European Cooperation in the field of Scientific and Technical Research
- COST -

Brussels, 14 November 2014

COST 083/14

MEMORANDUM OF UNDERSTANDING

Subject :	Memorandum of Understanding for the implementation of a European Concerted Research Action designated as COST Action BM1405: Non-globular proteins - from sequence to structure, function and application in molecular physiopathology (NGP-NET)

Delegations will find attached the Memorandum of Understanding for COST Action BM1405 as approved by the COST Committee of Senior Officials (CSO) at its 191th meeting on 12-13 November 2014.

MEMORANDUM OF UNDERSTANDING
For the implementation of a European Concerted Research Action designated as

COST Action BM1405
NON-GLOBULAR PROTEINS: FROM SEQUENCE TO STRUCTURE, FUNCTION AND APPLICATION IN MOLECULAR PHYSIOPATHOLOGY (NGP-NET)

The Parties to this Memorandum of Understanding, declaring their common intention to participate in the concerted Action referred to above and described in the technical Annex to the Memorandum, have reached the following understanding:

1. The Action will be carried out in accordance with the provisions of document COST 4114/13 "COST Action Management" and document 4112/13 "Rules for Participation in and Implementation of COST Activities", or in any new document amending or replacing them, the contents of which the Parties are fully aware of.

2. The aim of the Action is to create an effective European network, seeking synergies to improve understanding of non-globular proteins (NGPs). A consensus among experts will provide guidelines for the classification of different NGP phenomena.

3. The economic dimension of the activities carried out under the Action has been estimated, on the basis of information available during the planning of the Action, at EUR 80 million in 2014 prices.

4. The Memorandum of Understanding will take effect on being accepted by at least five Parties.

5. The Memorandum of Understanding will remain in force for a period of 4 years, calculated from the date of the first meeting of the Management Committee, unless the duration of the Action is modified according to the provisions of Section *2. Changes to a COST Action* in the document COST 4114/13.

A. ABSTRACT AND KEYWORDS

Non-globular proteins (NGPs) encompass different molecular phenomena that defy the traditional sequence-structure-function paradigm. NGPs include intrinsically disordered regions, tandem repeats, aggregating domains, low-complexity sequences and transmembrane domains. Although growing evidence suggests that NGPs are central to many human diseases, functional annotation is very limited. It was recently estimated that close to 40% of all residues in the human proteome lack functional annotation and many of these are NGPs. While a better understanding of NGPs is crucial to fully comprehend human molecular physiopathology, progress has been hampered so far by the lack of a systematic approach to their study.

This Action Proposal aims to create a pan-European scientific network of groups that work on NGPs to strengthen, focus and coordinate research in this field. It proposes to develop a novel classification of NGPs by consensus among interested experts that will be showcased on a newly developed web site, along with meetings, training schools and scientific missions on NGP-related topics.

Keywords: Computational biology, proteome, intrinsic disorder, tandem repeats, amyloid aggregation,

B. BACKGROUND

B.1 General background

Protein structure determination, employing techniques such as NMR and X-ray crystallography, is one of the centrepieces of biological research across Europe. Protein function has been traditionally derived in terms of unique 3D structure, where the fold of a globular protein determines its function. The classic example would be enzymes, autonomously folding domains of specific function. Their rigid structure combined with limited flexibility in the active site is crucial for catalytic function. Nowadays, more than 90,000 different crystal structures are available for the scientific community. Ironically, during the last 10 years structural research suggests that a well defined globular structure is only one of the different forms in which a protein can exist. Recently, there has been a growing interest in a number of different phenomena, such as intrinsically disordered proteins (IDPs), repeat proteins, amyloid aggregation, low complexity regions and transmembrane domains. All together

these unconventional phenomena are known as non-globular proteins (NGPs). NGPs have been linked to several human diseases sharing an unclear physiopathology and high social cost, such as the “mad cow” (Creutzfeldt–Jakob) disease, neurodegeneration and cancer.
NGPs have been an emerging concept in protein science for the last decade, with many publications highlighting their ubiquitous nature especially for higher organisms. In particular, NGPs are important for eukaryotic regulatory and signalling processes, interaction hubs and scaffold elements. The identification of many NGPs also enabled the development of sophisticated bioinformatics algorithms for their prediction from sequence, e.g. disorder and aggregation propensity, which further advanced the field. Currently, different research groups are investigating the peculiar properties that arise from unconventional non-globular folds and disordered protein regions. Rapidly, many groups in Europe developed tools for predicting and analysing the different aspects connected with NGP structure and biology, very often taking a leading role in the field. However, the same groups are working independently producing new results based on purely (or principally) local knowledge.
However, in order to expand and consolidate the excellence of European research on NGPs, large and inter-operative teams are needed. This Action’s principal aim is to join the efforts of many of these different local European entities into a large, solid research network. Most NGPs are well known for their intrinsic dynamic properties and the same applies to European researchers. The opportunity provided by COST to share pre-existing know-how and to develop an unprecedented interchange of ideas will benefit this young and dynamic research area. This Action will combine expertises ranging from bioinformatics, molecular biology, and genetics to pharmaceutical companies. The knowledge to be developed in this Action will be at the forefront of our efforts to understand NGP-related molecular processes e.g. cancer development and degenerative diseases. As local leaders in their respective research fields, the participants have their own research funding, sometimes as part of interdisciplinary collaborations with other scientists but often limited to national borders. COST will be critical to increase coordination of this Europe-wide effort.

