PRIN Report - UniTO

Integrative tools for defining the molecular basis of the diseases:
Computational and Experimental methods for Protein Variant Interpretation.

Giovanni Birolo, Tiziana Sanavia and Piero Fariselli

Wednesday September 9, 2020

Report 2019/2020 – UniTo (Unit 1)
PRIN-201744NR8S

MINISTERO DELL' ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA

Computational and Experimental
Protein Variant Interpretation
https://ceprovi.github.io/

Department of Medicine
University of Torino

Project aims

The main aim of our project consists in filling the gap between thermodynamic data and disease-related information on protein variants.

1. integrate theoretical/computational approaches with experimental validations to assess the impact of amino acid variations on protein structure, function and protein-protein binding affinity.
2. We will generate a comprehensive database collecting all the structural and functional protein variants associated to diseases, with specific reference to predicted or experimentally available thermodynamic data.
3. Then, we will use those data to implement customized methods for predicting the impact of variants on proteins associated to cancer and to genetic diseases affecting calcium signalling.

Project

Definition of the objectives and deliverables in the light of the 38% cut

Project Units

- Unit1: unito
- Unit2: unibo
- Unit3: univr
- Unit4: uniroma1
- Unit5: infn roma2

PRIN Meeting - Sep 9, 2020
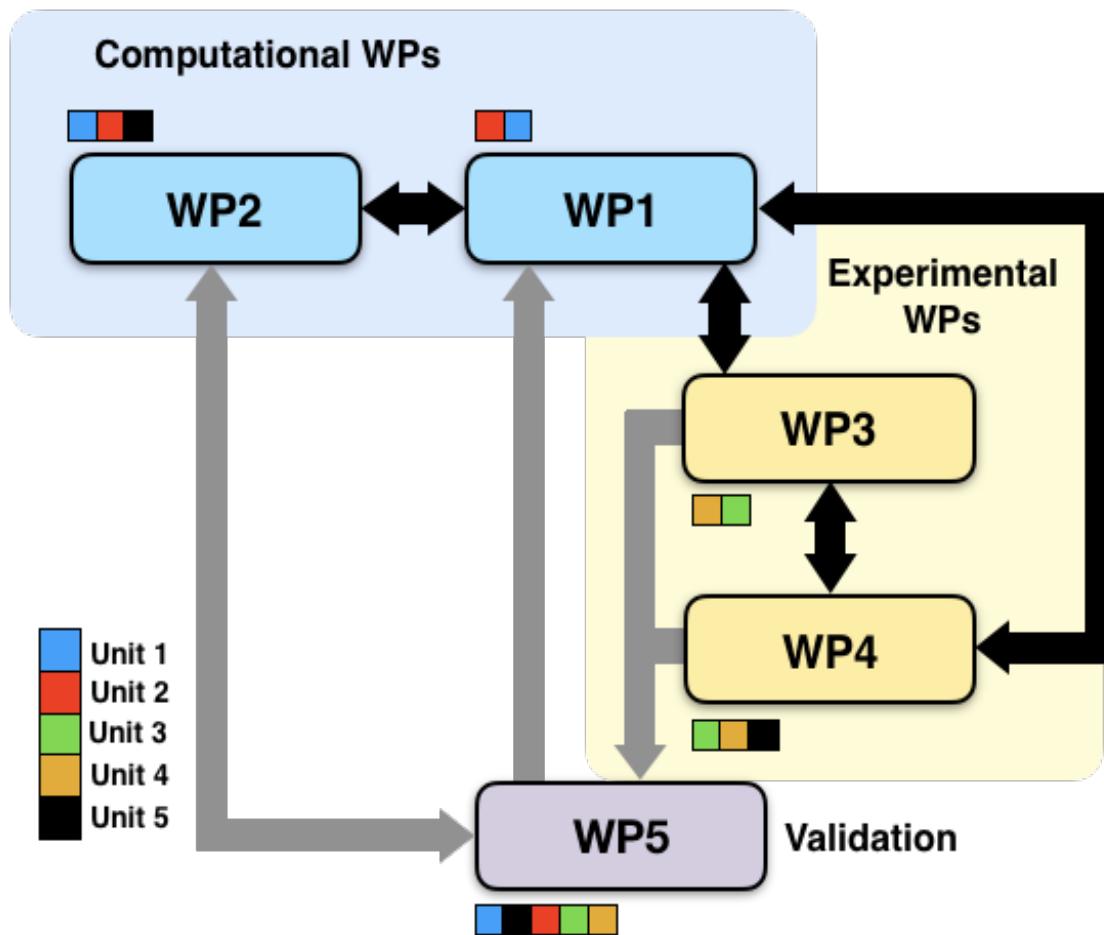
Project WP

- WP1: Development of an integrated protein variation database (months 1:36)

- WP2: Development of the Predictors (months 5-30)

- WP3: Generation of new experimental data: structural, functional and stability (months 1-32)

