Part A.
Proposal Data and Obligations – Project Proposals

Helmut Grubmüller, Göttingen
Holger Dachsel, Jülich
Berk Hess, Stockholm

I. Proposal data

1. Type of proposal

Programme

Priority Programme SPP 1648
SPPEXA — Software for Exascale Computing
Coordination proposal.

Proposal category

New proposal

2. Proposal information

2.1. Title/duration

Title (in German)
“Flexible und skalierbare Potentialberechnung für realistische Molekulardynamik-Simulationen mit dynamischer Protonierung”
Title (in English)
“Unified Long-range Electrostatics and Dynamic Protonation for Realistic Biomolecular Simulations on the Exascale”
36 months

2.2. Subject classification

Subject area: Biophysics

2.3. Keywords

Keywords (in German):
Simulationen, langreichweitige Wechselwirkungen, schnelle Multipolmethode

Keywords (in English):
simulations, long-range interactions, fast multipole method

2.4. Countries

None

2.5. Summary

English Summary

In this proposal, we target a flexible, portable and ultra-scalable solver for potentials and forces, which is a prerequisite for exascale applications in particle-based simulations with long-range interactions in general. As a particularly challenging example that will prove and demonstrate the capability of our concepts, we use the popular molecular dynamics (MD) simulation software GROMACS. MD simulation has become a crucial tool to the scientific community, especially as it probes time- and length scales difficult or impossible to probe experimentally. Moreover, it is a prototypic example of a general class of complex multiparticle systems with long-range interactions.

MD simulations elucidate detailed, time-resolved behaviour of biology's nanomachines. From a computational point of view, they are extremely challenging for two main reasons. First, to properly describe the functional motions of biomolecules, the long-range effects of the electrostatic interactions must be explicitly accounted for. Therefore, techniques like the particle-mesh Ewald method were adopted, which, however, severely limits the scaling to a large number of cores due to global communication requirements. The second challenge is to realistically describe the time-dependent location of (partial) charges, as e.g. the protonation states of the molecules depend on their time-dependent electrostatic environment.

Here we address both tighly interlinked challenges by the development, implementation, and optimization of a unified algorithm for long-range interactions that will account for realistic, dynamic protonation states and at the same time overcome current scaling limitations.

Deutsche Zusammenfassung

Ziel dieses Projektes ist es, einen portablen, flexiblen und extrem skalierbaren Algorithmus für Potentiale und Kräfte zu konzipieren. Ein solcher Algorithmus ist unabdingbare Voraussetzung für die Simulation von unterschiedlichsten Vielteilchensystemen mit langreichweitigen Wechselwirkungen auf zukünftigen Parallelrechnern. An der verbreiteten Molekulardynamik (MD) Software GROMACS, die ein typisches und anspruchsvolles Anwendungsbeispiel für einen solchen zukünftigen Algorithmus ist, werden wir die Leistungsfähigkeit unserer Konzepte testen, entwickeln und demonstrieren. Molekulardynamik-Simulationen sind heutzutage ein elementares Werkzeug der Wissenschaft, insbesondere weil sie die Dynamik von Biomolekülen auf Zeitskalen erschließen, welche mit experimentellen Methoden nicht erreichbar sind.

MD Simulationen ermöglichen es, das Verhalten biologischer Nanomaschinen detailliert und mit hoher Zeitauflösung zu untersuchen. Aus computertechnischer Sicht sind diese Berechnungen jedoch außergewöhnlich herausfordernd, und zwar aus folgenden Gründen. Zum einen muß für eine korrekte Beschreibung funktioneller Bewegungen die Langreichweitigkeit elektrostatischer Kräfte explizit berücksichtigt werden. Dazu werden heutzutage Methoden wie "Particle-Mesh Ewald" eingesetzt; diese sind jedoch aufgrund ihrer globalen Kommu-

nikationsanforderungen generell nicht auf beliebig hohe Prozessorzahlen skalierbar. Zum anderen ist die realistische Beschreibung der Ladungsverteilung besonders anspruchsvoll, da beispielsweise die Protonierungszustände von Biomolekülen von ihrer elektrostatischen Umgebung abhängen und sich somit zeitlich verändern.

In diesem Projekt begegnen wir beiden eng miteinander verknüpften Herausforderungen mit der Entwicklung, Implementierung und Optimierung eines vereinheitlichten Algorithmus für langreichweitige Wechselwirkungen, welcher, als eine von vielen Anwendungen, realistische dynamische Protonierungszustände in Biomolekülen ermöglicht und gleichzeitig die gegenwärtigen Skalierungs-Einschränkungen überwindet.

3. Participating individuals

3.1. Applicants

Academic degree/title:	Hon. Prof. Dr.
First name:	Helmut
Last name:	Grubmüller
Nationality:	German
Gender:	m
Date of birth:	31.07.1965
German-speaking:	y
E-mail address:	hgrubmu@gwdg.de
Phone:	05 51 / 2 01 23 01
Address:	Max Planck Institute for biophysical Chemistry Theoretical and computational biophysics Am Faßberg 11, 37077 Göttingen

Academic degree/title:	Dr.
First name:	Berk
Last name:	Hess
Nationality:	Dutch
Gender:	m
Date of birth:	08.11.1972
German-speaking:	n
E-mail address:	hess@kth.se
Phone:	+46 / 8 55 37 80 27
Address:	Department of Theoretical Physics School of Engineering Sciences Royal Institute of Technology SE-100 44 Stockholm

Academic degree/title:	Dr.
First name:	Holger
Last name:	Dachsel
Nationality:	German
Gender:	m
Date of birth:	20.03.1961
German-speaking:	y
E-mail address:	h.dachsel@fz-juelich.de
Phone:	0 24 61 / 61 65 69
Address:	Forschungszentrum Jülich Institute for Advanced Simulation Jülich Supercomputing Centre, Mathematics Division Leo-Brandt Str. 52425 Jülich

Address of the institution that will host the proposed project:

Max Planck Institute for Biophysical Chemistry
Theoretical and Computational Biophysics
Am Fassberg 11, 37077 Göttingen

3.2. Other participating individuals

Academic degree/title:	Dr.
First name:	Carsten
Last name:	Kutzner
Nationality:	German
Gender:	m
Date of birth:	28.06.1972
German-speaking:	y
E-mail address:	ckutzne@gwdg.de
Phone:	05 51 / 2 01 23 13

Academic degree/title:	Dr.
First name:	Ivo
Last name:	Kabadshow
Nationality:	German
Gender:	m
Date of birth:	22.06.1980
German-speaking:	y
E-mail address:	i.kabadshow@fz-juelich.de
Phone:	0 24 61 / 61 87 14

4. Participating institutions

No other cooperating institutions or industry partners.

II. Obligations

In submitting a proposal to the Deutsche Forschungsgemeinschaft (DFG), all applicants agree to

- adhere to the **rules of good scientific practice**;
- use the grant exclusively and in a targeted manner to realise the funded project, to conform to the relevant regulations of the DFG in the use and accounting of funds, and in particular not to use the grant to finance core support.
- submit research progress reports according to the dates specified in the award letter and to present financial accounts to the DFG detailing the use of funds.
- have adhered to the guidelines regarding publication lists and bibliographies.
- and — if applicable —
  - inform the DFG immediately if funding for this project is requested from a third party. Proposals requesting major instrumentation and/or those previously submitted to a third party must be mentioned in the *Project Description* under *Additional Information.*
  - inform your university's DFG liaison officer about the proposal submission;
  - inform the head office of the Max Planck Society about the proposal submission;
  - plan and conduct any **experiments involving humans**, including identifiable samples taken from humans and identifiable data, in compliance with the Declaration of Helsinki (Ethical Principles for Medical Research Involving Human Subjects) concluded by the World Medical Association (WMA) in June 1964, as last revised; as well as the regulations of the German Embryo Protection Law (Embryonenschutzgesetz), Stem Cell Act (Stammzellgesetz), Medicines Law (Arzneimittelgesetz, sec. 40–42) and Medical Devices Act (Medizinproduktegesetz, sec. 17–19), in their current forms.
  - plan and conduct any **animal experiments** in compliance with the German Animal Protection Law (Tierschutzgesetz) and refrain from commencing such research until approval has been obtained.
  - plan and conduct any **experiments involving recombinant DNA** in compliance with the Genetic Engineering Act of 20 June 1990 (Gesetz zur Regelung von Fragen der Gentechnik, Bundesgesetzblatt, 1990 I, page 1080) and refrain from commencing research work until the approvals required under this law and its related ordinances have been obtained.
  - obtain approval required for embryonic stem-cell research and refrain from commencing such research until it has been officially authorised; the funds earmarked for work with human embryonic stem cells remain locked until the appropriate approvals have been submitted to the DFG.
  - if the research project, or parts thereof, are subject to the Convention on **Biological Diversity**, to follow the *Guidelines for Funding Proposals Concerning Research Projects within the Scope of the Convention on Biological Diversity (CBD)*.