B.2 Current state of knowledge

While the widely accepted protein structure–function paradigm has dominated protein biology for more than 100 years, an ever-increasing number of non-globular but biologically active proteins is being discovered and, in some cases, functionally characterised. Only around the early years of the

new millennium was it formally raised in several conceptual papers that many proteins, or regions of proteins, can exist in an alternative non-globular form. Intrinsically disordered proteins, repeats and amyloids under native, functional conditions pointed out the limits of the structure–function paradigm. NGPs, initially only addressed as a specialization of a few exotic proteins, were connected with several human diseases and important cellular functions. Starting from a dozen initially classified non-globular proteins, thousands of PDB (Protein Data Bank) structures are nowadays known to contain disordered (intriguingly sometimes also repeated) chains that become structured only in presence of an interaction partner. Many regions missing from electron density maps also complete the scenario, reinforcing the idea that the structure–function paradigm is only one of several forms in which a protein can exist. NGPs are characterized by peculiar properties. It was demonstrated that IDPs provide a larger interaction surface area than globular proteins of a similar length. Due to their conformational flexibility coupled with the tendency to expose short linear peptides (i.e. functional motifs), IDPs interact with numerous other proteins providing plasticity and regulation of protein activity. IDPs are also more prone to undergo post-translational modifications, facilitating functional regulation (i.e. cell cycle regulation). Another NGP class, commonly present in higher organisms, are protein tandem repeat sequences, coming in various flavours. From single amino acid homorepeats, e.g. poly-Q, to cryptic repeats of up to forty residues per unit and even entire domains, there is a plethora of shapes and functions. Simple sequence repeats can be detected for their lack of sequence complexity. Homorepeats of some residues are more frequent than others and vary between proteomes. At the other extreme, long cryptic repeats can be classified into solenoid, toroid and single domain repeats. Solenoids, e.g. LRR or Armadillo repeats, form open and extended structures requiring the presence of several units to stabilize the fold. Toroids, e.g. beta-propellers, are composed of repeats forming closed circular structures. Single domain repeats, e.g. fibronectin domains, behave like “chains of pearls” where the single domain is large enough to be detected with classic similarity searches. A third phenomenon linked to NGP regions is the tendency of certain sequence stretches to misfold and self-assemble into cytotoxic aggregates. While some weak indications are in the literature about the overlap between disorder, repeats and aggregation, so far there has been no attempt to grapple with this problem in a coherent framework. Very recently, it was found that out of the 38% of human proteome residues with no Pfam domain annotation, an important fraction exhibits compositional bias. This stresses the importance to develop improved methods for accurate classification of NGP regions to fully and accurately annotate the human proteome. Many human diseases involve poorly characterized NGPs.

Understanding the functional NGP repertoire can yield molecular hypotheses explaining the effect of mutations in very different conditions such as e.g. cancer, cardiovascular disease, Parkinson's and epilepsy. The gain in basic knowledge from the Action Proposal can contribute to research in several biomedical problems constituting major health challenges.
Research on NGP phenomena is hampered by the difficulty to address these proteins with suitable experimental techniques. NMR and X-ray crystallography, although essential for early-stage NGP identification, have their technical limits. Participants in this action have provided progress by developing computational methodologies for NGP identification and analysis as well as large and detailed databases of different NGP flavours.

B.3 Reasons for the Action

This Action is aimed at both scientific/technological consolidation of the current European research leadership in the field of NGPs and preparing the ground for exporting valuable outcomes at the European economic/societal level. NGPs have great potential in the medical, biotech and pharmaceutical industries due to their implication in human health. However, translating this potential into valuable tools for the community requires a cooperative effort between all European players. The classic example would be the definition of NGPs. A plethora of different names are currently used to define similar or even the same phenomena. At a larger scale, a similar multiplication of tools, methodologies and applications are surfacing in the field. The technology platform which this Action intends to develop is a highly connected community around NGPs with the final aim to disseminate the derived benefits to the European countries and over. It also aims to enlarge our understanding of biomolecular mechanisms driving NGPs and to improve collaboration with the structural biology community as well as biotech and pharma industries. There is also an urgent need for integration of approaches, e.g. complex modeling algorithms to model proteins and protein complexes based on different experimental data. A key objective is to provide a common platform for collaborations useful for consolidation of expertise and know-how as well as training of the next generation of users. Knowledge merging will speed up the development of methodology, new tools and will provide a solid opportunity for academia to work with the health-care industry.
In summary, this Action will provide several productive outcomes, e.g.:

- Fostering a better understanding of overlapping NGP phenomena and joint consensus classification.

- Developing guidelines in the NGP field by a coordinated consensus, including experts and fostering scientific discussion.
- Creation of an effective European network, seeking synergy between participants for joint activities and enabling the pooling of a large number of fragmented resources.
- Dissemination to the scientific community of best practices and stimulating a wider scientific discussion on NGPs, while asserting European leadership.

B.4 Complementarity with other research programmes

The collaboration with ELIXIR, the European sustainable bioinformatics infrastructure initiative, and INSTRUCT, the structural biology initiative will serve to discuss the best ways to ensure the long-term availability of NGP-related information beyond this COST Action. While NPG research may be adequately funded at the national level, there is a clear lack for alignment and coordination of the various activities across Europe. Particular emphasis on knowledge dissemination and training (student/staff exchange) will be posed. Complementary and valuable links can be established to the following COST Actions: BM1403; BM1203; CM1207; CM1402. All of these are interested in specific systems, whereas this Action focuses on the underlying NGP phenomena.

