 | Correction 08 July 2021

Turning transient structures into drug targets

Start-up Sibylla Biotech has developed a drug-discovery platform to look for protein folding intermediates to target therapeutically.

Future

Sibylla is closing its Series-A round!
(formal announcing expected in the Summer)

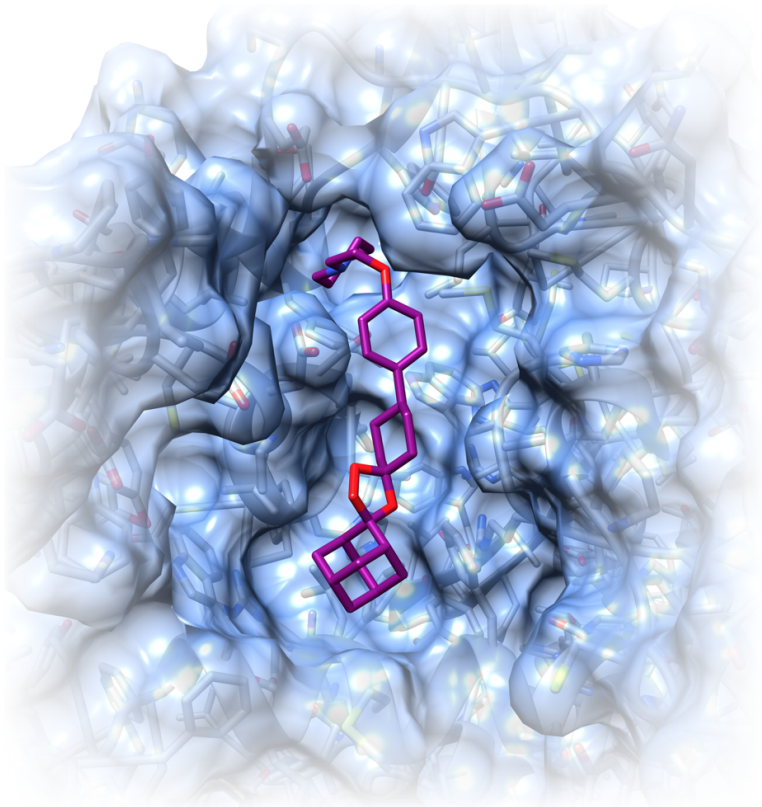
Sibylla is **currently hiring** several computational physicists!

WHAT DO NEW/EMERGING TECHNOLOGIES HAVE TO OFFER?

Partners:

U. Trento, Space Pharma, CJD Foundation (Israel), U. Tel Aviv, U. Santiago de Compostela, INFN

A MAIN LIMITING FACTOR



Impossible to crystallize folding intermediates on Earth

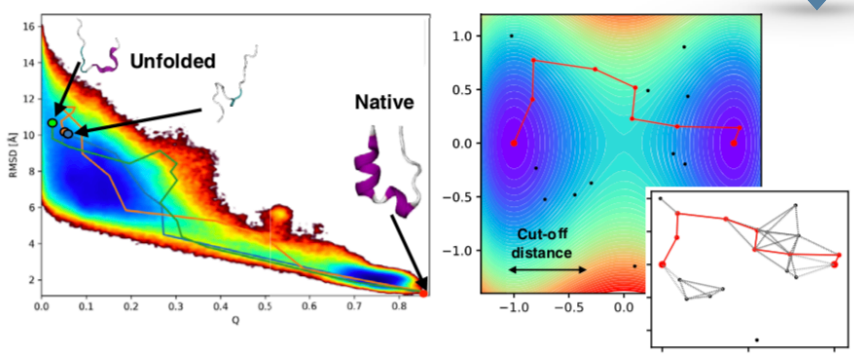
Microgravity conditions may provide the solution!

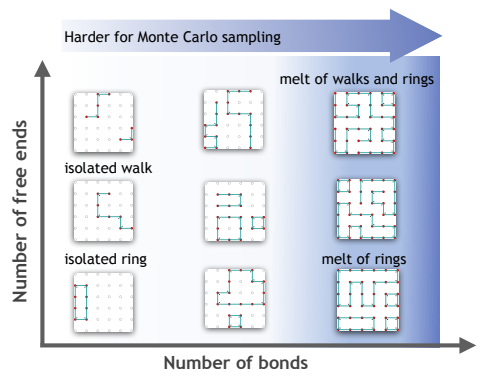
WHAT DO NEW/EMERGING TECHNOLOGIES HAVE TO OFFER?

Quantum Computing

Theoretical Physics

Artificial Intelligence





Partners:

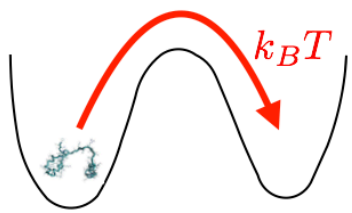
U. Trento, Q@Trento, INFN, SISSA, BEC-CNR

Final considerations (very subjective!)

Fundamental science can breed new ideas

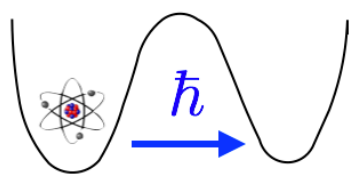
Thermal activation

$(\beta = (K_B T)^{-1})$



$$P(x_f, t|x_i) = \int_{x_i}^{x_f} \mathcal{D}q \; e^{-\frac{\beta}{4M\gamma}\int_0^t d\tau (M\ddot{q}+M\gamma\dot{q}+\nabla U(q))^2}$$

Quantum tunneling



$$K_E(x_f, t|x_i) = \int_{x_i}^{x_f} \mathcal{D}q \; e^{-\frac{1}{\hbar}\int_0^t d\tau \left(\frac{M}{2}\dot{q}^2+U(q)\right)}$$

Cross-disciplinarity is key to tackle **complexity**.
Seek for colleagues with different background…
…..and **learn to talk a lot!**

Final considerations (very subjective!)

If you have a good idea… **”money is not the limiting factor issue”!**

Patenting is not the Enemy of Science! …

..but just don’t wait too much to look out !!!

Acknowledgments

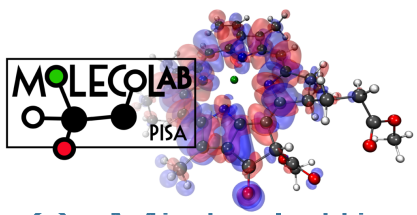
TIFPA Trento Institute for Fundamental Physics and Applications

FONDAZIONE Telethon

Italy:

Trento: *E. Biasini*, A. Ianeselli (2014-2017), G. Spagnolli S. A Beccara (2009-2017), S. Orioli (2014-2018), E. Schneider (2012-2015), M. Carli (2017), M. Turelli (2018), F. Mascherpa (2014), *G. Garberoglio, F. Pederiva, M. Sega*, R. Covino (2012-2015),

Pisa: B. Mennucci, L. Cupellini, S. Jurinovich



Perugia: L. Barreca

SISSA: C. Micheletti, A. Laio

Europe:

U. Zurich: B. Schuler

U. Compostela: J. Raquena

CEA-Saclay: H. Orland

USA:

U. Maryland: P. Wintrode

U. Mass.: A. Gershenson