I/We accept the foregoing conditions and obligations.

I/We agree to:

- the DFG's electronic processing and storage of data provided in conjunction with this proposal. This information may be passed to reviewers and the DFG bodies as part of the DFG's review and decision-making process.
- having all address and contact information (telephone, fax, e-mail, internet website), as well as information on the content of this research project (e.g. title, summary, keywords, international cooperation), if approved, published in the DFG's project database GEPRIS (http://www.dfg.de/gepris) and — in excerpts (grant holder's name, institution and location) — in the "Programmes and Projects" section of the DFG's electronic annual report (http://www.dfg.de/jahresbericht). I/We understand that the electronic publication of this information may be opposed by contacting the appropriate programme contact no later than four weeks from receipt of the award letter.

I/We accept the foregoing conditions.

Signature(s) (all applicants)

City:	Göttingen	Stockholm	Jülich
Date:			
	Helmut Grubmüller	Berk Hess	Holger Dachsel

I/We agree to:

- the DFG's electronic processing and storage of data provided in conjunction with this proposal. This information may be passed to reviewers and the DFG bodies as part of the DFG's review and decision-making process.
- having all address and contact information (telephone, fax, e-mail, internet website), as well as information on the content of this research project (e.g. title, summary, keywords, international cooperation), if approved, published in the DFG's project database GEPRIS (http://www.dfg.de/gepris) and — in excerpts (grant holder's name, institution and location) — in the "Programmes and Projects" section of the DFG's electronic annual report (http://www.dfg.de/jahresbericht). I/We understand that the electronic publication of this information may be opposed by contacting the appropriate programme contact no later than four weeks from receipt of the award letter.

I/We accept the foregoing conditions.

Signature(s) (all applicants)

City:	Göttingen	Stockholm	Jülich
Date:		09-05-2012	
	Helmut Grubmüller	Berk Hess	Holger Dachsel

I/We agree to:

- the DFG's electronic processing and storage of data provided in conjunction with this proposal. This information may be passed to reviewers and the DFG bodies as part of the DFG's review and decision-making process.
- having all address and contact information (telephone, fax, e-mail, internet website), as well as information on the content of this research project (e.g. title, summary, keywords, international cooperation), if approved, published in the DFG's project database GEPRIS (http://www.dfg.de/gepris) and — in excerpts (grant holder's name, institution and location) — in the "Programmes and Projects" section of the DFG's electronic annual report (http://www.dfg.de/jahresbericht). I/We understand that the electronic publication of this information may be opposed by contacting the appropriate programme contact no later than four weeks from receipt of the award letter.

I/We accept the foregoing conditions.

Signature(s) (all applicants)

City:	Göttingen	Stockholm	Jülich
Date:			
			
	Helmut Grubmüller	Berk Hess	Holger Dachsel

Part B.
Project Description – Project Proposals

Project Description

1. State of the art and preliminary work

One of the hallmarks of complex systems is that most of them are large, strongly interacting multiparticle systems. Examples are ubiquitous, e.g., in cosmology (interacting galaxies), in astrophysics (interacting stars in galaxies or stellar clusters), plasma physics (interactions via coulomb forces), solid state physics (atoms and defects in crystal lattices), complex fluids (strongly interacting molecules), biological macromolecules and many others. At the level of highest complexity, even the brain may be described as a stongly interacting system of many neurons or synapses.

Large scale simulations of all these systems pose computational challenges which share many similar characteristics and strategies particularly at the exascale level. An in-depth analysis and treatment of one well chosen example therefore not only benefits the specific field, but also holds the promise to provide a very general tool box for a wide range of exascale applications.

In all cases, both users and code developers face mainly two lines of challenges. First, the inter-particle interactions are typically long-range, which calls for special numerical techniques to cut down computational complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \cdot \log N)$ or even $\mathcal{O}(N)$. Depending on the specific application, different schemes have been developed and applied, such as PME [DYP93] [EPB+95] [DH98] [BYQS03] [CDDL09] [WDH10], all facing communication bottlenecks, and thus achieving various levels of scalability on large parallel machines. The second ubiquitous challenge is that of time-dependent interaction functions, i.e., changing masses of the interacting galaxies or stars, changing ionization states in plasmas, or changing titration states in fluids or macromolecules. This second challenge poses severe limitations as to which multiparticle numerical method can be applied, and it bears the risk of creating further communication bottlenecks, thus calling for sophisticated computation/communication balancing algorithms.

In all cases, scalable codes at the exascale level require a strict co-design approach. In particular, different levels of physics and modeling need to be traded off against available numerical and computational strategies and algorithms.

To more specifically describe this very general computational, we here consider the example of biomolecular systems, which (a) exhibit a comparably high level of complexity and (b) have attracted a large number of physicists, computer scientists, mathematicians, chemists, and biologists to tame this complexity. As a result, several codes have been developed [CHB+05] [EWGP06] [HKvdSL08] [CCD+05] [BCX+06] [PBW+05] and are currently employed by a very broad user base.

Proteins and protein complexes are biology's nanomachines, performing the vast majority of all cellular functions. Ranging from transporters to enzymes, from motor to signalling proteins, conformational transitions are frequently at the heart of protein function, rendering a detailed understanding of the involved *dynamics* indispensable. From a physics point

of view, proteins represent tightly coupled, heterogeneous many-particle systems, which exhibit highly complex dynamics at time scales ranging from femtoseconds to seconds. Experimentally, the atomistic dynamics on submillisecond timescales are notoriously difficult to access, making computer simulations the method of choice. Molecular dynamics (MD) simulations of biomolecular systems have come of age and are routinely used to study the mechanisms underlying biological function on the atomistic level. Individual proteins and other biomolecules such DNA, as well as assemblies consisting of multiple biomolecules and membranes [THS+06], up to whole virus particles [ZG10] and subcellular organelles are currently accessible to such simulations. While many biological processes take place on timescales of milliseconds or slower, the current simulation state of the art for moderately sized MD systems, using accurate protocols, lies in the microsecond timescale [LLF+08] [PLP+11]. Hence, there is a clear gap between the scientists' needs and the capabilities of current hard- and software. Apart from the system size that poses a natural limit to such simulations, there are two effects that limit the scope of biomolecular simulations: Accuracy and simulation length. Both are topics of this proposal, and are limitations faced by all types of many-particle simulations mentioned above. Hence, also in this respect, the strategies developed in this proposal will be of very general applicability.

It has been shown that particularly the accurate calculation of the forces resulting from the electrostatic interactions in biomolecular complexes is essential for a proper description of their functional motions. These interactions are numerically challenging in two respects.

First, their long-range character (the potential drops off slowly with $1/r$ with distance r) renders traditional cut-off schemes prone to artefacts, such that mesh-based Ewald summation methods were introduced, which provide an accurate solution for 3d periodic boundaries. The current standard is the particle-mesh Ewald (PME) algorithm [EPB+95] that makes use of fast Fourier transforms to result in an efficient long-range solver that scales as $N \cdot log\, N$ with N the number of particles. Due to the nature of PME, however, which requires multiple all-to-all communication steps per time step, one of the inherent challenges is the scaling of this algorithm to a large processor count p, since the number of involved messages is p^2 [KvdSF+07]. For the PME algorithm included in the highly efficient, open-source MD package GROMACS, much effort has been made by the developers to reduce as much as possible the all-to-all bottleneck, e.g. by partitioning the parallel computer in long-range and short-range processors, which reduces the number of messages involved in all-to-all communication [HKvdSL08]. GROMACS is one of the fastest, best-scaling and most versatile MD engines worldwide. Despite all these efforts, however, even for multimillion atom MD systems simulated on modern hardware, the performance levels off beyond 10,000 cores due to the inherent scaling limitations of the PME algorithm [SLPS09].