C. OBJECTIVES AND BENEFITS

C.1 Aim

The main aim of the Action is to create an effective European network, seeking synergy between participants towards joint activities, enabling the pooling of otherwise fragmented resources and uniting the voices of a large number of European labs, in the better understanding of overlapping NGP phenomena and their joint consensus classification. Various researchers will join forces to develop guidelines in the NGP field by a coordinated consensus approach, including experts in the field, and fostering scientific discussion on the subject. Dissemination to the scientific community of best practices and stimulating a wider scientific discussion on NGP phenomena will thus enhance the quality, translation and impact of European research in a highly competitive and relevant area of biology. Although the Action covers a large number of problems at different levels (as structured in the WGs), they all carry as their central theme the problems that come from extensive fragmented

resources and conflictual definitions in this field. The network will allow participants to share data and experiences, as well as modeling and solving issues in this competitive and challenging area.

C.2 Objectives

The primary scientific objective of this Action is to foster the creation of a European network of experts aiming to better understand the role of NGPs through the development of widely accepted definitions and functional annotations of each specific phenomenon. Specifically:

- To define and develop guidelines, best practices and critical assessments of methods for NGPs in agreement among researchers by a coordinated consensus approach. Understanding this phenomenon will enhance our comprehension of the complex relation between structural mobility or aggregation and function.

- To evaluate the quality and performance of existing NGP prediction methods and lead the discussion among the integration of different approaches and methodologies, also providing measurable outcomes for quality assessment. A real improvement in this sector can be achieved only with a fruitful collaboration between experimentalists and computational researchers, by merging their complementary expertise to produce and develop novel strategies for the sequence-structure-function analysis of NGPs.

- The Action will support young European researchers by provision of Short Term Scientific Missions (STSMs) and organizing Scientific Meetings, Workshops, Conferences and Training Schools aiming to train a new generation of scientists in an interdisciplinary way. Enhancing early stage research activity could drive the researcher to become more productive. Moreover, the Action will promote gender balance at all levels of its organization.

- To boost European research efficiency and quality by disseminating the Action's results in international conferences and by publishing its work in peer-reviewed journals. The Action will create and promote a web-portal in order to share state-of-the-art tools, methods, databases and papers as well as experience and skills. The aim of such a portal is to create a central repository that can overcome dispersion, a drawback which currently results in a plethora of different definitions for the same phenomenon.

- To establish and extend state-of-the-art methods to evaluate predictors and encourage the establishment of comparative assessments in analogy to the successful CASP (Critical Assessment of Structure Prediction) and CAFA (Critical Assessment of Functional Annotation).

- To establish a multidisciplinary research network that will lead a cooperative action to develop, prepare and submit a proposal for Horizon2020 and other joint EC research initiatives.

C.3 How networking within the Action will yield the objectives?

This Action will be the first forum where NGPs will be discussed in a large community. To reach a definition and guidelines for a wide and complex phenomenon like NGPs, synergistic collaboration is fundamental to prevent tool duplications and quality measurements. The results could address fundamental improvements not only for the scientific community, but also for pharmaceutical companies and patients with NGP related diseases. More specifically:

- The Action objectives logically lead to a series of Work Group (WG) activities. Four WGs will be established (see D.2 below).

- WG coordinators and Committees, and details of the management strategy will be defined on 1st MC Meeting.

- A budget will be made available within the COST Action to fund a large number of workshops and training, both for partner education and for end-user training. All educational material will be made freely available through the dissemination portal.

- The use of Short-Term Scientific Missions (STSM) to meet specific challenges and to provide a vehicle to support the training of early stage researchers. Specific attention will be posed on young researchers for a comprehensive and multidisciplinary career development. Every year the Action will open a competitive call to ensure meritocratic access to Action funds.

- Every tool developed within the Action that can be fruitful for the community will be open software and hosted in the Action's web-portal together with a comparative and qualitative analysis

- Final conference at the end of the Action where all the partners present their results, arranged to maximize dissemination of Action knowledge.

C.4 Potential impact of the Action

Supporting the establishment of a strong European network in a current topic like NGPs will open up new research fields by enhancing efficiency and focus of research. Increasing understanding of NGPs will have tremendous benefits for many biomedical researchers working in more applied fields, providing molecular insights for inexplicable phenomena in human physiopathology.
The Action benefits will lead to:

- A better understanding this phenomenon can increase the number of novel, unconventional targets for pharmaceutical companies as well as, in case of success, a long term benefit on the public health. As a minimum impact, the identification process of novel protein targets for NGP diseases will be accelerated, e.g. Parkinson’s or Alzheimer’s, Epilepsy, cardiovascular diseases and various cancers.

- The development and implementation of standardized definitions for the various NGP topics and dissemination to the scientific community will be technically beneficial for every future improvements and references.

- Scientifically, this Action will boost research in the field, avoiding duplication of methods and tools, aiming to maximize the efficiency of related scientific projects. Moreover, this Action will create novel opportunities for the establishment of innovative spin-off companies.

- It can be roughly estimated, considering the participating research groups, that this Action will produce ca. 40-60 new PhD positions in the NGP field.

- The Action will improve the overall quality of European research in the NGP field. This will provide worldwide visibility of the researchers involved in the Action. Furthermore, this will enlarge the slice of Europe based knowledge in the field of age-related and neurodegenerative diseases connected with NGP phenomena.