- WP4: Generation of new experimental data: binding affinity variations (months 5-32)

- WP5: Data validation and evaluation of the predictors (Months: 25-36)

Project Workflow

Project GANTT

	Year I												Year II												Year III											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
WP1	Database implementation and development																																			
WP2					Developments of the Predictors																															
WP3	Generation of new experimental data: structural, functional and stability																																			
WP4					Generation of new experimental data: binding affinity variations.																															
WP5																									Data validation and evaluation of the predictors											
Deliverables WP1																																				
Deliverables WP2																																				
Deliverables WP3																																				
Deliverables WP4																																				
Deliverables WP5																																				
Meetings																																				
Seminars																																				
Dissemination																																				
Public Engagement																																				
Reports																																				

Project WP1

WP1: Development of an integrated protein variation database (months 1:36)
- D1.1 (month 6): Implementation and design of the database.
- *D1.2 (month 12): Database refinement and web-based internal releas*

Project WP2

WP2: Development of the Predictors (months 5-30)
- *D2.1 (month 10->16): Development of the new DDG predictor.*

Project WP3

WP3: Generation of new experimental data: structural, functional and stability (months 1-32)

- D3.1 (month 3): Selection of protein kinases MAPK1, 3, 6, 8, 11 and phosphatases PTPN4, 11, 14 nsSNVs found in cancer tissues from databases.
- D3.2 (month 10->16): Expression and purification of wild type and selected nsSNVs of MAPK1, 3, 6, 8, 11 and phosphatases PTPN4, 11, 14 found in cancer tissues.
- D3.3 (month 12->18): Expression and purification of wild type and mutant CaM (at least 11 variants for LQTS /CPVT conditions; selected mutants (at least 7) for cancer-related variants) in normal and 15N-labelled media.

Project WP4

- **WP4**: Generation of new experimental data: binding affinity variations (months 5-32)
No deliverable first year

Project WP5

- WP5: Data validation and evaluation of the predictors (Months: 25-36)
- *No deliverable first year*

Presentations

1. Morante "Dealing with "damage" in X-ray absorption spectroscopy experiments", EBS workshop on "Sample Modulation by High Photon Densities: Desired and Undesired Effects", ESRF, Grenoble, France - 11th December 2019
2. Morante in XXIV INTERNATIONAL SCHOOL OF PURE AND APPLIED BIOPHYSICS "Applications of X-rays and Neutron Scattering in Biology" Venice (Italy), Palazzo Franchetti - 27-31 January 2020 1
3. 3/12/2019: "Physiological and pathological regulation of vertebrate phototransduction: from molecules to networks", Invited seminar, Institut de Biologie Valrose Université Nice Côte d'Azur, Nice, France (speaker. D.Dell'Orco)
4. 19/09/2019: PhD week 2019, University of Trieste, Institute of Medical Genetics. Invited talk entitled: "Molecular basis of retinal dystrophies: from individual proteins to networks" (speaker. D.Dell'Orco)
5. 30/09/2019: 98th Meeting of the German Physiological Society, Ulm, Germany, invited lecture entitled: "Calcium binding proteins and the regulation of the visual sensory system: from molecules to networks (speaker. D.Dell'Orco)
6. Benevenuta S, Capriotti E, Fariselli P. (2020) Calibrating variant-scoring methods for clinical decision making. VarI-COSI 2020 Meeting - International conference on Intelligent Iystems for Molecular Biology (ISMB). Iirtual Conference. 13-16 July 2020.Speaker: Beneveuta Silvia

- Future plan:
- Meeting with an open workshop: 2021 summer-autumn?
- Seul?
- ?

Computational and Experimental
Protein Variant Interpretation
https://ceprovi.github.io/

Publications

1. M. Banchelli et al., "Nanoscopic insights into the surface conformation of neurotoxic amyloid β oligomers" RSC Advances, 37 (2020)
2. Dinh Quoc Huy Pham et al., "Computational Model to Unravel the Function of Amyloid-β Peptides in Contact with a Phospholipid Membrane" J. Chem. Phys. B124 (16) 3300 (2020)
3. S. Morante, G. La Penna, G.C. Rossi, F. Stellato, "SARS-CoV-2 virion stabilization by Zn binding", Frontiers in Molecular Biosciences (2020)
4. G. La Penna and S. Morante, "Aggregates sealed by ions" in "Computer Simulations of Aggregation of Proteins and Peptides", Eds. Mai Suan Li, Marek Cieplak, Maksim Kouza and Andrzej Kloczkowski. Methods in Molecular Biology, Springer 2020.
5. Sanavia T, et al (2020). Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. Computational and Structural Biotechnology Journal. 18: 1968-1079.
6. Benevenuta S, Fariselli P. On the Upper Bounds of the Real-Valued Predictions. Bioinform Biol Insights. 2019 Aug 23;13:1177932219871263.
7. Savojardo C et al. On the critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. Brief Bioinform. 2019 Dec 28:bbz168. doi: 10.1093/bib/bbz168.