The second challenge is the tight and non-local coupling between the electrostatic potential and the location of charges on the protein, in particular protonable groups the charge of which changes with their electrostatic environment. Hence, all protonation states are closely coupled, depend on pH, and therefore the protonation/deprotonation dynamics needs to be taken into account during the simulation. Particularly for larger functional, conformational, or electrostatic changes, the lack of appropriate algorithms for this processes poses one of the most severe limitations for realistic biomolecular simulations.

Currently, most MD simulations therefore focus on *static* protonation states, based on heuristic estimates. Recently, a *dynamic* scheme based on λ-dynamics has been designed by one of the applicants that allows for a realistic equilibrium between protonated and unprotonated states, thereby allowing for constant-pH, rather than constant protonation state simulations [DTGG11]. The constant-pH concept requires that different protonation states are considered during the course of the simulation, each one of them consisting of to two or more

locally differing charge distributions. PME, however, requires the separate evaluation of the Fourier coefficients for each charge alternative. Each of these evaluations involves expensive all-to-all communication when using PME, which severly limits the parallel scaling.

Here, we propose to set up a unified and general framework both on the software level and on the human/logistics level, to develop and provide an "exascale toolkit" that serves a very broad community of diverse applications in supporting the transition to the exascale level. In contrast to the ongoing efforts of isolated communities, here we will identify the general challenges and features which the numerical tasks and solutions share. Using biomolecular simulations both as a role model and proof of concept system, we not only aim at exa-scalable codes, but also at a co-design type collaboration structure between users and developers, which enables efficient maintenance, generalization, evaluation, optimization, and distribution of exascale codes. We anticipate that new paradigms for the exascale level will emerge.

1.1. Project-related publications

1.1.1. Articles published by outlets with scientific quality assurance, book publications, and works accepted for publication but not yet published

[Dac06] H Dachsel. **Fast and accurate determination of the Wigner rotation matrices in the fast multipole method**. *J Chem Phys*, 124:144115, 2006.

[Dac10] H Dachsel. **An error-controlled fast multipole method**. *J Chem Phys*, 132(11):119901, 2010.

[DTGG11] S Donnini, F Tegeler, G Groenhof, and H Grubmüller. **Constant pH Molecular Dynamics in Explicit Solvent with λ-Dynamics.** *J Chem Theory Comput*, 7:1962–1978, 2011.

[HKvdSL08] B Hess, C Kutzner, D van der Spoel, and E Lindahl. **GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation**. *J Chem Theory Comput*, 2008.

[KD11] I Kabadshow and H Dachsel. **An Error Controlled Fast Multipole Method for Open and Periodic Boundary Conditions**. *IAS series, Fast Methods for Long-Range Interactions in Complex Systems*, 6(11):85–114, 2011.

[KGdGZ11] C Kutzner, H Grubmüller, BL de Groot, and U Zachariae. **Computational Electrophysiology: The Molecular Dynamics of Ion Channel Permeation and Selectivity in Atomistic Detail**. *Biophys J*, 101(4):809–817, 2011.

[KvdSF+07] C Kutzner, D van der Spoel, M Fechner, E Lindahl, UW Schmitt, BL de Groot, and H Grubmüller. **Speeding up parallel GROMACS on high-latency networks**. *J Comput Chem*, 28(12):2075–84, 2007.

1.1.2. Other publications

[PLP+11] S Pronk, P Larsson, I Pouya, G Bowman, I Haque, K Beauchamp, B Hess, VS Pande, P Kasson, and E Lindahl. **Copernicus: A new paradigm for parallel adaptive molecular dynamics**. *Proceedings of SC'2011*, pages 60–70, 2011.

1.1.3. Patents

None pending, none issued.

2. Objectives and work programme

2.1. Anticipated total duration of the project

3 Years.

2.2. Objectives

Generic approach

In this project, we aim to develop an ultra-scalable, versatile solver for complex long-range interactions that is applicable to the broad class of particle-based simulations in general. The main goal concerning the method used to derive forces from $1/r$-type potentials will be

(a) to optimize the scalability and

(b) to broaden the applicability

to the general situations and geometries needed by all types of particle-based application codes abundant in astrophysics, fluid dynamics, biomolecular simulation, and other fields.

To reach these ambitious goals, we will

1. characterize the scaling behavior of different long range methods under exascale conditions,
2. in a close co-design loop, identify the method that proves most generalizable (we anticipate that will be the fast multipole method (FMM) and structure-adapted FMM),
3. generate a toolbox of how to advance the FMM to the exascale level for various test-systems,
4. generate a toolbox of how to overcome the communication bottleneck for time-dependent interactions,
5. evaluate the toolbox by one or two sample implementations with the characterization of the scaling behavior for, e.g., biomolecular simulations and astrophysical simulations, both for (a) long range FMM and (b) dynamic charge/mass changes.

In detail, we will first evaluate algorithms for long-range potentials with respect to their ability to (a) overcome the current scaling bottleneck by avoiding expensive global communication and to (b) allow for the efficient treatment of local alternative charge or mass distributions. From our current point of view, a promising candidate to fulfil both conditions is the fast multipole method (FMM). The latter scales by construction much better than PME, which is in addition not well suited for alternative charge distributions, since it yields a scaling behaviour linear with n_λ in the long-range part of the potential. As n_λ usually grows with system size, a solver largely independent of n_λ is clearly needed to reach exascale. Opposed to PME, additional multipole expansions for local charge alternatives can be computed and communicated with the FMM. This makes the communicated volume (extra multipole components) somewhat larger, but no global communication steps are involved as in PME, where the global communication volume grows linearly with n_λ and quadratic with the number of processes p. Once the ideal method to reach goals (a) and (b) has been identified, suitable measures are taken to optimize the scalability. This will be done by redesigning the meth-

ods' communication model in order to reduce global communication to an absolute minimum. Additionally, support for time-dependent interactions will be build into the method. Assuming the ideal method is the FMM, this will be done in the form of alternative multipole expansions for the local environment. To emphasize the co-design aspects, we will continuously benchmark the scaling behaviour of our algorithm(s) using actual scientific challenges from different fields of simulation, e.g. molecular dynamics and astrophysics applications. A one result, a constant-pH biomolecular simulation engine based on λ-dynamics will be available for massivly parallel simulations. Since the FMM scales linearly with particle number N, it paves the way to efficient computation of systems significantly larger than a few million particles. For our MD example, our goal is to push the current limit of 10^4 cores which can be simultaneously used in parallel by a factor of 100. This step holds the promise to advance first principles drug development (as opposed to current docking approaches) for industrial large scale exploitation.

Generalizability and co-design aspects

The project as a whole is intended to support all types of complex particle simulations. A "particle" can be anything from an atom or a small bead, as in molecular simulations, up to a whole star or even a cluster of stars, as in astrophysical simulations. For increased flexibility, we plan to support efficient parallel calculation of r^a-type potentials not only for $a = -1$, but also for any real a. To allow the straightforward adaptation for and/or incorporation into future exascale applications, the general and robust yet customizable algorithms will be made available in a modular, freely available software library. Furthermore, we will put the co-design paradigm into practice by incorporating users and other developers into the design process by providing a web-based communication platform and a flexible concurrent version control system. The success story of 15 years of GROMACS development has taught us that freely distributed, open source software packages are generally most beneficial to the scientific community. These packages enjoy large user bases (in total, the three main GROMACS publications of the years 2001 [LHvdS01], 2005 [vdSLH$^+$05], and 2008 [HKvdSL08] have been cited ≈6300 times due to Google scholar, ISI, and Scopus) not only because they are free of charge but mainly because there is a strong link between the users (scientists who incorporate these concepts in form of algorithms into their own code), and the developers. Therefore, in these projects there is much feedback in the form of code contributions, bug reports/fixes, feature requests, and optimizations from the user base to the project itself. Often, users over time contribute more and more and sometimes get part of the group of developers. This type of co-design has been successfully established for the GROMACS project, where feedback is facilitated by mechanisms like, e.g.,

- a mailing list for users and developers, where project-related questions are discussed,
- a code submission and quality control server to which code modifications and additions can be submitted,
- workshops for users and developers,
- a homepage with a wiki to which users and developers contribute.