- The action will create a collaborative and fruitful environment, definitely improving creativity and providing ideas, which will be then converted into European grant proposals.

C.5 Target groups/end users

The target groups for this Action are:

- **European researchers** as well as the entire academic community, working towards answers to biomedical questions, who will gain access to the right tools, and to newly emerging applications. Bioinformaticians that develop new algorithms and make these accessible by user-friendly software tools. Documents, guidelines and definitions produced as results of expert discussion and debate during this Action will be of importance to the researchers in this field. It will allow moving towards a set of standardized conditions for experimentation and comparison between results obtained from different labs. This will be provided by continuous quality assessment of the existing as well as newly developed tools, thereby minimizing waste of European funds by preventing duplication of tools by different laboratories. The European research community will gain direct access information via the Action's Web-portal, providing a fast growth on the subject.

- **Early Stage Researchers (ESRs)** interested in learning about NGPs, as primarily involved in the research covered by this Action. They will be provided with the possibility of participating in a comprehensive set of formative and networking activities, promoting the development and strengthening what will be the future research and academic community.

- **Patients** suffering from diseases directly or indirectly associated with NGPs, as well as physicians involved in disease management, expanding the comprehension of these phenomena.

- **Pharma and biotech industries**, which will be provided the basic background and knowledge on this subject. This will outline new research branches, extending the potential target user base for the development of new chemical entities, and helping in exiting the current standstill scenario for drug targets.

- **European society**, eventually, will be the largest beneficiary of this Action proposal, as the mean age of the population is increasing and this Action will provide new insights in the understanding of age-dependent and neurodegenerative diseases, reducing the cost for the health-care system.

D. SCIENTIFIC PROGRAMME

D.1 Scientific focus

The **scientific program** is focused on 3 research areas: intrinsic disorder, tandem repeats and aggregation, with the following main tasks:

- Initial state-of-the-art report, framing the NGP problem with available information

- Establishing a widely accepted definition and classification of NGP phenomena

- Development of novel approaches for detection and functional annotation of NGP types

The scientific programme is focused on three research areas that exhibit strong links and interrelationships:

1) Intrinsic Disorder

Intrinsic disorder may be considered one of the challenging paradoxes of the last decade as the mainstream belief that protein function implies structure is turning into a more complex scenario. Plenty of question marks are, at time of writing, surrounding the rising number of available protein structures in the PDB with missing electron density for some residues. The same questions are in turn arising from particular protein purification experiments and high proteolysis sensitivity. Protein

flexibility is an emerging phenomenon comprehensively answering these questions, enlightening the “dark side” of structural biology. Through the past years, the terminology seems to have settled on “intrinsically disordered” for proteins/regions that exist as dynamic ensembles. An entire class of polypeptides lacking rigid structures but possessing biological function was under-appreciated and ignored for a very long time despite numerous examples scattered in the literature. Growing interest towards these proteins is impacting current protein research and yielding the development of various sequence-based predictors, facilitating functional annotation of these proteins. These prediction methods are based on different concepts, providing results from different types of analysis. Listing all the discoveries and findings made in this field is beyond the scope of this description, although the amount of data generated during the past 15 years on specific features related to the structural properties of IDPs, their abundance, distribution, functional repertoire, regulation, involvement in disease pathogenesis, and so forth, is vast and seems to be just the tip of the iceberg. The subject is producing almost daily results and even more important breakthroughs are expected over the next few years. New models explaining various functions of IDPs, their evolution, and involvement in diseases are in great demand, together with a general theory updating current knowledge on protein structure and function, and with novel experimental and computational tools for investigations focused on IDPs. Low complexity sequences (e.g. polyQ) will also be considered in this category.

2) Repeat Proteins

Proteins showing repetitive motifs such as structural or functional units are called repeat proteins. Many techniques have been developed to identify repetitions in protein sequences, exploiting different approaches. These approaches are generally working on sequences, but the identification of repeats can be demanding due to repeat units often being degenerated. The main reason for this barrier is the fact that a consistent amount of point mutations do not affect repetitive structures. Several repeat structure families have been studied and classified so far. They are relevant for protein engineering and connections to human diseases, such as neurodevelopment issues to name a few. One of the main fields concerning repeat proteins is the existence, and thus detection, of other common structures that may be undetected at the moment. The most common way to detect and classify repeat proteins is, so far, sequence annotation followed by visual inspection of the structure, and as the PDB and sequence databases are massive, the systematic description of repeats is becoming a question of using automated methods to detect them. This field is relatively new and the available methods are still few. Transmembrane proteins may also be grouped in this category as

they are often repetitive in structure (e.g. porin beta barrels).

3) Aggregation

A broad range of human diseases arise from the failure of a specific peptide or protein to adopt, or remain in, its native functional conformational state. These pathological conditions are generally referred to as protein misfolding diseases. In many cases, misfolding of the wild type protein is associated with a late disease onset, whereas pathogenic familial variants, often single mutants, cause an early onset and more severe symptoms. The largest group of misfolding diseases is associated with the conversion from a soluble functional form to highly organized fibrillar aggregates, generally described as amyloid fibrils. Amyloid fibrils are often related to a plethora of human diseases, such as Alzheimer syndrome, Parkinson's disease, the “mad cow” disease (Creutzfeldt–Jakob disease), and more in general neurodegeneration and cancer. One hallmark of the amyloid structure is a specific supramolecular architecture called cross-beta structure, held together by hydrogen bonds extending repeatedly along the fibril axis, exhibiting non-globular characteristics.