- Future plan:
- Common paper ?
- Challenge with new mutants?
- …

PRIN Report - UniBO

Integrative tools for defining the molecular basis of the diseases:
Computational and Experimental methods for Protein Variant Interpretation.

Emidio Capriotti and Paola Turina

Wednesday September 9, 2020

Report 2019/2020 - UniBO (Unit 2)
PRIN-201744NR8S

MINISTERO DELL' ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA

Computational and Experimental
Protein Variant Interpretation
https://ceprovi.github.io/

Department of Pharmacy
and Biotechnology (FaBiT)
University of Bologna

Research Activity

- Database analysis: ClinVar, SwissVar and ProTherm

- Database curation: ProTherm, ThermoScan

- Server implementation: ProTherm. ProtDDG-Bench, ThermoScan

Database Analysis

Analysis of databases of variant effect protein stability and function

- Analysis of the databases ClinVar and SwissVar for generating disease specific datasets based on OMIM and Orphanet classification

- Analysis of the ProTherm database to define curated and unbiased sets for benchmarking new methods for predicting the ΔΔG upon single point mutation.

ProTherm
Thermodynamic Database
for Proteins and Mutants
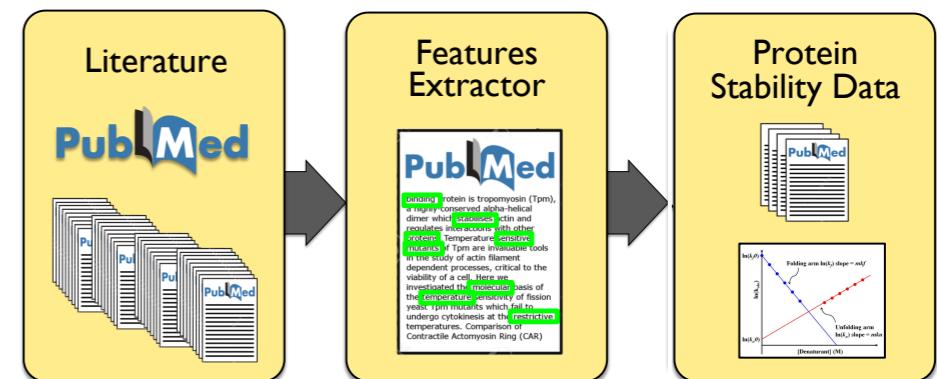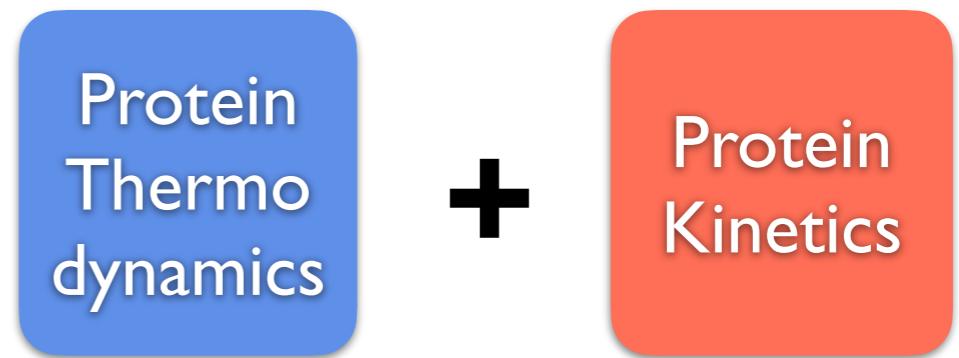
Database curation

New implementation and update of ProTherm database

- Development of an updated version ProTherm web server including kinetic data

- Semi-automatic data-mining method based on the analysis of fulltext html files

Server implementation

Development of web server for data sharing, collection and prediction

- Development of ProtDDG-Bench server on GitHub for dataset "FAIRification"

- Implementation or web server version of ThermoScan for scanning PMC articles

- New version of ProTherm database

- New version of I-mutant for predicting protein variant ΔΔG

ProtDDG-Bench
Benchmarking resource for predictors of protein stability change
https://protddg-bench.github.io/
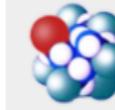
ThermoScan
BioFolD Scan biomedical publications to retrieve thermodynamic data.
https://folding.biofold.org/ithermoscan/

ProTherm
BioFolD Thermodynamic Database for Proteins and Mutants
https://folding.biofold.org/iprotherm/

I-Mutant+
BioFolD Predicting protein stability change upon mutation
https://folding.biofold.org/i-mutant+/

Computational and Experimental Protein Variant Interpretation
https://ceprovi.github.io/

PRIN Meeting - Sep 9, 2020

Dissemination

- **Publication:**

Sanavia T, et al. (2020). Computational and Structural Biotechnology Journal. 18: 1968-1079. DOI: 10.1016/j.csbj.2020.07.011 (Open Access)

- **Presentation:**

Benevenuta S, Capriotti E, Fariselli P. (2020) VarI-COSI 2020 Meeting - ISMB 13-16 July 2020.

