**Integrating DDGun into IPRO**

DDGun1 is an untrained method for predicting the ΔΔG (Gibbs free energy change of unfolding between wild type and variant proteins) for single and multiple variations from sequence or structure information.

DDGun predicts the score for M site variations, Ss= (s1, s2, …, sM) as follows:

|  |  |
| --- | --- |
| 𝑆𝑚𝑢𝑙𝑡=max(𝑆𝑠)+min(𝑆𝑠)−𝑚𝑒𝑎𝑛(𝑆𝑠) | (1) |

Variants with are stable, while are unstable. In order to prevent the selection of destabilizing mutations (negative ) inside IPRO2, we implemented the linearized form of the equation (1) within IPRO as additional constraints (equations (7)-(16)) by computing the DDGun scores for all possible mutations.

**MILP formulation**

Sets:

|  |  |
| --- | --- |
|  | set of all design positions |
|  | set of rotamers at position i |
|  | universal set of all feasible residue position and amino acid rotamer combinations |
|  | set of residue position and amino acid rotamer combinations for the design obtained from any iteration |

Parameters:

|  |  |
| --- | --- |
|  | interaction energy between substrate and backbone |
|  | interaction energy between rotamer r at position i and backbone |
|  | interaction energy between rotamer r at position i and substrate |
|  | interaction energy between rotamer r at position i and rotamer s at position j |
|  | prespecified cutoff value for the total energy of the protein and substrate complex |
|  |  |
|  | minimum of DDGun scores |
|  | maximum of DDGun scores |
|  |  |

Binary variables:

|  |
| --- |
|  |
|  |
|  |
|  |

*Continuous variables:*

|  |
| --- |
|  |
|  |
|  |
|  |

MILP formulation

Subject to:

|  |  |
| --- | --- |
|  | (2) |
|  | (3) |
|  | (4) |
|  | (5) |
|  | (6) |
|  | (7) |
|  | (8) |
| , | (9) |
|  | (10) |
|  | (11) |
|  | (12) |
|  | (13) |
|  | (14) |
|  | (15) |
|  | (16) |

The objective function of the MILP minimizes the binding score between the substrate and the protein. Equation (2) ensures that exactly one rotamer is selected at any given position i along the sequence (whether it is a design position or not). Equations (3) and (4) ensure that is only one if both and have values of one. Equation (5) prevents the same amino acid rotamer combination from being chosen at the same design position in follow up iterations. Equation (6) ensures that the mutations or rotamer alterations do not have a detrimental effect on the protein’s overall stability by requiring that the total energy of the protein be below a prespecified cutoff value, Ecutoff. Equation (7) and (8) ensure that mir assumes a value of one when mutation happens at position i. Equation (9) and (10) describe the variables and , respectively. Equation (11), (12) impose the upper bound and lower bound on and . Collectively, equations (13), (14), and (15) are the linearized form of the definition of mentioned in equation (1). Finally, equation (16) precludes destabilizing mutations from being selected at the design positions.

References:

1. Montanucci, L., Capriotti, E., Frank, Y., Ben-Tal, N. & Fariselli, P. DDGun: An untrained method for the prediction of protein stability changes upon single and multiple point variations. *BMC Bioinformatics* **20**, 1–10 (2019).

2. Saraf, M. C. *et al.* IPRO: An iterative computational protein library redesign and optimization procedure. *Biophys. J.* **90**, 4167–4180 (2006).