D.2 Scientific work plan methods and means

WG1 **“NGP Classification and Coordination”**

- **Availability.** A dedicated website/portal will be established to provide information about the COST Action as well as to document its progress and the achievement of its objectives. This will also serve as a vehicle for discussion and development of ideas and as a primary tool, providing best practice guidelines and recommendations for improvement and quality of the obtained results. The website will provide access to state-of-the-art scientific literature, algorithms and software tools and their application.

- **Knowledge**. To expand knowledge on NGPs, in particular functional annotations for overlapping subtypes. A better characterization of NGPs and their comparison will provide interesting functional conclusions about proteins, as well as annotations for thousands of sequences. Knowing the relationship between different NGP phenomena can help to explain the functional mechanism of complex proteins, which are highly relevant in many signalling and regulatory processes.

- **Consensus classification of NGPs.** To create consensus for a joint definition and classification of NGPs. To date, there is only sketchy information on distinct NGP classes. From a preliminary analysis, it appears that thousands of proteins in the human proteome are under-annotated. A well-defined classification scheme, developed with a consensus approach, will enable a thorough description of the different classes and open up further areas of research.

- **International synergy**. Seeking synergies with ELIXIR to maintain a sustainable NGP bioinformatics infrastructure in Europe. The chosen topic is unique in Europe, with no directly competing initiatives. Given the relevance of NGPs, it will be highly relevant to expand ELIXIR services in this area.

- **Continuous Assessment.** To coordinate the comparison of newly developed methodologies with those already existing by the other WGs, providing common guidelines and recommendations. All newly implemented methodologies will be compared with state-of-the-art protocols, in order to assess the quality of the improvements achieved.

- **Scientific workshops/conferences.** Organizing every year a scientific workshop/conference with the other WGs, in order to share the amount of generated knowledge and facilitate and strengthen cooperation.

WG2 "**Intrinsic Disorder"**

- **Definition of IDPs.** Taking the existing literature as a starting point, rules for the IDP and low complexity definition will be updated and applied iteratively as soon as new knowledge will be generated. Specific efforts are planned to generate a robust definition of IDPs. In particular, a dedicated sub-section will address the interpretation of low complexity and folding upon binding phenomena. These two transiently structured states are extremely important for protein-protein interaction, representing the way cells have to improve protein activation. The expected result of this task will shed light on this still relatively unclear subject.

- **Assessment of existing methods.** Existing methodologies result extremely powerful for interpreting single cases but coherence between methods is still lacking when explicitly compared. Thus, all existing methods for ID and low complexity prediction will be evaluated and compared, to provide guidelines for IDP investigation. This will also serve to provide suggestions for the improvement of already existing tools, and to reach a consensus between them.

- **Improvement of existing methods.** The consensus will provide a quality standard, which will then be used as a basis for the improvement of disorder and low complexity prediction methods towards more thorough analysis and predictions. The final aim is to move the pre-existing predictive tools from theory to the real world. Reliable methodologies will serve as *de facto* standard for IDP predictions.

- **Defining roles of IDPs.** As a consequence of the new definitions provided by point 1 and the assessment provided by point 2, existing protein sequences will be evaluated, providing new insights into the different roles of IDPs in NGP related diseases. The most relevant consequence will be a better comprehension of molecular details driving disease outcome, thus helping health-care companies in developing new specialized therapeutics.

- **Yearly meeting.** To organize every year a scientific workshop/conference for the work group. This will allow knowledge sharing between different group members, and will at the end be integrated with the generated knowledge of the whole Action community. A specific task will curate an easy impact communication for a non-prepared audience, allowing development of an informed social conscience. This will help to export knowledge outside the scientific community and advertising the active role of EU in science.

WG3 "**Repeats"**

- **Improvement of Repeat classification.** Starting from up-to-date literature and structures, the existing classification will be critically assessed, refined and tuned. Repeats are a large family of proteins presenting many different shapes. Borrowing a

term from botanic science, during this Action as many dichotomic keys as necessary for each repeat flavour will be formulated.

- **Automation of the classification process.** Algorithms and machine learning rules for repeat classification will be developed and assessed. The classification will move towards automatic structure based characteristics, which will be implemented for sequence based classification. The aim is to obtain an algorithm able to classify repeat proteins at the accuracy level of manual curation.

- **PDB structure updates.** All newly published PDB repeat structures and TM regions will be automatically classified and the classification made accessible for the scientific community.

- **Yearly meeting.** To organize every year a scientific workshop/conference for the work group. This will allow knowledge sharing between different group members, and will be integrated with the generated knowledge of the whole Action community. Repeat proteins are of interest for scientists of different disciplines e.g. neurologists and experimental biologists. Specific sessions focusing on relevant medical aspects will be planned.

WG4 "**Aggregation"**

- **Pathological vs. constitutive aggregation.** Taking the existing literature as a starting point, the difference between different protein aggregates will be discussed. Meetings between the experimental and computational sides to discuss current issues in aggregation. For example, the lack of experimental data to validate computational approaches. This will hopefully promote a quality standard for the data. A better aggregation definition is expected in this task.

- **Improvement of existing methods.** The computational techniques used to analyze fibrils are diverse and of varying performance. The Action would like to discuss how to create a standard to evaluate all the tools. How to combine them to achieve the most accurate technique. The idea is to stimulate the bioinformatics community to develop useful and reliable methods for medical applications.