These co-design paradigms will likewise be set up for a general exascale software toolbox containing the optimized parallel FMM and possibly other exascale algorithms developed in this project, within the SPPEXA, or contributed by other scientits working in that area. This framework should serve as a nucleus to encourage future users and developers to actively contribute to the design of tools for exascale applications.

As an example of co-design, we show how the developed concepts can be integrated into a simulation code to advance it towards exascale computing.

Relevance within SPPEXA

This project responds to the exponentially increasing computational demands of complex systems research, as well as to the end of Moore's law at the single-core level. By providing an extremely scalable, latency-tolerant solver for long-range interactions with optimal algorithmic complexity, data locality, and minimal communication, we address efficiency in computation as well as in energy consumption, scalability, developer/user communication and generalizability in an integrated co-design framework.

In addressing the exascale demands at the very general level of complex, strongly interacting multiparticle systems, this project serves a very broad user community. It will combine a general, efficient and exa-scalable toolbox of modules framed in a co-design approach with a proven world-wide 15 years' track record for biomolecular simulation. As intended by the SPP, the project brings together groups focusing on fundamental HPC methodology (Jülich) with developers working on the most important European MD code (Stockholm and Göttingen). Therefore, trans-disciplinary research activities are bundled, which will enable all three groups to significantly advance the state of the art of scientific HPC applications towards the exascale.

In particular, we concentrate on the SPPEXA suptopics #1 "Computational Algorithms" and #3 "Application Software". We address extremely large particle-based simulations as a general class of applications (#3). Our flexible and unified algorithm will boost the capability of scientific computing in general and of biomolecular simulation in particular by enhancing both the model depths and the accessible time scales. The developed ideas and concepts will be readily and publicly available for scientific use.

Since our algorithms and collaboration structures will be quite general, we expect them to be valuable for various particle-based simulations, as e.g. proposed in the "ExaMPIeS–Exascale Multiphase Particle Simulations" project, and for many astrophysical particle-based simulation codes (e.g. the GADGET code for cosmological simulations). Within SPPEXA, we are looking forward to collaborations with several other projects that should nicely interlink with our proposal, like, among others, the "PPAT–Performance-aware Programming for Automatic Tuning" and "EXIOS–A Unified Exascale I/O System" projects. EXIOS can offer optimized I/O routines including data compression algorithms for the large data volumes expected from exascale particle simulations while our program would in turn provide one or even several real-world application(s) as a test case for the I/O system. We have already initiated a close collaboration with the applicants of the EXIOS and ExaMPIeS projects. Especially the ExaMPIeS and GromEx projects will mutually benefit substantially, since the proposed methods (molecular dynamics with fast multipole methods) are the same but the area of application is an entirely different one (biomolecules vs. microfluidics) — very much following the lines of generalizability laid out above. We believe that our combined experiences will enable powerful, flexible, and universal algorithms for a diverse range of physical applications.

2.3. Work programme incl. proposed research methods

The whole project consits of five tasks, of which tasks 2–4 can be started independently and in parallel, while all tasks involve a very close collaboration between all three groups (compare Table 1). While the Stockholm group is not asking for own funding, they will provide training and support both Göttingen candidates (M. Hoefling, MH and T. Graen, TG) as well as the Jülich candidate (R. Halver, RH) throughout the whole project with workshops and

Table 1: Estimate of distribution of work-packages by half-year and group

Year	**2013**		**2014**		**2015**	
Taskset G PhD Göttingen (TG)	T1: benchmarks	T4: enhance FMM for alternative particle properties				T5: showcase application(s)
Taskset H PDoc Göttingen (MH)	T1: define & build interface	T3: alternative particle properties			T3: correctness tests T5: set up collaboration framework	
Taskset J PDoc Jülich (RH)	T1: define & build interface	T2: strong scaling FMM			T4: adapt FMM far field to local changes	T5: Documentation
Stockholm	training and assistance to Gö./Jü			Incorporation in GROMACS, help on T5		

internships.

Task 1: Evaluation and benchmark phase (shared among TG/Göttingen, MH/Göttingen and RH/Jülich)

As a first task, the scaling behaviour of different long-range interaction methods will be characterized.

To this aim, we will use a library developed by the Jülich team called “Scalable Fast Couloumb Solver” (ScaFaCoS). This freely available library offers various electrostatic solvers including parallel versions of PME, P^3M, multigrid, and the FMM. The version of the FMM includes full error control, allowing a runtime optimization [Dac06] [Dac10] [KD11]. In combination with memory optimization and an efficient parallelization the FMM was already successfully applied in a simulation of the biggest particle system up to now ($N > 3 \times 10^{12}$), taking full account of long range interactions between all particles [w08].

MH will build a preliminary interface between ScaFaCoS and our example particle simulation package (GROMACS). No general or flexible interface will be needed at this time since the goal at this stage is just to allow benchmarks of various solvers for the (electrostatic) potential in a realistic environment.

On the GROMACS side, the calculation of the Coulomb forces will be dissected from the rest of the dynamics engine since the van der Waals and the Coulomb forces on a particle are usually computed in a combined kernel due to optimization puposes. Subsequently, the particle positions and charges will be communicated via the interface to the ScaFaCoS library and forces will be communicated back to the MD application in turn. This way, van der Waals, bonded forces, and constraints are still calculated by the MD application while the Coulomb forces are calculated by the Jülich library.

Since both codes are parallelized, the distribution of the particles among the CPU cores will be adapted: The variable-sized MD domains will be mapped onto cubic multipole boxes. Then, benchmarks with variable system size N and processor count p will be made to evaluate which electrostatics method performs best for a given N and p. If several different methods turn out to be needed along the (N, p)-range, an automatic performance tuning can be implemented in a later phase of the project so that for any scenario, automatically the best performing method among all available ones is chosen.

The interface between the two codes will be implemented by RH (for the ScaFaCoS part)

and MH (for the GROMACS part, with the assistance of the Stockholm developers). The benchmarking will be done on the compute cluster of the Göttingen group for up to ≈ 200 cores, or, for larger number of cores, on the cluster of the GWDG in Göttingen or on one of the clusters of the Jülich supercomputing center.

To be able to detect (and remove) scaling problems at a very early stage in the design process, we intend to benchmark the FMM embedded in a sample application. Such scaling problems could occur for instance when particle redistributions are necessary to adapt the MD domains to FMM boxes, since these need not be identical.

Task 2: Strong scaling FMM (RH/Jülich)

In the second task, RH will strive for strong scaling in the FMM and will revise the communication model and memory management.

Using one-sided, point-to-point communication with a prefetching strategy will allow to hide most communication behind computation. Global communication will thereby be reduced to a minimum. This includes local updates of particle positions or charges without global updates of far field contributions, which is necessary for processes like ionization, dissociation or force fields like variable charge polarization models. This extension of the FMM to $\mathcal{O}(N)$ operational complexity is considered an important contribution to combine optimal algorithms with largest data locality. To improve the strong scaling properties and therefore the efficiency of simulations of systems on long time scales, the FMM will also be optimized to a small number of particles per core.

Also asynchronous runtime behavior will be explored and consequently synchronisation points in the hierarchy of tree levels will be removed, i.e. every process computes and sends information mostly independent from all the other processors by local computations and one-sided send-operations and traversing the hierarchical tree without global synchronization. To accelerate the computations the current memory model will be changed, so that necessary values like elements of rotation matrices are precomputed and stored. Also send-buffers will be reorganized such that it will not be necessary to calculate the number of buffer-elements exactly beforehand by analysing all data. In order to reduce the runtime, the necessary data-sorting will be changed into local sorts, which avoids global testing for data transfer. Local updates will be done between processes, neighbored on a Morton-curve, and particles will be sorted locally into data structures. To further improve the scaling of the FMM, workload information of the previous timestep will be utilized to balance the workload of the current time step. Additionally, if resources permit, geometries other than cubic unit cells (these are at the moment supported by ScaFaCoS) for 3d periodic boundary conditions will be implemented. These enhancements of the FMM will be done by the Jülich group, which already provides the basic FMM functionality in the ScaFaCoS library.

