

Recombinant protein quality evaluation: proposal for a minimal information standard

Ashley M. Buckle^{1,15}, Mark A. Bate¹, Steve Androulakis², Mario Cinquanta³, Jerome Basquin⁴, Fabien Bonneau⁴, Deb K. Chatterjee⁵, Davide Cittaro³, Susanne Gräslund⁶, Alicja Gruszka⁷, Rebecca Page⁸, Sabine Suppmann⁹, Jun X. Wheeler¹⁰, Deborah Agostini³, Mike Taussig¹¹, Chris F. Taylor¹², Stephen P. Bottomley¹, Antonio Villaverde¹³, Ario de Marco^{14,*}

¹The Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Australia

²Monash eResearch Centre, Monash University, Clayton, Victoria, Australia

³Cogentech, Protein Chemistry Unit, Milan, Italy

⁴Max Planck Institute of Biochemistry, Department of Structural Cell Biology, Martinsried, Germany

⁵Protein Expression Laboratory, SAIC-Frederick Inc., National Cancer Institute, Frederick, MD USA

⁶Structural Genomics Consortium, Karolinska Institutet, Department of Medical Biophysics and Biochemistry, Stockholm, Sweden

⁷IEO, Milan, Italy

⁸Brown University, Department of Molecular Biology, Cell Biology and Biochemistry, Providence, RI, USA

⁹Max Planck Institute of Biochemistry, Microchemistry Core Facility, Martinsried, Germany

¹⁰National Institute for Biological Standards and Control, Health Protection Agency, Hertfordshire, UK

¹¹Protein Technologies Group, Babraham Bioscience Technologies, Cambridge UK

¹²The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK.

¹³Institute for Biotechnology and Biomedicine and Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain

¹⁴Department Environmental Sciences, University of Nova Gorica, Nova Gorica, Slovenia.

*Corresponding authors: Ario de Marco, University of Nova Gorica (UNG), Rožna Dolina (Nova Gorica), Slovenia. Tel. 0039.3493542056; ario.demarco@ung.si

Presentation of the MIPFE checklist

A proposal for the introduction of the Minimal Information (MI) platform dedicated to the acquisition and annotation of data concerning recombinant proteins (Minimal Information for Protein Functionality Evaluation – MIPFE) was recently published [1] and discussed at the 5th Recombinant Protein Production Conference (Alghero 2008) and the 2009 PEP Talk meeting (San Diego). The benefits of such standards are generally recognized, although there are concerns regarding its implementation as well as its perception of being too invasive for research freedom [2].

The meaning attributed to stored data is perceived differently within the MI community. The necessity of optimizing the quality of protein quality data annotation is generally acknowledged [3,4], since ontology and formal correctness are

crucial for unambiguous data reporting and comparison, and ignoring such rules would decrease the accuracy of curation, lead to the loss of valuable information for efficient data mining and prevent the assessment of the experimental methods [5]. However, in certain domains further orthogonal corroboration of the same material used in reported experiments is highly desired for the identification and recognition of artifacts and assessment of the final results. For instance, it is still very often the case that published biological data are obtained with starting material, the structural characteristics of which have not been evaluated or made available [6]. As a result, there is a pressing need for good practice guidelines within publications and databases, as for example in the evaluation of the native state of proteins used for *in vitro* interaction assays [1,7].

The reputation of journals, as well as funding bodies, depends on data quality. However, data quality is often hard to evaluate during the peer-review process. This has not gone unnoticed in the editorial context, where, for example, improvements to the peer-review process have been suggested that will facilitate the collection, submission and validation of proteomic, microarray and, more recently, imaging data [8]. In addition, funding agencies are becoming increasingly concerned about the reliability and accessibility of data collected by laboratories which they fund [8-10]. We therefore argue that it is time to implement similar policies for the transparent and rigorous reporting of data in all publications concerning proteins. For example, it is often ignored that recombinant proteins form not only insoluble precipitates, but also soluble aggregates, mostly when carriers are fused to improve solubility [11-13]. Such aggregates may retain some function [13,14] and therefore, without controlled experiments aimed at defining monodispersity and native structure, the interpretation of experimental results is weakened. Thus, the scientific community (editors, reviewers, readers) must have access to the raw data to assess the biophysical characterization and, accordingly, be able to judge the quality of the proteins used in the experiments. Ideally, it will remain the responsibility of editors and referees to check the robustness of controls and, where necessary, to request further experiments using the original material. Integration of annotated control experiments into the main text offers a useful complementary evaluation tool for reviewers and readers. We consider that information concerning aggregation status and secondary structure should be reported as a minimal requirement for publication under Supplementary Material. These controls should be available when authors describe protein production as well as protein interaction experiments (pull-down, surface plasmon resonance, antibody/protein microarrays, and isothermal titration calorimetry).