- **Long term community organization.** The task will promote the idea of a central repository for all the tools, databases and experimental data related to human health. Such diseases can be broadly grouped into neurodegenerative conditions, in which fibrillar aggregation occurs in the brain, non-neuropathic localised amyloidoses, in which aggregation occurs in a single type of tissue other than the brain, and non-neuropathic systemic amyloidoses, in which aggregation occurs in multiple tissues.

E. ORGANISATION

E.1 Coordination and organisation

The following organisational structures will be implemented:

Management Committee (MC)

A MC, convened by representatives of the participating countries as described in the COST framework guidelines, will coordinate the scientific and administrative actions. At its first (kick-off) meeting, the Chair, Vice-Chair as well as STSM and website coordinators will be elected by majority vote. The External Advisory Board and leaders and deputy leaders for each WG will also be appointed. The MC also has the following responsibilities and authorities: promotion of gender balance and ESR involvement; distribution of funds within the Action; draft scientific reports to COST; launch of calls for exchange visits; increase of visibility by publication of activities; recruit new members; organisation, planning and coordination of meetings, conferences, workshops and training schools; monitoring of scientific progress within WGs; establishment of a panel that reviews STSM applications; promoting and establishment of collaborations between Action members with other COST Actions, European and international programmes; promoting communication with societies; support of interactions with industrial partners, commercial exploitation and translation of research results.

Steering committee (SC)

The Action will establish a SC composed of the Chair and Vice Chair of the MC, the WG leaders, STSM and Website coordinators. This SC will facilitate efficient communication and decision-making by preparing MC meetings and coordination of Action activities. The SC will meet every month through internet-meetings. The SC will also monitor gender balance.

WG meetings – see also E.2

The Working Groups (WGs) will also meet during MC meetings and discuss the progress and

coordination of their specific WG objectives. WG leaders will report the progress of their respective WG based on the predefined timetable to the MC. The WG leaders are responsible for achieving goals and milestones of each WG and for reporting their achievements to the MC.

MC and WG meetings

The Management Committee will convene yearly to ensure efficient coordination, to evaluate progress and make specific plans for future activities. These meetings will, with the exception of the kick-off, coincide with WG meetings. They will take place at different locations, reflecting the geographical distribution of current and future members of the Action. Efforts will be made to align MC/WG meetings with conferences in the field (ISMB, ECCB, EMBnet Conferences, etc.), to increase the visibility of the Action, attract more participants, and save travel costs. Their duration is expected to be 2 full days. To ensure efficiency in meeting WG needs, promote communication and achieve the maximum possible exchange of information among the WGs, the meetings will include day-long WG-specific sessions and plenary sessions involving representatives of all WGs. These will include presentations of individual research results, and discussions on future developments in the field. Emphasis will be placed on the inclusion of young scientists and women in these activities (oral presentations, chairmanship of sessions, etc.). Based on these meetings, the MC will evaluate the progress of the Action towards reaching its objectives, and will decide on future plans accordingly.

Monitoring and evaluation

The Action will closely monitor its progress and activities by evaluation forms handed out after each MC and WG meeting as well as at training schools. The outcomes of this internal evaluation will be available on the website and an agenda item at the next respective meeting. Each STSM will conclude with a report generated by the guest and the host for the MC and WG. These reports will allow to improve the outcomes of STSM.

Website

A dedicated website/portal will be established to provide information about the COST Action as well as to document its progress and the achievement of its objectives. The website coordinator will be responsible for its establishment and maintenance (monthly update). It will be separated into a freely accessible domain and a password protected intranet. The website will entail information about the Action in general and its progress towards reaching its aims. It will list information on and links to the field of NGPs. One major goal is to communicate novel scientific insights to the broader scientific community and society in general, thus raising interest and awareness about NGPs.

Within the intranet, activities, decisions and documents generated by the MC and WGs will be uploaded. In addition, dedicated mailing lists will be established to keep different MC members informed regarding the progress of the COST Action and as a means of communication between different MC and WG members. A key feature of the website will be the development of a dedicated Wiki that participants can utilize to develop reports and whitepapers based on the outcome of workshops and discussions between participants as well as share experience and best practice through FAQs and blogs.

Scientific Conference

The Action will organise an international scientific conference at the end of the Action where the findings will be presented and discussed with international experts from academia, industry and other stakeholders.

Workshops

The planning of workshops in the different WGs will be such to stimulate exchange of information and ideas among WG. That is, workshops should be organized at the same time and sufficient time between consecutive workshops within each WG is required to build upon acquired knowledge and put forward new ideas at the next workshop.

Training Schools (TS)

Dedicated TSs, each devoted to a WG specific topic will give scientists a multi-disciplinary and international platform to access the expertise of internationally leading researchers. Each TS, organized by the WGs, will be several days long. The Action will make sure that leading experts share their knowledge, with particular focus on education of ESRs. Hot topic sessions are envisaged. TSs provide the opportunity for intensive training and networking, as supported by in depth discussions. These are the foundation for future scientific cooperation. For ESRs, special sessions will also be organised that train transferable skills, for example, bioethics, communication, scientific writing and presentation, as today's job market not only demands from ESRs that they know how to conduct good research, but also to have such transferable skills.

Short-Term Scientific Missions (STSM)

An ad hoc Short Term Scientific Mission (STSM) evaluation committee will be appointed by the MC with one coordinator and one representative of each WG. This committee will favor the mobility and training of young researchers. A number of STSMs will be financed each year for exchange of technologies and training between labs in different member states as defined in the rules of the COST framework. The applications for STSMs will be sent to the Chair of the

Evaluation Committee, with approval subject to COST rules and guidelines. STSMs are important to facilitate cooperation and initiating collaborations between Action participants. STSMs will provide ESRs with additional opportunities for participation within the Action and scientific and non-scientific, e.g. cultural, skill development.