Financial report

- Expenses:
  Server DELL EMC PowerEdge R7425

  TS2L3-UPS

  Apple MacBook Pro 13"

  Total ~16.000 Euro

- Human Resources:
  Co-Funded PhD Fellowship UniBO

  Data Science and Computation

Future activities

- Servers: Completing the server under development

- Manuscript: 1 manuscript under revision + 2/3 in preparation

- CAGI: Possible participation to CAGI

- Promoting collaborations: Write common manuscript or between units

PRIN Report – UniVR

Integrative tools for defining the molecular basis of the diseases:
Computational and Experimental methods for Protein Variant Interpretation.

Daniele Dell'Orco

Giuditta Dal Cortivo

Mariapina D'Onofrio

Carlo Barracchia

Wednesday September 9, 2020

Report 2019/2020 – UniVR (Unit 3)
PRIN-201744NR8S

MINISTERO DELL' ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA

Computational and Experimental
Protein Variant Interpretation
https://ceprovi.github.io/
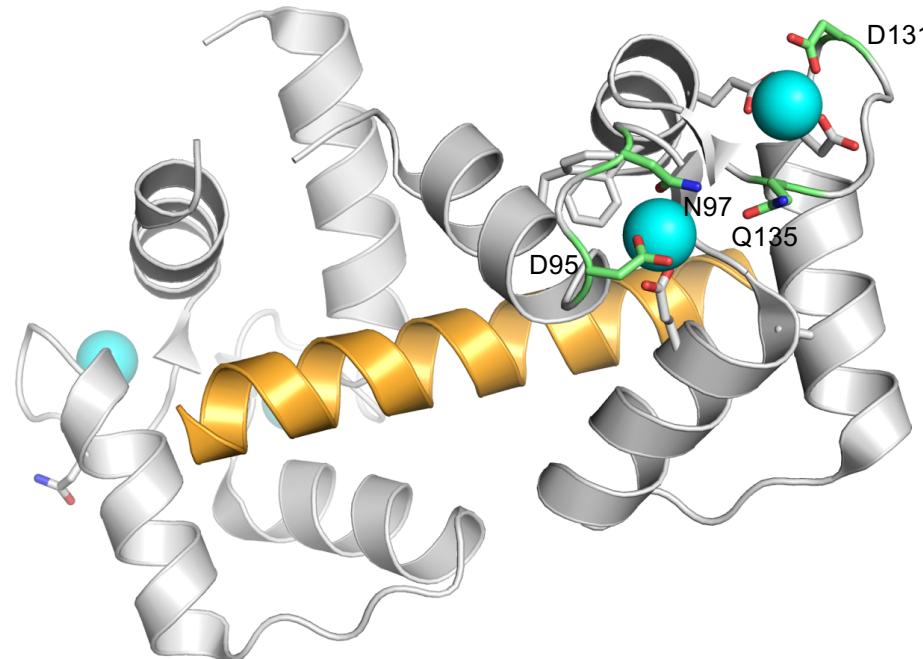
Department of Biotechnology
University of Verona

a) se ci sono stati cambiamenti (aggiunte/eliminazioni o spostamenti temporali) rispetto al previsto, illustrando le principali motivazioni;

WP3: Express and purify selected proteins and their variants (disease-associated SAVs deriving from SNV) *("min 11 arrhythmogenic+ 7 cancer variants in CaM"+ variants from Unit 4)*

➢ Ottenuti i *costrutti per le 17 varianti* di CaM associate ad aritmia (**NO varianti cancro**)

➢ Produzione e purificazione dei seguenti **otto** mutanti di calmodulina associati a tachicardia: **D95H**; **D95V**; **N97S**; **N97I**; **D131H**; **D131E**; **D131V**; **Q135P**

b) quale sia il reale progresso verso gli obiettivi previsti, indicando, altresì, gli eventuali risultati ottenuti;

➢ **Produzione del campione:** espressione di varianti di 6x-His CaM in *E.coli (con marcatura ^{15}N per studi NMR)*;
purificazione tramite cromatografia di affinità;
taglio del tag con TEV proteasi His-tagged (messa a punto del protocollo).

Risultati soddisfacenti.
Campioni ad elevata purezza

➢ **Caratterizzazione del campione:** Spettri CD nel lontano UV (struttura II) e vicino UV (struttura III)

Corretta risposta strutturale
a Ca^{2+} e peptide target

➢ **Spettroscopia NMR**: - messa a punto delle condizioni di acquisizione spettri mono e bi-
   dimensionali (HSQC) di varianti in presenza e assenza di Ca^{2+} e verifica del corretto folding
   - Spettri 3D per assegnamento del backbone, propedeutico agli esperimenti di   scambio H/D
volti a misurare il ΔG di unfolding delle varianti selezionate.

