BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Carazo, Jose Maria

eRA COMMONS USER NAME (credential, e.g., agency login): carazog

POSITION TITLE: Full Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
University of Granada, Spain	Ms.C	1981	Physics
University Autonoma, Madrid, Spain	Ph.D	1984	Structural Biology

NOTE: The Biographical Sketch may not exceed five pages. Follow the formats and instructions below.

A. Personal Statement

I specialize in the area of Structural Biology, with a very strong additional component of data management and data analysis. In this way, I am part of two key European Research Infrastructures: INSTRUCT (for Structural Biology) and ELIXIR (for BioInformatics), effectively bridging the world of data management and analysis with the one of the elucidation of the three-dimensional structure of Biological Macromolecules, paving the way to drug design. My impact in the field can be estimated quoting my more than 8260 citations (h index 49), according to ISI Web of Science on February 18, 2020, the 3500 different users from all over the world that have downloaded our software suites XMIPP and Scipion in the last two years, the 1700 unique users of a new software 3DBionotes since March 2017 or the 7 International projects I am partner of, including an ERC Synergy grant.

I would like to highlight our strategic involvement in the development of the very much used EMDataBank (EMDB) (started from the European Union “Bioimage” project that I coordinated from 1996 to 1999), one of the most unique and undisputed resources for structural information to date. Naturally, structure has to be necessarily complemented by other approaches, a position that explains our continued effort in information integration at all levels, with 3DBionotes, our information integration software, being selected in 2019 as one of very few ELIXIR Recommended Interoperability Resources.

I know very well the area of software development both in the public and the Corporate world, and the Bioinformatics company that was spinned off from my group in 2003, called Integromics, was an example of information integration during the more than a decade long it was active, winning the first National Prize of La Caixa Emprendedor XXI 2007 and the Frost and Sullivan 2008 Prize to the Most Innovative European Bioinformatics Company, before being acquired in 2014 by the US multinational Perkin Elmer.

B. Positions and Honors

Positions and Employment

07/1981-12/1984	Pre-doctoral fellow, IBM Research Center, Madrid, Spain
01/1985- 12/1986	Post-doctoral fellow, Centro de Biología Molecular Severo Ochoa, Madrid, Spain
01/1987-12/1988	Research Affiliate II, New York State Department of Health
01/1989 -	Head of the Biocomputing Unit, National Center of Biotechnology, Madrid, Spain
03/1990- 06/2001	Tenure Scientist, National, Center of Biotechnology, Madrid, Spain
09/1998 – 10/2002	Deputy Director for Research, National Center of Biotechnology, Madrid,
11/2002 – 10/2003	Deputy Director for Research Planning and Monitoring, Science and Technology Ministry

09/1995-06/2003 Adj. Professor of Computer Science, Autonoma University, Madrid, Spain
06/2001-06/2005 Senior Research Scientist National Center of Biotechnology, Madrid, Spain

Other Experience and Professional Memberships

07/1997- 06/2001 President of the Spanish Microscopy Society.
05/2005 - Senior member of the IEEE Computer Society
01/2003-07/2014 Principal Founder of the spin-off Integromics, developing software in the Bioinformatics area and adquired by Perkin Elmer on July 2014
06/2014 - Structural Biology Grid (SBGrid), Member of the Scientific Advisory Board
01/2015 - European Synchrotron Radiation Facility, Member of the Science Advisory Committee (SAC)
11/2015 - Australian Centre of Excellence for Advanced Molecular Imaging, Member of the International Scientific Advisory Committee (ISAC)
01/2016 - National Center for Protein Science (CAS), Shanghai, Member of the Top User Program

Honors

11/1984 PhD Excellence Award by the Spanish Academy of Science.
08/1986 Okazato Research Award, by JEOL.
06/1998 Rhone-Poulenc Excellence Award by the French Academy of Science
03/2008 Frost & Sullivan Most Innovative European Bioinformatics Company of the year to our spin-off company Integromics
11/2016 Member of Faculty of 1000