Task 3: Allow for alternative particle properties (MH/Göttingen)

The third task is the implementation of a scalable protocol that allows efficient force and potential calculation for alternative, time-dependent (lambda) properties of some of the particles.

For the proposed constant pH biomolecular simulations [DTGG11] this would be the charge, while for other applications it might be another property. In the constant pH case, for each titratable residue, typically two, sometimes more, alternative charge distributions have to be evaluated. From these charge alternatives, the force on the virtual lambda “particle” is calculated, which in turn defines the current protonation state of the residue. If we for sim-

plicity just look at a single proton at a specific site at the protein, then lambda describes the *continuous* coordinate ranging from fully protonated (lambda is zero) to fully deprotonated (lambda is one). A feature of the constant pH protocol is to allow for temporary (unphysical) mixed states—which would translate to only part of a proton being there—but which in turn allow the protonation state to react to changes in its electrostatic environment. Due to an adjustable potential barrier inbetween the physical states, the system effectively remains in intermediate states for short amounts of time only [DTGG11].

The work on the scalable version of the constant pH protocol can be begun right away, even if the FMM cannot yet feed back potentials for alternative charge distributions: We start by defining an interface to the library that will allow for charge alternatives. As soon as the interface specifications are worked out between Göttingen and Jülich, both groups can start to implement the interface as well as the features based on it. MH can therefore directly start to redesign the dynamics code to send positions and charges to the FMM interface and to accept back the potential and the forces. Until the FMM is ready to provide the potential and the forces for alternative charge distributions, for testing purposes, we will use the forces generated by the existing proof-of-concept const-pH protocol.

One of the challenges of this task is to improve the dynamics engine in that it must be able to use the FMM for large-scale simulations, while for small-scale simulations (up to ≈10,000 cores) still the traditional PME approach may be preferable. Ideally, the simulation engine should decide upon measured performance which algorithm to choose. Depending on what algorithm is used for a particular simulation, either the PME method in combination with the combined GROMACS nonbonded kernels for van der Waals and Coulomb forces have to be called, or, when in combination with a FMM, only kernels calculating the van der Waals interactions are needed. Care has to be taken here that for all combinations the results are correct. Therefore, thorough tests in which the numerical correctness of the combined algorithms is checked, have to be designed and carried out.

With the possibility of the enhanced multipole algorithm to evaluate various, only locally differing charge distributions, the electrostatic potential can be efficiently computed for each physical state (i.e. how the site is protonated) for each of n_λ titratable sites i, which in turn makes the dynamic adjustment of all i protonation states possible.

Task 4: Advance the FMM code for charge alternatives (TG/Göttingen, with the help of RH/Jülich)

In the fourth task, TG will advance the FMM such that it allows efficient calculation of locally alternative charge distributions for which the potentials and forces are communicated back to the main simulation engine.

This task will be started in the framework of the existing ScaFaCoS libary, but for a tight integration and full exploitation of all optimization potentials it is likely necessary at a later stage to integrate the long-range solver directly into the main application. For optimal efficiency, the following design considerations will be realized:

(i) On the lowermost layer, where the interactions between the particles are calculated as direct interactions (i.e. without multipole expansions) double evaluations have to be avoided. While in an naive implementation one could simply calculate all interactions in the volume of interest for $\lambda = 0$ and $\lambda = 1$, we want to perform alternative evaluations only for interactions between two atoms of which at least one is a lambda atom. Instead of one multipole for this volume on the lowermost layer (as in normal FMM), two multipoles will have to be constructed for each lambda atom i, one for $\lambda_i = 0$ and one for $\lambda_i = 1$. Typically, a box on the lowermost level will contain at most one lambda atom.

(ii) On the multipole layers up the hierarchy, once for each lambda atom i the $\lambda_i = 0$ and $\lambda_i = 1$ multipole alternatives have to be constructed, which is feasible since the main part of the computations in FMMs is needed on the lowermost level.

Task 4 can be started independently of tasks 3 and 2 since only the definition of the common interface between ScaFaCoS and GROMACS is needed.

Task 5: Integration (TG/Göttingen, MH/Göttingen, RH/Jülich)

Towards the end of the funding period, our three groups will jointly integrate and disseminate the unified algorithm(s) and scaling methods, and set up the collaboration framework.

In particular, the highly portable HPC software toolkit will be made available to the scientific community as an open-source, free-of-charge software. Similar to how GROMACS is distributed and administered, and with the help of the Stockholm group, MH will set up the framework that will allow scientists from all over the world to download the software as well as to contribute to it. Part of the collaboration framework will be a concurrent code versioning system, a bug-tracking system, a wiki for documentation, and a mailing list for communication between future users and developers.

TG will carry out one or two large-scale showcase and benchmark simulations to demonstrate the capabilities of the concepts for one or two particle simulation applications from the fields of, e.g., molecular dynamics, astrophysics and/or microfluidics.

Also, the proper documentation and publication of the results will be taken care of. We expect publishable results for each of the five tasks individually as well as for the project as a whole.

Risk assessment

We have identified two potential major risks for the project as a whole, and have developed proper fallback stategies.

First, it could turn out that the FMM implementation is not efficient enough to scale to a significantly larger number of cores than PME. Since, compared to PME, the FMM has a superior scaling behaviour with particle number N, this is very unlikely for the large N we want to address. If unforeseen essential difficulties arise during the optimization of the FMM scalability, we will resort to other efficient methods for long-range interactions, as e.g. hierarchical tree or multigrid methods. However, it is more likely that FMM scaling difficulties can be tackled by introducing an adaptive update scheme: In such a scheme, faraway multipoles would be updated less frequently than near multipoles. As, compared to system size, the particles do not move large distances in a few time steps, this adaptive scheme would not introduce significant force errors but has the power to dramatically reduce the communication volume, depending on the requested update frequencies.

Second, for the cases in which a large number of alternative charge distributions have to be calculated by the FMM, it could happen that the volume of the communicated multipole coefficients gets too large and therefore limits the scaling of the unified method. In that case, as a fallback project, a Monte-Carlo-based protonation/deprotonation scheme will be implemented, similar to the nonequilibrium switching protocol demonstrated in [KGdGZ11]. For MD systems under biological conditions, we do however not expect extremely large numbers of alternative charge distributions per volume.

2.4. Data handling

Code will be made publicly available.

2.5. Other information

None

2.6. Descriptions of proposed investigations involving experiments on humans, human materials or animals

No such investigations.

2.7. Information on scientific and financial involvement of international cooperation partners

3. Bibliography

[BCX+06] KJ Bowers, E Chow, H Xu, RO Dror, MP Eastwood, BA Gregersen, JL Klepeis, I Kolossvary, MA Moraes, FD Sacerdoti, JK Salmon, Y Shan, and DE Shaw. **Scalable algorithms for molecular dynamics simulations on commodity clusters**. *SC 2006 Conference, Proceedings of the ACM/IEEE*, pages 43–43, 2006.

[BNVL09] P Bjelkmar, P Niemelä, I Vattulainen, and E Lindahl. **Conformational Changes and Slow Dynamics through Microsecond Polarized Atomistic Molecular Simulation of an Integral Kv1.2 Ion Channel**. *PLoS Comp Biol*, 5(2):e1000289, 2009.

[BYQS03] D Barash, L Yang, X Qian, and T Schlick. **Inherent speedup limitations in multiple time step/particle mesh Ewald algorithms**. *Journal of Computational Chemistry*, 2003.

[CCD+05] DA Case, TE Cheatham, T Darden, H Gohlke, R Luo, KM Merz, A Onufriev, C Simmerling, B Wang, and RJ Woods. **The Amber biomolecular simulation programs**. *J. Comput. Chem.*, 26(16):1668–1688, 2005.

[CDDL09] DS Cerutti, RE Duke, TA Darden, and TP Lybrand. **Staggered mesh Ewald: an extension of the smooth particle-mesh Ewald method adding great versatility**. *Journal of chemical theory and computation*, 5(9):2322–2338, 2009.

[CHB+05] M Christen, PH Hünenberger, D Bakowies, R Baron, R Bürgi, DP Geerke, TN Heinz, MA Kastenholz, V Kräutler, C Oostenbrink, C Peter, D Trzesniak, and WF van Gunsteren. **The GROMOS software for biomolecular simulation: GROMOS05** . *J. Comput. Chem.*, 26(16):1719–1751, 2005.