➢ **Interazione con il target:** Stechiometria e termodinamica dell'interazione CaM-RyR1/RyR2 (peptide) tramite
   PAGE nativa, calorimetria ITC, Surface Plasmon Resonance e dicroismo circolare (Ca^{2+} saturante).

Stechiometria 1:1; dati
qualitativamente soddisfacenti

> **WORK IN PROGRESS:**

- Assegnamento spettri NMR per ottenere $\Delta\Delta G_f° = \Delta G_f^{°mut} - \Delta G_f^{°wt}$ di varianti selezionate tramite scambio H/D e confronto con stabilita' termiche relative (CD/DSC)

- Caratterizzazione SPR/ITC per ottenere $\Delta\Delta G_b°$ (binding CaM/RyR1 e RyR2) e parametri cinetici

c) come i risultati gia ottenuti verranno sfruttati nell'ambito delle attivita in corso di svolgimento…

Manoscritto sulla caratterizzazione strutturale e biofisica delle varianti **D95H/V** e **D131H/E/V** di CaM, localizzate su EF-hand differenti, e della loro interazione con i target RyR1 e RyR2.

d) se sono sopraggiunte particolari difficoltà che mettano a rischio il conseguimento degli obiettivi minimi previsti.

- COVID-19
- Inaspettata **impressionante produttività di gruppi competitors**
- Ridimensionamento obiettivi (riduzione numero varianti) per limiti budget

PRIN Report – UniRM1

Integrative tools for defining the molecular basis of the diseases:
***C**omputational and **E**xperimental methods for **Pro**tein **V**ariant **I**nterpretation.*

Leonore Novak, Maria Petrosino, Alessandra Pasquo
Roberta Chiaraluce, Valerio Consalvi

Wednesday September 9, 2020
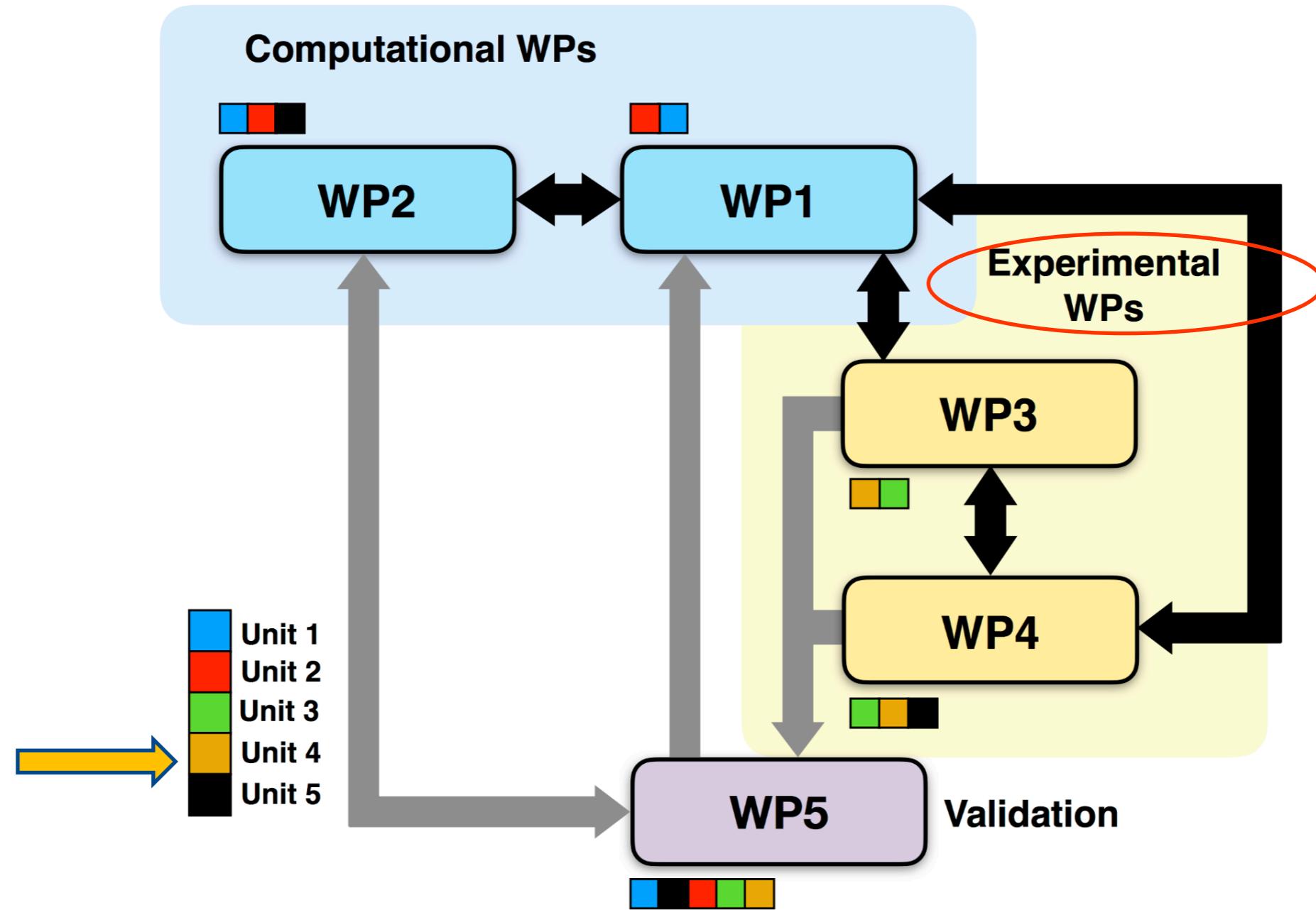
MINISTERO DELL' ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA

Computational and **E**xperimental
Protein **V**ariant **I**nterpretation
https://ceprovi.github.io/

Department of Biochemistry "A. Rossi Fanelli"
Sapienza University of Rome

Experimental
Generation of new experimental data: structural, functional and stability

➢**WP3: Generation of new experimental data: structural, functional and stability (months 1-32)**
Leading unit: UNIT4
Participants: UNIT3 and UNIT5

➢Generate data for the development and testing of the new predictors.

➢Effect of somatic nsSNVs found in cancer tissues on the structural properties in solution of protein kinases MAPK and phosphatases PTPN and CaM variants

➢Determination of $\Delta\Delta G$ values (ΔGvariant- ΔGwt).

➢The results and data obtained in this WP will be inserted in the database WP1, and used for the development and testing of the predictors (WP2, WP5).

Production and characterization of MAPK missense variants

➤MAPK1 and MAPK3 missense variants found in cancer tissues have been selected from COSMIC database

➤7 MAPK1 missense variants have been obtained with specific mutagenesis primers by polymerase chain reaction, using wild type MAPK1 as a template

➤MAPK1 wild type and variants have been expressed in E. coli and purified.

➤ The structural conformation of MAPK1 variants have been compared to that of the wild type by monitoring the near and far-UV circular dichroism and intrinsic fluorescence spectra.

➤The thermodynamic stability of MAPK1 wild type and variants has been measured at different concentrations of denaturant by monitoring the spectral changes (far-UV circular dichroism and intrinsic fluorescence emission) induced by the denaturant.

➤The thermal stability of MAPK1 wild type and variants has been measured by monitoring the changes of the dichroic activity as a function of temperature

Future prospective

➢ Funtional activity of MAPK1 variants will be measured and compared with that of wild type protein

➢ characterization of other MAPK variants

PRIN Meeting - Sep 9, 2020

PRIN Report - INFN

Integrative tools for defining the molecular basis of the diseases:
Computational and Experimental methods for Protein Variant Interpretation.

Gaetano Salina, Silvia Morante, Giovanni La Penna

Wednesday September 9, 2020

Report 2019/2020 – INFN (Unit 5)
PRIN-201744NR8S

MINISTERO DELL' ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA

Computational and Experimental
Protein Variant Interpretation
https://ceprovi.github.io/

Sezione INFN - University of Rome «Tor Vergata»

Istituto Nazionale di Fisica Nucleare
SEZIONE DI ROMA TOR VERGATA

People to be recruited
call for a PostDoc scholarship (1+1 years, call around October 15th)

Already acquired instrumentation
1) Assembled Workstation CoolerMaster PC-Ryzen7-RTX2080 –1431.06 euro
2) Mac Air for general purposes - 1495.05 euro

Publications
1) M. Banchelli et al., "Nanoscopic insights into the surface conformation of neurotoxic amyloid β oligomers" RSC Advances, 37 (2020)
2) Dinh Quoc Huy Pham et al., "Computational Model to Unravel the Function of Amyloid-β Peptides in Contact with a Phospholipid Membrane" J. Chem. Phys. B124 (16) 3300 (2020)
3) S. Morante, G. La Penna, G.C. Rossi, F. Stellato, "SARS-CoV-2 virion stabilization by Zn binding", Frontiers in Molecular Biosciences (2020)
4) G. La Penna and S. Morante, "Aggregates sealed by ions" in "Computer Simulations of Aggregation of Proteins and Peptides", Eds. Mai Suan Li, Marek Cieplak, Maksim Kouza and Andrzej Kloczkowski. Methods in Molecular Biology, Springer 2020.