C. Contribution to Science

1.-Analysis of conformational flexibility of macromolecules

Macromolecular machines carry on their functions by multiple interactions and "adaptations", often presenting a marked structural flexibility. Among all technologies capable to shed light into this complicated process, Electron Microscopy under cryogenic conditions (cryoEM) stands up offering the unique capability to directly visualize individual macromolecules down to the Angstrom scale. However, instead, this blessing often came as a curse, since the resolution achieved in these cases in the final cryoEM maps was often substantially degraded. Much of the problem originated on the fact that image processing, in particular image/map classification, was not being properly handled, motivating my research in this area through my entire career (the enclosed bibliographical selection has been chosen so that it reflects more than 20 years of accomplishments). Probably the most impacting contribution was (2), where we developed the Maximum Likelihood approach that is right now at the heart of the vast majority of current work on macromolecular architecture by cryoEM. This contribution changed totally the field. However, the very recent contribution in (1) describes a new approach to resolution with a very wide application range, from very precise map validation to model tracing. Going back almost 10 years, I would highlight the work on neural networks referred in (3), which was the first of a long series where we analyzed from the mathematical foundations to the practical applicability to the cryoEM field. Finally, I note how my first work in this area already started during by first postdoc with Dr. J.Frank, when I studied fuzzy sets techniques on ribosome images (4).

1.- Measuring local-directional resolution and local anisotropy in cryo-EM maps. Vilas JL, Tagare HD, Vargas J, **Carazo JM**, Sorzano COS. Nat Commun. 2020 Jan 2;11(1):55.
2.-Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Scheres SH, Gao H, Valle M, Herman GT, Eggermont PP, Frank J, **Carazo JM**. Nat Methods. 2007 Jan;4(1):27-9.
3.-Pattern recognition and classification of images of biological macromolecules using artificial neural networks. Marabini R, **Carazo JM**. Biophys J. 1994 Jun;66(6):1804-14
4.-Fuzzy sets-based classification of electron microscopy images of biological macromolecules with an application to ribosomal particles. **Carazo JM**, Rivera FF, Zapata EL, Radermacher M, Frank J. J Microsc. 1990 Feb;157(Pt 2):187-203.

2.-Development and provision of cryoEM digital infrastructures

Not only we need a robust mathematical formulation capable to correctly perform image and map classification, but we also need to address all the steps before and after, removing bottlenecks and creating reliable software systems. It is in this context in which we developed first XMIPP (2), and then Scipion (1), as an EM image processing suite –the former- and a workflow software integrator –the latter-, as the cryoEM field evolved. As in the case before, the bibliographic selection goes over 20 years of research. Currently we are the Image Processing Center of the European Infrastructure for Structural Biology and, as such, in charge of providing efficient access to efficient software for cryoEM image processing. Note that more than 2000 download of our software in the last 5 years.
Along the same lines, I coordinated the European project "BioImage" (4), that was the seed for the very successful EMDB (Electron Microscopy Data Base), which is "the" cryoEM database for map deposition (3)

1.- Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. de la Rosa-Trevín JM, Quintana A, Del Cano L, Zaldívar A, Foche I, Gutiérrez J, Gómez-Blanco J, Burguet-Castell J, Cuenca-Alba J, Abrishami V, Vargas J, Otón J, Sharov G, Vilas JL, Navas J, Conesa P, Kazemi M, Marabini R, Sorzano CO, **Carazo JM**. J Struct Biol. 2016 Jul;195(1):93-9
2.- Xmipp: An Image Processing Package for Electron Microscopy
Marabini R, Masegosa IM, San Martin MC, Marco S, Fernandez JJ, de la Fraga LG, Vaquerizo C, **Carazo JM**. J Struct Biol. 1996 Oct;116(1):237-40
3.-New electron microscopy database and deposition system. Tagari M, Newman R, Chagoyen M, **Carazo JM**, Henrick K. Trends Biochem Sci. 2002 Nov;27(11):589.
4.-The BioImage Database Project: organizing multidimensional biological images in an object-relational database. **Carazo JM**, Stelzer EH.J Struct Biol. 1999 Apr-May;125(2-3):97-102