[dGG01] BL de Groot and H Grubmüller. **Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF**. *Science*, 294(5550):2353–7, 2001.

[DH98] M Deserno and C Holm. **How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines**. *The Journal of Chemical Physics*, 1998.

[DYP93] T Darden, D York, and L Pedersen. **Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems**. *The Journal of Chemical Physics*, 1993.

[EHHT10] N Eicker, W Homberg, and S Höfler-Thierfeldt. **Towards exascale: Foundation of EIC and ECL**. *Inside*, 8:48, 2010.

[EPB+95] U Essmann, L Perera, ML Berkowitz, T Darden, H Lee, and LG Pedersen. **A smooth particle mesh Ewald method**. *J Chem Phys*, 103(19):8577–8593, 1995.

[EWGP06] M Eleftheriou, T Ward, M Giampapa, and M Pitman. **Blue Matter: Approaching the Limits of Concurrency for Classical Molecular Dynamics**. *IBM RC23956*, 2006.

[GHT96] H Grubmüller, B Heymann, and P Tavan. **Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force**. *Science*, 271(5251):997–9, 1996.

[GR87] L Greengard and V Rokhlin. **A fast algorithm for particle simulations**. *J Comput Phys*, 73(2):325–348, 1987.

[HBBF97] B Hess, H Bekker, HJC Berendsen, and JGEM Fraaije. **LINCS: A linear constraint solver for molecular simulations**. *J Comput Chem*, 18(12):1463–1472, 1997.

[Hes08] Berk Hess. **P-LINCS: A parallel linear constraint solver for molecular simulation**. *J Chem Theory Comput*, 4(1):116–122, 2008.

[LHvdS01] E Lindahl, B Hess, and D van der Spoel. **GROMACS 3.0: a package for molecular simulation and trajectory analysis**. *J Mol Model*, 7:306–317, 2001.

[LLF+08] OF Lange, N-A Lakomek, C Farès, G F Schröder, KFA Walter, S Becker, J Meiler, H Grubmüller, C Griesinger, and BL de Groot. **Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution**. *Science*, 320(5882):1471–5, 2008.

[PBW+05] JC Phillips, R Braun, W Wang, J Gumbart, E Tajkhorshid, E Villa, C Chipot, RD Skeel, L Kalé, and K Schulten. **Scalable molecular dynamics with NAMD**. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.

[SLPS09] R Schulz, B Lindner, L Petridis, and JC Smith. **Scaling of Multimillion-Atom Biological Molecular Dynamics Simulation on a Petascale Supercomputer**. *J Chem Theory Comput*, 5(10):2798–2808, 2009.

[SWK+07] JJ Sieber, KI Willig, C Kutzner, C Gerding-Reimers, B Harke, G Donnert, B Rammner, C Eggeling, SW Hell, H Grubmüller, and T Lang. **Anatomy and dynamics of a supramolecular membrane protein cluster**. *Science*, 317(5841):1072–6, 2007.

[THS+06] S Takamori, M Holt, K Stenius, EA Lemke, M Grønborg, D Riedel, H Urlaub, S Schenck, B Brügger, P Ringler, SA Müller, B Rammner, F Gräter, JS Hub, BL de Groot, G Mieskes, Y Moriyama, J Klingauf, H Grubmüller, J Heuser, F Wieland, and R Jahn. **Molecular anatomy of a trafficking organelle**. *Cell*, 127(4):831–46, 2006.

[vdSLH+05] D van der Spoel, E Lindahl, B Hess, G Groenhof, and AE Mark. **GROMACS: Fast, flexible, and free**. *J Comput Chem*, 26(16):1701–1718, 2005.

[w01] `http://www.mpibpc.mpg.de/home/grubmueller`.

[w02] `http://www.mpibpc.mpg.de/groups/de_groot`.

[w03] http://www.mpibpc.mpg.de/home/grubmueller/projects/MethodAdvancements/Gromacs/.

[w04] http://www.fz-juelich.de/SharedDocs/Pressemitteilungen/UK/EN/2011/11-11-15deep.html.

[w05] http://www.montblanc-project.eu.

[w06] http://www.eesi-project.eu.

[w07] http://www2.fz-juelich.de/jsc/scafacos.

[w08] http://www.fz-juelich.de/SharedDocs/Pressemitteilungen/UK/DE/2011/11-07-22algorithmus.html.

[w09] http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUDGE/JUDGE_node.html.

[w10] http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUGENE/JUGENE_node.html.

[w11] http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUROPA/JUROPA_node.html.

[w12] http://www.mpibpc.mpg.de/home/grubmueller/projects/MethodAdvancements/ConstpH.

[WDH10] H Wang, F Dommert, and C Holm. **Optimizing working parameters of the smooth particle mesh Ewald algorithm in terms of accuracy and efficiency**. *The Journal of Chemical Physics*, 133(3):034117, 2010.

[ZG10] M Zink and H Grubmüller. **Primary changes of the mechanical properties of Southern Bean Mosaic Virus upon calcium removal**. *Biophys J*, 98(4):687–95, 2010.

4. Requested modules/funds

4.1. Funding for staff

Since the whole project as well as most of its subprojects is quite ambitious, a rather high level of expertise will be needed both on the implementation as well as on the physics side for all the three main subprojects. Therefore, we request a postdoc position for the Jülich part and a postoc plus a Ph.D. position for the Göttingen group as justified below.

Graen, Timo (Taskset G) (Göttingen) This candidate's main assignment will be to adapt the fast multipole method to allow for alternative properties of some of the particles, and at a later stage to incorporate the concepts in a sample particle dynamics application. In addition, he will perform benchmark measurements to help to evaluate which long-range solver is best suited as well as carry out the final showcase simulation examples. This candidate must therefore have a background in informatics or physics. He must be a talented programmer, being able to program in C and at least to read and understand Fortran code. He must be familiar with the MPI library. This candidate must spend several research internships at our partners in Stockholm and Jülich, in total for a time of 6 months. Therefore, we apply for a 75% TVöD 13 position for this part. Timo Graen, who has just finished his Masters Thesis in the Göttingen Group, would be a qualified PhD canditate.

Hoefling, Martin (Taskset H) (Göttingen) The main responsibility of the candidate to realize taskset H will be two-fold: On the one hand, he must develop and implement simulation concepts supporting alternative particle properties, and on the other hand

he must coordinate future users and developers of the resulting library by setting up suitable software tools to support the collaboration framework. Candidates must have a strong background in physics, physical chemistry, biomolecular simulation as well as offer the relevant IT expertise. They must have profound programming skills in C, in which GROMACS is written, and must be familiar with parallel programming models, including MPI and OpenMP. The candidate must be familiar with performance monitoring and performance optimization of highly parallel applications. Moreover, the candidate must have experience in using, administering, and installing web-based tools needed for collaboration frameworks; among those tools are concurrent versioning systems like GIT or CVS, code review systems like Gerrit and software for managing mailing lists and operating a wiki. Since a PhD student would not provide this broad range of required expertise, we apply for a full TVöD 13 position for this part of the project, for which Dr. M. Hoefling would be a highly qualified candidate.

Halver, Rene (Taskset J) (Jülich) This candidate's main commitment will be the development and implementation of concepts for a general, extremely scalable solver for the potential and the forces in multi-processor particle simulations. Candidates for this ambitious task must have a strong mathematical background. Additionally, expert programming skills in C and Fortran are required, as the library to be build upon is written in these languages. Profound experience with parallel programming is needed (MPI/OpenMP), including in-depth parallel performance monitoring and optimization techniques. Therefore, we apply for a TVöD 13 position for this part. Rene Halver would be a possible candidate.

4.2. Funding for direct project costs

Equipment up to 10,000 €, software and consumables

Travel expenses

The GROMACS developers have an own interest in the success of this project and therefore are offering full cooperation. Since a collaboration including personal interaction is most effective, we request travel funds that will enable each year two meetings of one of the GROMACS developers with the three involved researchers, at one of the sites Jülich, Göttingen, or Stockholm, for several days. Therefore we require about 5000 € per year to cover hotel costs, flights, train tickets, and other transportation.