Seminars
1) S. Morante "Dealing with "damage" in X-ray absorption spectroscopy experiments", EBS workshop on "Sample Modulation by High Photon Densities: Desired and Undesired Effects", ESRF, Grenoble, France - 11th December 2019
2) S. Morante in XXIV INTERNATIONAL SCHOOL OF PURE AND APPLIED BIOPHYSICS "Applications of X-rays and Neutron Scattering in Biology" Venice (Italy), Palazzo Franchetti - 27-31 January 2020

Computational and Experimental Protein Variant Interpretation
https://ceprovi.github.io/

PRIN Meeting - Sep 9, 2020

Folding of wild-type vs. mutant proteins
Protein complexes and aggregates

Numerical Approach

Variational Method for the computon of free energy ΔG using
«Altruist Metadynamics» in combination with «Constrained Maximal Entropy»

Computing resources

1) ISCRA B (CINECA) Project "Modelling the polymorfism of frataxin - PolyFrat"

2) GPU PC-Ryzen7-RTX2080
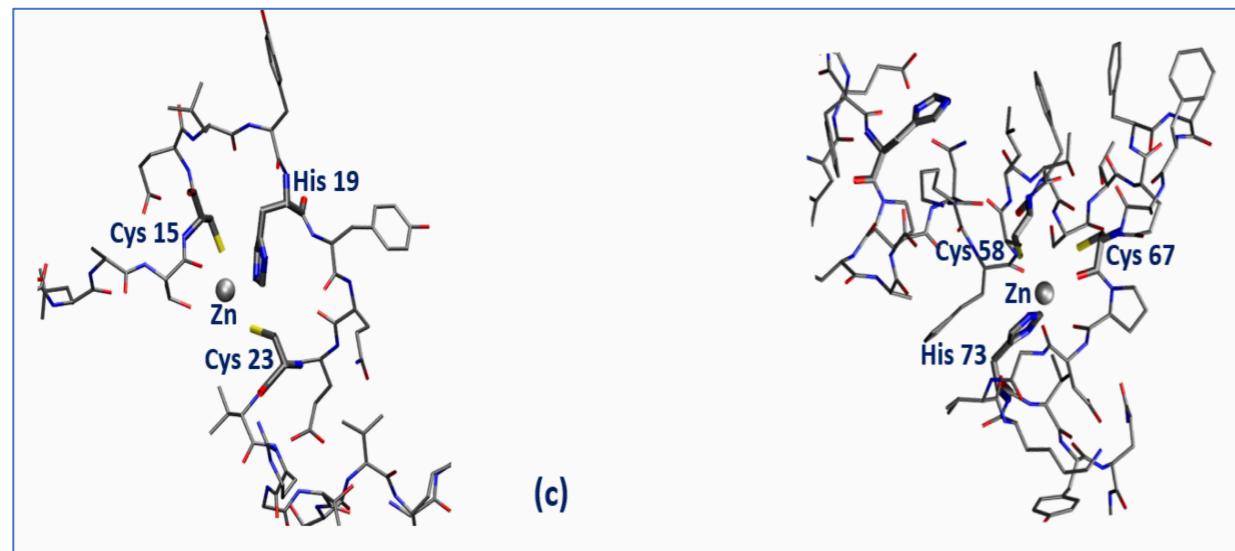
Assembled Workstation CoolerMaster
Processor: Ryzen 7 3700X 2x 8GB 3600MHZ
GPU: GTX 2080 8GB DDR6 ram
HD 1: 500 GB SSD 2.5"
HD 2: 1 TB HD 2.5"
Power supply: 1200W 80Plus M/B X570

Computational and Experimental
Protein Variant Interpretation
https://ceprovi.github.io/

PRIN Meeting - Sep 9, 2020
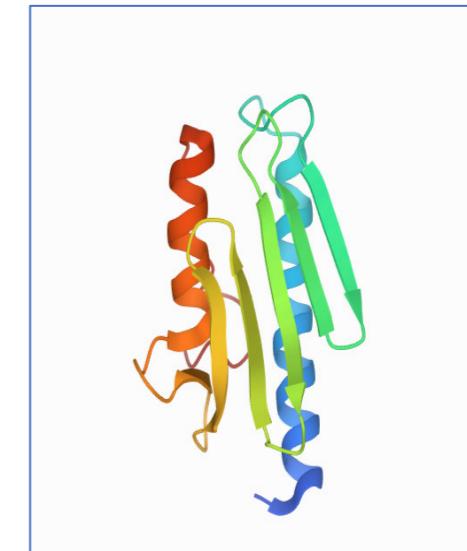
Targets

1) Comparison between the folding ΔG of the wt protein with that of proteins having single point mutations
2) Comparison between aggregation ΔG of viral proteins and CD in various configurations