E.2 Working Groups

The structure of the COST Action will be based on four main WGs (see D.2). Each WG represents a specific sub-topic to be analyzed as part of ongoing research. Action members will be assigned to WGs according to their expertise and interests. At the first meeting, WG leaders and deputy leaders will be nominated by the MC. The main tasks of the WG leaders are:

- Participating in MC sessions;
- Reporting WG progress to the COST Action;
- Planning and coordinating scientific meetings and activities;
- Coordinate the activities within their WG in order to meet scientific objectives
- Encouraging WG members to prepare common publications;
- Joint research programmes (e.g. through STSMs) within and between WGs;
- Inviting non-members to WG meetings and recruiting additional participants;

E.3 Liaison and interaction with other research programmes

Several participants are also involved in ELIXIR, the European sustainable bioinformatics infrastructure initiative. In order to coordinate both information exchange and joint activities (e.g. training schools and workshops), a liaison officer will be appointed by the MC. This will also serve to discuss the best ways to ensure the long-term availability of NGP-related data beyond this COST Action. Similarly, interactions will be sought with several international scientific societies in the fields of computational biology (e.g. ISCB, ISB) and protein science (e.g. FEBS, Protein Society),

especially in light of the possibility to organize joint workshops. Synergies and inter-COST meetings can also be sought with some existing COST Actions. The following COST Actions share some objectives on the analysis of specific protein phenomena: BM1403; BM1203; CM1207; CM1402. In all cases, the approach chosen by this Action is different than other existing programs, focusing on broad classes of NGP phenomena rather than specific systems, making the programs complementary rather than duplicating efforts.

E.4 Gender balance and involvement of early-stage researchers

This COST Action will respect an appropriate gender balance in all its activities and the Management Committee will place this as a standard item on all its MC agendas. The Action will also be considerably committed to involve ESRs, also placing it as a standard item on all MC agendas.

In more detail, the Action will ensure that women and ESRs are included in its scientific and organizational activities. With respect to gender balance, the MC will implement a policy to recruit female scientists to the labs of Action members. Special attention to the gender balance will be given for the Chairs of the MC and WGs. It is worth emphasising that female scientists have already played a major role in the design and delineation of the objectives of the Action. It is therefore expected that young and female scientists will play leading roles in the management of the Action. They will be invited and encouraged to present and chair sessions. As described above, gender balance will be observed in the workshops, teaching activities and STSMs, in which early-stage researchers are also expected to form the majority of participants. Each WG will nominate an ESR, who will communicate with other ESRs to seek feedback and ideas. This will then be presented to the MC.

Special attention for the development of ESRs will be given in the organization of training schools (see section E.1) that will particularly address the needs of ESRs. In the selection process of STSM proposals and admittance to workshops and training courses, priority will be given to favor gender balance and Early-Stage Researchers. Particular focus will be placed on including interdisciplinary exchanges across labs. These will facilitate the exchange of methods, and train ESRs in new techniques that they cannot learn at their home labs. Whilst offering excellent international/interdisciplinary research training experiences to ESRs, these exchanges will also underpin the inoculation of new research collaborations/ideas, which will lead to joint research

publications as well as prompt new and competitive collaborative grant applications both at the national, international and European level.

F. TIMETABLE

The duration of this Action will be four years. The kick-off meeting will mark the start-point, where the chair, co-chair, secretary, WG co-ordinators and Web-site coordinator will be selected. The Action's Website will be generated in the first trimester and will be continuously updated. Reports will be presented each year. Details of the frequency and timing of MC meetings, WG meetings, workshops, STSMs and teaching activities are indicated in the table below. Note that WG/MC meetings will take place at least once a year. Dependent on the available budget and the specific needs of the Action, their frequency may increase to twice a year. Regarding scientific deliverables, the Action expects that by the end of Y4 the targets (see Section D) will be fulfilled or considerably progressed.

Activity	Year 1				Year 2				Year 3				Year 4			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1st MC meeting	x															
Dissemination plan set up	x															
Web site establishment	x															
Web site updates		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
MC meetings					x				x				x			
SC video meetings		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
WG meetings (>= 1/year)	x				x				x				x			

Training schools (>= 1/year)		x				x				x				x	
STSMs	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Newsletter			x				x				x				x
Progress reports			x				x				x				x
Workshop	x				x				x				x		
Final conference															x

G. ECONOMIC DIMENSION

The following COST countries have actively participated in the preparation of the Action or otherwise indicated their interest: AT, BE, CH, CZ, DE, EL, ES, FR, HU, IE, IT, LT, NL, PL, PT, RO, RS, SE, TR, UK. On the basis of national estimates, the economic dimension of the activities to be carried out under the Action has been estimated at 80 Million € for the total duration of the Action. This estimate is valid under the assumption that all the countries mentioned above but no other countries will participate in the Action. Any departure from this will change the total cost accordingly.

H. DISSEMINATION PLAN

H.1 Who?