In practice, it is important to define what is to be considered mandatory and what may remain optional within the MI package. An overly rigid and demanding protocol will be perceived as interference in the scientific work and most likely would be rejected by the community on these grounds. Recently, an interesting attempt at identifying a version of the MI guideline for

describing proteins interacting in complexes has been reported [15]. However, it is difficult to judge the efficacy of the approach since the number of participants who volunteered to deposit the required information was limited to five.

In order to offer a workable solution for describing the MI for the evaluation of recombinant protein quality we propose a solution involving a repository to store the relevant results concerning protein construct features and biophysical characterization. Uploading of the information into the database is available through the MIPFE site [16]. We have designed a loosely structured text form allowing authors to describe the minimal information from an experiment which can be made available to reviewers, editors, and ultimately to other scientists. The proposed format requires little effort by the user (*e.g.* cut and paste using a simple text editor on any computing platform), and is human readable, yet sufficiently structured and formatted to allow data meta-analysis. Non-textual experimental results, such as gels and graphs, can be uploaded as image files alongside the form. In addition to its simplicity, the form can be copied and re-used by the authors and indeed the scientific community. Once deposited and validated, the dataset is given a unique handle which can be referred to in published manuscripts (for instance, as Supplementary Material), and possibly as a DOI tagged entity, as suggested recently [17].

Only the essential amount of obligatory information concerning the construct must be provided by the authors in the MIPFE form, in order to avoid possible misinterpretation of any annotation [18,19]. The fields concerning characterization experiments remain optional and are intended as guidelines for controlled experiments that are run in order to evaluate protein structural quality.

Although our approach is designed to capture the minimal amount of data from the user as quickly and effortlessly as possible, the form does allow for raw data to be described and deposited, encouraging users to provide as complete an entry as possible. MI platforms evolve progressively to match needs and overcome limitations [20] and the logical future development of the one we propose could be the implementation of the MIBBI standardization guidelines for annotation [21,22], allowing more extensive annotation and ultimately data mining and bioinformatic analyses.