Road-Map

1. Test of the method on short peptides: polyalanin with the substitutions A9R/A14R/A19R
2. Extension to
   2A) Frataxin and its 8 single-point mutants
   2B) NSP's orf7a and orf8 (Covid19) in complex with BST2 (CD317) tetherin and Zn
3. Comparison with some published results (Steinbrecher et al. 2017: 10 proteins, 741 sp mutations)
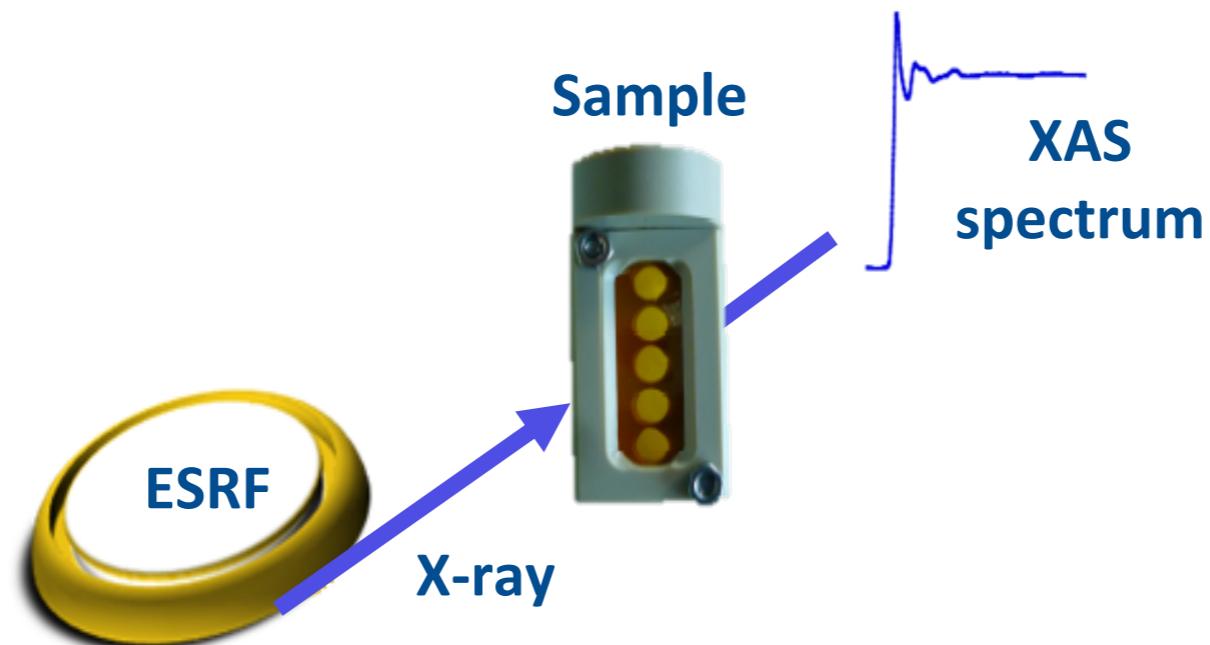
Frataxin PDB available structures

PDB ID	Mutation(s)	# Mutations	Mutation res (vs 3S4M)	# residues	Method	resolution	Reference
1EKG	No			119	XP	1.8	1
1LY7	No			121	SOLUTION NMR		2
3S4M	No			121	XP	1.3	3
3S5D	Yes	1	74 (W->A)	121	XP	1.5	3
3S5E	Yes	1	74 (W->R)	121	XP	1.31	3
3S5F	Yes	1	74 (W->F)	245	XP	1.5	3
3T3L	Yes	1	72 (Q->A)	121	XP	1.15	4
3T3J	Yes	1	65 (N->K)	123	XP	1.7	4
3T3K	Yes	1	67 (Q->R)	122	XP	1.24	4
3T3T	Yes	1	67 (Q->G)	480	XP	1.38	4
3T3X	Yes	1	84 (R->C)	243	XP	1.57	4

1. **Musco, G.**, **Stier, G.**, **Kolmerer, B.**, **Adinolfi, S.**, **Martin, S.**, **Frenkiel, T.**, **Gibson, T.**, **Pastore, A.** (2000) Structure **8**: 695-707
   DOI: 10.1016/s0969-2126(00)00158-1
2. **Dhe-Paganon, S.**, **Shigeta, R.**, **Chi, Y.I.**, **Ristow, M.**, **Shoelson, S.E.** (2000) J Biol Chem **275**: 30753-30756 **DOI:** 10.1074/jbc.C000407200
3. **Tsai, C.L.**, **Bridwell-Rabb, J.**, **Barondeau, D.P.** (2011) Biochemistry **50**: 6478-6487 D**OI:** 10.1021/bi200666h
4. Bridwell-Rabb, J., Winn, A.M., Barondeau, D.P. (2011) Biochemistry 50: 7265-7274 **DOI:** 10.1021/bi200895k

>TCGA

Computational and Experimental
Protein Variant Interpretation
https://ceprovi.github.io/

X-ray Absorption Spectroscopy (XAS)

Sample

XAS spectrum

ESRF

X-ray

Projects that have been already approved by the ESRF panel

1) Frataxin and its mutants in the presence of Fe or Co
2) orf7a/orf8 + tetherin BST2 complexes in the presence of Zn

(spectra will be recorded at ESRF in the week October 7-14 2020)