3.- Soft X-rays Cellular Tomography

This área of work was motivated by the construction of the Spanish Synchrotron ALBA in the early 2000's and the subsequent Call for Ideas for new beam lines. In this way, in 2006 I presented to ALBA Scientific Advisory Board the proposal for a beam line dedicated to soft X-rays cellular cryo tomography. Indeed, the so called "water window" seemed to offer the opportunity to study whole cells preserved in cryo conditions with high contrast and a resolution in the order of the tens of nanometer. At the time, only two microscopes of this type existed in the world, at Berkely and Berlin. The proposal was accepted, and the design and construction of the microscope started. However, soon became clear that our models for image formation were falling short, and new studies were carried out to further characterize this new image modality (1), leading to a first proposal for accurate tomographic reconstruction algorithms (2 and 3) and, most recently (4) to a new image acquisition approach that correct to a large degree for the reduced depth of field associated to the diffractive optics (4). In this way, our developments in this area have totally changed the way 3D maps are being obtained, allowing the field to develop from qualitative to quantitative structural information

1.- Image formation in cellular X-ray microscopy. Oton J, Sorzano CO, Pereiro E, Cuenca-Alba J, Navarro R, Carazo JM, Marabini R. J Struct Biol. 2012 Apr;178(1):29-37
2.- The Soft x-ray transform, J. Klukowska, G. T Herman, J. Otó, R. Marabini[2] and J M.Carazo, Inverse Problems, 2014, 30(12).
3.- Characterization of transfer function, resolution and depth of field of a soft X-ray microscope applied to tomography enhancement by Wiener deconvolution, J. Oton, E.Pereiro, A.J. Berna, L. Millach, C.O.S.Sorzano, R.Marabini and J.M.Carazo, BioMedical Optics Express, in press
4.- XTEND: Extending the depth of field in soft X-ray tomograph, J. Oton, E.Pereiro, A.J.Perez-Berna, L.Millach, C.O.S Sorzano, R. Marabini and J.M. Carazo, under review.

4-Structural analysis of DNA replicative helicases

In this case, rather than focusing into methods and infrastructures, I refer to the biological system that I chose to study after my postdoc: DNA replicative helicases. Not only it was a fascinating research topic, but also one characterized by the occurrence of large structural flexibility, rendering it intractable (at the time) for X rays crystallography. I chose two specimens: DnaB, for a bacterial model, and SV40 Large T Antigen, for eukaryotes. As for DnaB, we were able to characterize part of its dynamical stages, as well as the "freezing" it occurs upon interaction with its loading partner DnaC (1, 2). However, probably our most impacting research happened on LTA (3, 4), with special emphasis in (4), where we provided the first

average images on the system together with Ab labeling, allowing us to propose a detailed molecular architecture of the system. Note that I have always been very aware that my expertise was not so much on detailed biological systems, but on their approach by cryoEM, so that during the almost 20 years that this area was a major topic of my research I always worked with molecular biologists collaborators with large experience in the field (my current activity in this field is reduced).

1-The DnaB.DnaC complex: a structure based on dimers assembled around an occluded channel. Bárcena M, Ruiz T, Donate LE, Brown SE, Dixon NE, Radermacher M, **Carazo JM**. EMBO J. 2001,15;20(6):1462-8.
2.-Three-dimensional reconstructions from cryoelectron microscopy images reveal an intimate complex between helicase DnaB and its loading partner DnaC. San Martin C, Radermacher M, Wolpensinger B, Engel A, Miles CS, Dixon NE, **Carazo JM**. Structure. 1998 Apr 15;6(4):501-9.
3.-Conformational rearrangements of SV40 large T antigen during early replication events.
Cuesta I, Núñez-Ramírez R, Scheres SH, Gai D, Chen XS, Fanning E, **Carazo JM**. J Mol Biol. 2010 Apr 16;397(5):1276-86
4.-Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. Valle M, Gruss C, Halmer L, **Carazo JM**, Donate LE. Mol Cell Biol. 2000 Jan;20(1):34-41.

5.-Data analysis in Bioinformatics

Armed with my strong bases on pattern recognition and large data sets mining, for more than twenty years I have approached the field of information integration, trying to find ways to enrich structural information. A recent and representative approach is the one presented in (4) –now accessible from the EM Data Bank at the European Bioinformatics Institute- where we develop an interactive framework to integrate multiple data sources into a structural context. Still, in earlier times my research explored other areas in the realm of genomics and proteomics (1-4), always using machine learning approaches to derive new biomedical knowledge, Note that this field of work is behind the launch in 2003 of a bioinformatics company (Integromics), that in 2008 won a European prize to the most innovative product and that in 2014 became a PerKin-Elmer company.