References

- de Marco A. Minimal Information: an urgent need to assess the functional reliability of recombinant proteins used in biological experiments. *Microb Cell Fact* 2008; **7**:20. [PubMed](#) [doi:10.1186/1475-2859-7-20](https://doi.org/10.1186/1475-2859-7-20)
- de Marco A, Stevastyanovich YR, Cole JA. Minimal information for protein functional evaluation (MIPFE) workshop. *New Biotechnol* 2009; **25**:170. [PubMed](#) [doi:10.1016/j.nbt.2008.12.006](https://doi.org/10.1016/j.nbt.2008.12.006)
- Orchard S, Taylor CF. Debunking minimum information myths: one hat need not fit all. *New Biotechnol* 2009; **25**:171-172. [PubMed](#) [doi:10.1016/j.nbt.2008.12.001](https://doi.org/10.1016/j.nbt.2008.12.001)
- Sherman DJ. Minimum information requirements: neither bandits in the Attic nor bats in the belfry. *New Biotechnol* 2009; **25**:173-174. [PubMed](#) [doi:10.1016/j.nbt.2008.12.002](https://doi.org/10.1016/j.nbt.2008.12.002)
- Taylor CF. Standards for reporting bioscience data: a forward look. *Drug Discov Today* 2007; **12**:527-533. [PubMed](#) [doi:10.1016/j.drudis.2007.05.006](https://doi.org/10.1016/j.drudis.2007.05.006)
- de Marco A. Reagent validation: an underestimated issue in lab practis. *J Mol Recognit* 2011; **24**:136. [PubMed](#) [doi:10.1002/jmr.1060](https://doi.org/10.1002/jmr.1060)
- Burgoon LD. The need for standards, not guidelines, in biological data reporting and sharing. *Nat Biotechnol* 2006; **24**:1369-1373. [PubMed](#) [doi:10.1038/nbt1106-1369](https://doi.org/10.1038/nbt1106-1369)
- Standardizing data. *Nat Cell Biol* 2008; **10**:1123-1124. [PubMed](#) [doi:10.1038/ncb1008-1123](https://doi.org/10.1038/ncb1008-1123)
- Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, Cavalieri D, Gaasterland T, Hingamp P, Holstege F, et al. The underlying principles of scientific publication. *Bioinformatics* 2002; **18**:1409. [PubMed](#) [doi:10.1093/bioinformatics/18.11.1409](https://doi.org/10.1093/bioinformatics/18.11.1409)
- Ball CA, Sherlock G, Brazma A. Funding high-throughput data sharing. *Nat Biotechnol* 2004; **22**:1179-1183. [PubMed](#) [doi:10.1038/nbt0904-1179](https://doi.org/10.1038/nbt0904-1179)
- Philo JS. Is any measurement method optimal for all aggregate sizes and types? *AAPS J* 2006; **8**:E564-E571. [PubMed](#) [doi:10.1208/aapsj080365](https://doi.org/10.1208/aapsj080365)
- Nominé Y, Ristriani T, Laurent C, Lefevre JF, Weiss E, Travé G. A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. *Protein Eng* 2001; **14**:297-305. [PubMed](#) [doi:10.1093/protein/14.4.297](https://doi.org/10.1093/protein/14.4.297)
- Schrödel A, de Marco A. Identification and characterization of recombinant protein aggregates. *BMC Biochem* 2005; **6**:10. [PubMed](#) [doi:10.1186/1471-2091-6-10](https://doi.org/10.1186/1471-2091-6-10)
- Martínez-Alonso M, Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A. The functional quality of soluble recombinant polypeptides produced in *Escherichia coli* is defined by a wide conformational spectrum. *Appl Environ Microbiol* 2008; **74**:7431-7433. [PubMed](#) [doi:10.1128/AEM.01446-08](https://doi.org/10.1128/AEM.01446-08)
- Ceol A, Chatr-Aryamontri A, Licata L, Cesareni G. Linking entries in protein interaction database to structured text: the FEBS Letters experiment. *FEBS Lett* 2008; **582**:1171-1177. [PubMed](#) [doi:10.1016/j.febslet.2008.02.071](https://doi.org/10.1016/j.febslet.2008.02.071)
- Minimal Information for Protein Functionality Evaluation. <http://www.mipfe.org>
- Credit where credit is overdue. *Nat Biotechnol* 2009; **27**:579. [PubMed](#) [doi:10.1038/nbt0709-579](https://doi.org/10.1038/nbt0709-579)
- Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill DP, Kania R, Schaeffer M, St Pierre S, et al. Big data: The future of biocuration. *Nature* 2008; **455**:47-50. [PubMed](#) [doi:10.1038/455047a](https://doi.org/10.1038/455047a)
- Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, Rual JF, Borick H, Braun P, Dreze M, et al. Literature-curated protein interaction datasets. *Nat Methods* 2009; **6**:39-46. [PubMed](#) [doi:10.1038/nmeth.1284](https://doi.org/10.1038/nmeth.1284)
- Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK, Jr., Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, et al. The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol* 2007; **25**:887-893. [PubMed](#) [doi:10.1038/nbt1329](https://doi.org/10.1038/nbt1329)
- Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol* 2008; **26**:889-896. [PubMed](#) [doi:10.1038/nbt.1411](https://doi.org/10.1038/nbt.1411)
- Kettner C, Field D, Sansone SA, Taylor C, Aerts J, Binns N, Blake A, Britten CM, de Marco A, Fostel J, et al. Meeting report from the second "Minimum Information about a Biological or Biomedical Investigation" (MIBBI) workshop. *Stand Genomic Sci* 2010; **3**:259-266. [PubMed](#) [doi:10.4056/sigs.147362](https://doi.org/10.4056/sigs.147362)