Detailed exemplary travel costs for first year: Trip 1: three candidates visit GROMACS team in Stockholm, flights Frankfurt–Stockholm each $\approx$ 350 €, hotel costs for 1 week $\approx$ 560 €/person, trains and other transportation $\approx$ 150 €/person, food 150 €/person, in total $3\times$ 1210 € = 3630 €.

Trip 2: One GROMACS developer visits Göttingen for a week, Jülich candidate joins. Flight Stockholm–Frankfurt $\approx$ 350 €, hotel costs for 4 days $\approx$ 275 €/person, trains and other transportation $\approx$ 150 €/person, food 85 €/person, in total 1370 €.

Visiting researchers

Project-related publication expenses

Funding to cover expenses for publications resulting from this project is needed. Since publishing in open-access journals is most beneficial for the community but comparatiely

expensive, we request 2000 € per year per group.

4.3. Funding for instrumentation

Not applied for.

5. Project requirements

5.1. Employment status information

5.2. First-time proposal data

5.3. Composition of the project group

Göttingen team: The Department of Theoretical and Computational Biophysics (Director: Helmut Grubmüller) [w01] [w02] has an internationally renowned track record in computational structural biology and pushing the limits of biomolecular simulations. The group has developed, established, and applied new molecular dynamics methods on high performance parallel computers during the past 15 years [GHT96] [SWK+07]. Through many international collaborations, the group has contributed to the elucidation of numerous functional mechanisms in proteins such as ATPase or aquaporins [dGG01]. The department operates an own high performance 9000-core parallel linux cluster, and has access to HPC (high performance computing) centers such as the Göttingen Computer Center (GWDG), the central Max Planck Computing Centre Garching (RZG), and the Jülich HPC centre. The department also participates in an active network of ongoing national and international collaborations with leading experts worldwide. Among those is a pertinent collaboration with the GROMACS developers which led to several optimizations of GROMACS's parallel PME algorithm [HKvdSL08] [KvdSF+07] (see also [w03]). Furthermore, the department has developed the λ-dynamics const-pH algorithm and implemented it in a proof-of-concept fashion in GROMACS [DTGG11] (see also [w12]).

Stockholm team: The division of Theoretical & Computational Biophysics (chair: Erik Lindahl) is internationally leading in research on membrane protein modelling and simulation, as well as molecular simulation methodology, and currently holds two ERC grants. The division is responsible for the development of a number of biomolecular applications, including GROMACS. The team has made critical discoveries in membrane protein biogenesis, protein folding, ion channel functionality [BNVL09], and numerous advances in molecular simulation development — GROMACS is for instance the engine of Folding@Home, and part of the DEISA and PRACE benchmark suites. The group has a private 4000-core Linux cluster, access to a local 36,000-core Cray XE6 at KTH, and strategic collaborations with access to the 705,000-core K-computer facility in Kobe. The division has a long-term collaboration on GPU acceleration with NVIDIA and scaling with teams at Oak Ridge. Over the last 15 years, the GROMACS code coordinated by David van der Spoel, Berk Hess & Erik Lindahl [LHvdS01] [vdSLH+05] has turned into one of the fastest available parallel molecular dynamics implementations [HKvdSL08] [HBBF97] [Hes08], and we are currently focusing on enabling molecular simulation applications on future Exascale resources [PLP+11].

Jülich team: The Jülich Supercomputing Centre (JSC) operates supercomputers of the highest class. Currently the first European Petaflop computer (Jugene [w10]) is operated by JSC, which is intended for capability computing of selected projects and large scale applications. Next year the successor system, a Blue Gene/Q with 8 racks will start operating.

The 200 TFlop parallel cluster architecture Juropa [w11] addresses a broader class of parallel applications. In addition a 122 nodes CPU/GPU cluster (Judge [w09]) with two GPU cards per node is operating. The present and future architectures will be available to the project for development and benchmarking. JSC has coordinated the BMBF funded project ScaFaCoS [w07] in which highly scalable algorithms for long range interactions in complex systems have been developed, optimized and integrated into a parallel library. JSC is partner in several exascale projects, like Mont Blanc [w05], DEEP [w04] or EESI [w06]. It is closely connected to hardware development via the Exascale Cluster Laboratory (ECL), a cooperation between Jülich, Intel and ParTec, and the Exascale Innovation Center (EIC) [EHHT10], a cooperation between IBM and Jülich. Large experience exists in parallel particle simulations, parallel algorithms and methods for long range interactions, especially fast multipole methods (FMM) [GR87]. The JSC team is just finishing a project where various electrostatic solvers will be offered as a freely available Fortran library. This library includes parallel versions of PME, P^3M, multigrid, and the fast multipole method (FMM) The JSC version of the FMM includes full error control, allowing a runtime optimization [Dac06] [Dac10] [KD11]. In combination with memory optimization and a most efficient parallelization the FMM was applied in a simulation of the biggest particle system up to now ($N > 3 \times 10^{12}$), taking full account of long range interactions between all particles [w08].

5.4. Cooperation with other researchers

5.4.1. Researchers with whom you have agreed to cooperate on this project

None

5.4.2. Researchers with whom you have collaborated scientifically within the past three years

Applicant	Cooperation partner
Helmut Grubmüller	**International** Herman Berendsen, University of Groningen, the Netherlands José Carrascosa, University of Madrid, Spain Mathias Gautel, King's College London, UK Elisha Haas, BIU Jerusalem, Israel Berk Hess, Royal Institute of Technology, Sweden Brent P. Krueger, Hope College, Holland, USA Erik Lindahl, University of Stockholm, Sweden Siewert-Jan Marrrink, University of Groningen, the Netherlands Daniel Müller, ETH Zurich, Switzerland Ben Schuler, ETH Zurich, Switzerland Gert Vriend, Radboud University Nijmegen, the Netherlands Dixon J. Woodbury, Brigham Young University, Provo, USA Ulrich Zachariae, University of Edinburgh, UK **National** Bernd Abel, University of Leipzig Jörg Enderlein, University of Göttingen Ralg Ficner, University of Göttingen Wolfgang Fischle, MPI for Biophysical Chemistry, Göttingen Hermann Gaub, LMU München Dirk Görlich, MPI for Biophysical Chemistry, Göttingen Christian Griesinger, MPI for Biophysical Chemistry, Göttingen Bert de Groot, MPI for Biophysical Chemistry, Göttingen Stefan Hell, MPI for Biophysical Chemistry, Göttingen Claudia Höbartner, MPI for Biophysical Chemistry, Göttingen Reinhard Jahn, MPI for Biophysical Chemistry, Göttingen Stefan Jakobs, MPI for Biophysical Chemistry, Göttingen Adam Lange, MPI for Biophysical Chemistry, Göttingen Reinhard Lührmann, MPI for Biophysical Chemistry, Göttingen Thomas Meier, MPI for Biophyiscs, Frankfurt Marcus Müller, University of Göttingen Robert Preissner, Institute for Molecular Biology and Bioinformatics, Berlin Marina Rodnina, MPI for Biophysical Chemistry, Göttingen Gunnar Schröder, Forschungszentrum Jülich Dirk Schwarzer, MPI for Biophysical Chemistry, Göttingen Claus Seidel, University of Düsseldorf Holger Stark, MPI for Biophysical Chemistry, Göttingen Robert Tampé, University of Frankfurt Hans-Jürgen Thiesen, Institute for Immunology, Rostock

Applicant	Cooperation partner
Holger Dachsel	**International** Jeff Hammond, Argonne National Laboratory, Argonne, USA **National** Pablo Garcia Risueno, Humboldt University of Berlin Ulf Saalmann, MPI for the Physics of Complex Systems, Dresden

Applicant	Cooperation partner
Berk Hess	**International** Anna-Karin Tornberg, KTH: Improving particle-mesh Ewald electrostatics methods. Gustav Amberg, KTH: Coupling continuum and molecular models for contact line dynamics. Barcelona Supercomputing Center: Analyzing and improving parallel simulations. Roland Schulz, ORNL, USA: Improving performance of massively parallel simulations. **National** Elmar Bonacurso, CSI, TU Darmstadt: Understanding complex wetting phenomena

5.5. Scientific equipment

The Max-Planck Society invests approx. 200 kEUR per year to maintain the existing cluster at the MPI for Biophysical Chemistry with a current configuration consisting of 800 nodes, ca. 10 500 processor cores. The newest types of nodes are 100 nodes with 48 cores AMD Magny-Cours, 1.9 GHz and 32 GB-RAM interconnected by Gigabit-Ethernet. The total compute power is about 70 Tera FLOPS.