The activities of the Action will be disseminated as widely as possibly to the following target audiences:

- Researchers: The Action is an interdisciplinary network of European scientific experts and ESRs providing critical mass to perform internationally competitive R&D in a way that would not be achievable by small-scale cooperations. The dissemination methods

will trigger the interest of further scientists contributing further expertise, insights, technologies and tools. This provides opportunities that cannot be foreseen at this stage;

- Pharmaceutical and medical industries: The Action expects increased competitiveness and leadership in preclinical discovery, thereby facilitating the growth of biotechnology and pharmaceutical industries in Europe as well as stimulating the establishment of new SME building on the scientific deliverables;

- EU member states, policy makers and regulators: The Action will provide excellent training options for the future scientific leaders. It is confident to enhance the competitiveness by joining forces of outstanding European scientists in a large network to achieve high-level objectives that would not be achievable at an individual level. Thereby, the Action will contribute to European reputation as a place of excellence in biomedical and translational research;

- Scientific societies, e.g. International Society for Computational Biology (ISCB), International Society for Biocuration (ISB), Protein Society and FEBS in getting access to the latest, ground-breaking R&D results in the field and support of its conferences.

H.2 What?

The Action will develop a strategic communication, dissemination, and education plan for non-globular proteins (NGPs) by using the following dissemination methods:
Web portal:

- General information about the project, partnership, organisation and contact addresses, high-level goals, and information for non-scientist stakeholders (journalists, undergraduate students, etc.);

- The public part website will inform about the latest research results;

- A communication area that holds E-mail distributors, a FAQ, a bulletin board, etc;

- A calendar with meetings, workshops, courses, conferences (including those from non-partners);
- The actual educational material from courses and workshops, and the actual seminars presented at conferences, where possible including a video recording; when applicable also the proceedings of conferences, etc;
- Lists of available software products, sorted by category (e.g. Web server, Database) and topic (e.g. repeats, disorder);
- Lists of ongoing software projects;
- Lists of job-openings in NGPs;
- A small password protected area for the exchange of sensitive material like pre-submission versions of articles, minutes of meetings, etc.

Publications:

- Original papers and reviews jointly published in peer-reviewed journals;
- One or more brochures that the partners can distribute nationally, or at conferences;
- Educational material for workshops and courses;
- Lay information, including citizen science projects (e.g.: http://scistarter.com/) and Wikipedia (see below).

Coordinate / stimulate a series of activities:

- Training courses developed around thematic areas, tailored to participants that require maximum functional skills in a narrow area in minimum time. This Action will provide such opportunities by engaging in high quality skill deployment methods and tuned delivery of materials, in collaboration with existing initiatives e.g. ELIXIR;
- Workshops, courses, and summer schools, in order to disseminate the latest developments in the NGP field and also combine hands-on practical training with

theoretical information. These will be offered mainly to young investigators. The Action proposes tailored training at moderate costs, including training facilities, lodging and subsistence at adequate levels for ESRs. A few participating countries have indicated that they can provide such facilities;

- Presentations at major international conferences and seminars at academic institutions by Action members. Industry representatives will be invited and research results presented to further increase industry involvement;

- Link into societies: Some Action participants are society members (e.g. ISCB) and support integration;

- A final international conference.

Training young people is a crucial dissemination aspect. The action will therefore stimulate young scientists to go on training sessions in the labs of Action partners, or beyond. Financial support will be provided for such Short-Term Scientific Missions (STSMs). Other European (e.g. Elixir) and national programs are thinking about intensive programming sessions, trying to ensure that hack-aton collaborations will be forged. Open source code will be developed, using the rules available in www.opensource.org/licenses. This will enable the participation of young developers in an organized way.
An innovative key feature of the COST Action will be the contribution to Wikipedia and critical assessments which will be used to:

- Provide thematic pages describing the state of the art in each NGP sub-field;

- Create high-level pages to describe the relationship between NGP classes and a consensus for their nomenclature;

- Summarize properties of existing methods and analysis tools;

- Disseminate relevant publications;

- Share experience of best practice and known issues and problems with designs and methodologies;

- Identify expertise within the European Scientific community;
- Identify priority areas for future research and development.

H.3 How?

The Action will work out a concrete dissemination plan in the first three months, to be regularly updated based on evaluations, using the following dissemination methods:

- Action Network logo used on all publications, documents, events and presentations;
- Action website (set up during the first quarter, see H.2);
- Twitter will be used to increase visibility;
- Presentations at major scientific international conferences and seminars at academic institutions by Action members;
- The Action will involve societies, e.g. to organize special (e.g. hot topic) sessions at conferences;
- Connection with other networks and service organisations will foster cross-sectional interactions, e.g. with ELIXIR (bioinformatics infrastructure). Further, the International Society for Computationall Biology (ISCB) and International Society for Biocuration (ISB) hold international conferences (such as ISMB, ECCB and Biocuration) and the Action plans to support speaker participation;
- Stalls at scientific conferences will promote the Action among international scientific audiences;
- Original papers and reviews in peer-reviewed journals;
- Training schools for ESRs;
- Exchange of ESRs between Action laboratories;

- A final international conference will bring together scientists, industry partners and other interest groups on non-globular proteins from all over the world. It will thus also promote the outcomes to non-European audiences.

The Management Committee (MC) will be responsible for activities that require a budget such as STSM requests, deciding on funding for workshops, courses, etc. The representative members of each country will be responsible for disseminating the activities of the Action to research groups within their countries, industrial partners, societies, and representatives of society. Each MC member is therefore expected to generate, regularly update, and circulate to other MC members a list of target groups with contact information. For regional meetings and other activities, the MC will delegate responsibilities to WG coordinators and members of the WGs depending on their specialty.