1.-Self-organizing tree-growing network for the classification of protein sequences. Wang HC, Dopazo J, de la Fraga LG, Zhu YP, **Carazo JM**. Protein Sci. 1998 Dec;7(12):2613-22.
2.-GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Carmona-Saez P, Chagoyen M, Tirado F, **Carazo JM**, Pascual-Montano A. Genome Biol. 2007;8(1):R3.
3.-ChIPCodis: mining complex regulatory systems in yeast by concurrent enrichment analysis of chip-on-chip data. Abascal F, Carmona-Saez P, **Carazo JM**, Pascual-Montano A. Bioinformatics. 2008 May 1;24(9):1208-9
4- 3DBIONOTES: A unified, enriched and interactive view of macromolecular information. Tabas-Madrid D, Segura J, Sanchez-Garcia R, Cuenca-Alba J, Sorzano CO, **Carazo JM**. J Struct Biol. 2016 May;194(2):231-4

D. Research Support

Ongoing Research Support

UE(810057):H2020 - ERC-2018-SyG 01/03/2019 – 28/02/2024
HighResCells: A synergistic approach toward understanding receptor signaling in the cell at very high resolution
Role: PI

UE(824087): H2020-INFRAEOSC-2018-2020 01/03/2019 – 28/02/2023
EOSC-Life: Providing an open collaborative space for digital biology in Europe.
Role: PI

UE (857203): H2020-WIDESPREAD-2018-03 01/09/2019 – 31/08/2022
IMPACT: Imaging life from Molecules to cells - building knowledge on Cryo-electron microscopy methodologies
Role: PI

UE (871037): H2020-INFRAIA-2018-2020 01/02/2020 – 31/01/2024
iNEXT-Discovery: Infrastructure for transnational access and discovery in structural biology

Its aim is the development of new ways of access structural biology (SB) infrastructures, and to implement and provide access
Role: PI

UE (654248): H2020- INFRADEV-1-2014-1 01/09/2015 – 31/005/2020
CORBEL: Coordinated Research Infrastructures Building Enduring Life-science services
Its goal is the generation of integrative approaches among Biomedical infrastructures
Role: PI

UE (731005): H2020-INFRADEV-2016-2017 01/01/2017 – 31/12/2020
INSTRUCT – ULTRA: Releasing the full potential of Instruct to expand and consolidate infrastructure services for integrated structural life science research
Role: PI

CAM S2017/BMD-3817 01/01/2018 - 31/12/2021
TomoXLiver-CM: TomoXliver: estudio de la disfunción del hepatocito desde un abordaje multidisciplinar
Role: Coordinator of seven research groups & PI

ISCIII: PT17/0009/0010 01/01/2018 – 31/12/2021
Instituto Nacional de Bioinformática (ISCIII)
This is a network-type of projects linking biomedical Platforms in Spain
Role. PI

Completed Research Support

UE (675858): H2020- EINFRA-2015-1 01/11/2015 – 31/10/2018
West life: World-wide E-infrastructure for structural biology
Its goal is the development and deployment of a Virtual Research Environment for SB
Role: PI

UE (654142): H2020-EINFRA-2014-2 01/03/2015 – 31/08/2017
EGI-Engage: Engaging the EGI Community towards an Open Science Commons
Its aim (as applied to our group) is the generation of Cloud cryoEM image processing pipelines. Role: PI

UE(676559): H2020-INFRADEV-3-2015 01/09/2015 – 31/08/2019
ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life-sciences.
It aims at generate hybrid approaches between SB and other disciplines
Role: PI

MINECO: BIO2016-76400-R 30/12/2016 – 29/12/2019
FLEX3D: Análisis en alto rendimiento de la flexibilidad estructural tanto por crio microscopia electrónica como por crio microscopia de rayos X blandos
This is the basal founding of the lab at the National level
Role: PI

ISCIII: PT13/0001/0009. 01/01/2014 – 31/12/2018
Plataforma de recursos biomoleculares y bioinformáticos, PRB2
This is a network-type of projects linking biomedical Platforms in Spain
Role. PI