The Research Centre Juelich operates two supercomputer, the IBM Blue/Gene P Jugene and the Linux Cluster Juropa.

Jugene consists of 73728 nodes and 294912 cores. The processor at Jugene is a Power PC 450 with a clock speed of 850 MHz. The entire system achieves an overall peak performance of 1 Petaflops and provides 144 TB of main memory.

Juropa consists of 2208 nodes and 17664 cores and is equipped with the Intel Xenon X5570 (Nehalem-EP) processor running at a clock speed of 2.93 GHz. The overall peak performance of Juropa is at 207 Teraflops. The systems provides 53 TB of main memory.

5.6. Project-relevant interests in commercial enterprises

None

6. Additional information

None

Part C.
Appendices

7. List of publications

The five most relevant publications of each applicant are listed in their respective Curriculum Vitae.

8. References

None

9. CV's

Helmut Grubmüller

Honorary Professor Dr. Helmut Grubmüller
Director for the Max Planck Institute for Biophysical Chemistry

Personal data

Address	Max Planck Institute for Biophysical Chemistry Department of Theoretical and Computational Biophysics Am Fassberg 11 37077 Göttingen
Phone	05 51 / 2 01 23 01
Fax	05 51 / 2 01 23 02
eMail	hgrubmu@gwdg.de
Birth	1965
Gender	male

Research topics

Theoretical Biophysics

Academic education and scientific degrees

2002	Habilitation, Physics, University of Göttingen
1994	Conferral of a doctorate, Physics, Technical University of Munich, Prof. P. Tavan
1985–1990	Studies in Physics, Technical University of Munich, Diploma, Prof. K. Schulten

Scientific career

2009–2010	Managing Director of the Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
2005–	Honorary Professor for Physics at the Georg-August-University Göttingen, Germany
2003–	Director, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany: Head of the Theoretical and Computational Biophysics Department
2003	Associate Professor for Biomolecular Sciences at the École Polytechnique Fédérale de Lausanne (EPFL), France
1998–2003	Head of the Theoretical Molecular Biophysics Group at the Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
1997	EMBO fellow at the Institute for Molecular Biology and Biophysics, Federal Institute of Technology (ETH) Zürich, Switzerland
1994–1998	Postdoctoral assistant at the Theoretical Biophysics Group, University of Munich, Germany

Awards and professional activities

2008–	Elected Member of the IUPAP Commission on Biological Physics
2007–	Spokesman of the International Max Planck Research School *Physics of Biological and Complex Systems*
2007–	Steering Committee of the Max Planck Digital Library
2007–	Elected Member of the Max Planck Society Perspective Commission
2005–	Advisory Board Member, Fritz Haber Minerva Research Center for Molecular Dynamics
2005–	Executive Committee Member, European Biophysical Societies' Association (EBSA)
2006–2011	Scientific Advisory Board Member, Göttingen Computer Center (GWDG)
2003–2012	Elected Member of the Reviewing Board, German Research Foundation
2002–	Biophysical Journal, Editorial Board
1997–2000	German Biophysical Society (DGfB, Panel Member)

Five most relevant publications

1. OF Lange, NA Lakomek, C Farès, GF Schröder, KFA Walter, S Becker, J Meiler, H Grubmüller, C Griesinger, BL de Groot. **Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution.** *Science* (2008) 320: 1471–1475
2. M Hoefling, N Lima, D Haenni, CAM Seidel, B Schuler, H Grubmüller. **Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach.** *PLoS One* (2011) 6: e19791
3. JJ Sieber, KI Willig, C Kutzner, C Gerding-Reimers, B Harke, G Donnert, B Rammner, C Eggeling, SW Hell, H Grubmüller, T Lang. **Anatomy and Dynamics of a Supramolecular Membrane Protein Cluster.** *Science* (2007) 317:1072–76
4. EM Puchner, A Alexandrovich, AL Kho, U Hensen, LV Schäfer, B Brandmeier, F Graeter, H Grubmüller, HE Gaub, M Gautel. **Mechanoenzymatics of titin kinase.** *Proc Natl Acad Sci USA* (2008) 105: 13385–13390
5. C Kutzner, H Grubmüller, BL de Groot, U Zachariae. **Computational Electrophysiology: The Molecular Dynamics of Ion Channel Permeation and Selectivity in Atomistic Detail.** *Biophys J* (2011) 101: 809–817

Holger Dachsel

Dr. Holger Dachsel
Forschungszentrum Jülich

Personal data

Address	Forschungszentrum Jülich Institute for Advanced Simulation Jülich Supercomputing Centre, Mathematics Division Leo-Brandt Str. 52425 Jülich
Phone	02 46 1 / 61 65 69
Fax	02 46 1 / 61 66 56
eMail	h.dachsel@fz-juelich.de
Birth	1961
Gender	male

Research topics

Software development for massively parallel systems
Strategies and algorithms in parallel computing
Data compression methods
Multipole Methods (FMM, CFMM)

Academic education and scientific degrees

1992	Conferral of a doctorate, University of Leipzig
1981–1986	Studies in Chemistry, University of Leipzig, Diploma

Scientific career

2000–	Research Scientist at Research Centre Jülich, ZAM, Germany
1999–2000	Research Scientist, Q-Chem, Inc. Pittsburgh, USA
1998–1999	Research Scientist at the Department of Theoretical Chemistry, University of Amsterdam, the Netherlands
1995–1998	Post Doctoral Fellow at the Department of Theoretical Chemistry, Universtiy of Vienna, Austria
1987–1992	Assistant Scientist at the Department of Informatics at University of Leipzig, Germany
1986–1987	Assistant Scientist at the Department of Physics at Universtiy of Leipzig, Germany

Awards and professional activities

1998	Supercomputing 98 Best Overall Paper Award, Orlando, USA

Five most relevant publications

1. H Dachsel. **An error-controlled fast multipole method.** *J. Chem. Phys.* (2010) 132:119901
2. H Dachsel. **Fast and accurate determination of the Wigner rotation matrices in the fast multipole method.** *J. Chem. Phys.* (2006) 124:144115
3. H Dachsel, RJ Harrison, DA Dixon. **Multireference Configuration Interaction Calculations on Cr–Cr: Passing the One Billion Limit in MRCI/MRACPF Calculations.** *J. Phys. Chem.* (1999) 152
4. T Müller, SS Xantheas, H Dachsel, RJ Harrison, J Nieplocha, R Shepard, GS Kedzora, H Lischka. **A systematic ab initio investigation of the open and ring structures of ozone.** *Chem. Phys. Lett.* (1998) 293:72
5. H Dachsel, H Lischka, R Shepard, J Nieplocha, RJ Harrison. **A massively parallel multireference configuration interaction program: The parallel COLUMBUS Program.** *J. Comput. Chem.* (1997) 18:430

Berk Hess

Associate Prof. Dr. Berk Hess
Royal Institute of Technology

Personal data

Address	Department of Theoretical Physics School of Engineering Sciences Royal Institute of Technology SE-100 44 Stockholm
Phone	+46 / 8 55 37 80 27
eMail	hess@kth.se
Birth	1972
Gender	male

Research topics

Physical chemistry of hydration
Microfluidics at interfaces
Developing algorithms for efficient molecular simulation
Developing algorithms for free energy calculations

Academic education and scientific degrees

2002	Conferral of a doctorate, Mathematics, University of Groningen, the Netherlands, Prof. H.J.C. Berendsen
1995	Studies in Applied Mathematics, University of Groningen, the Netherlands, Masters

Scientific career

2011 –	Associate Professor (Universitetslektor) at the Department for Theoretical Physics of the Royal Institute of Technology, Stockholm, Sweden
2009 – 2011	Senior researcher at the Center for Biomembrane Research, Stockholm, Sweden
2002 – 2009	Postdoc at the Max Planck Institute for Polymer Physics in Mainz, Germany

Awards and professional activities

2008 – 2011	Resource funding as principal investigator

Five most relevant publications

1. P Bjelkmar, P Larsson, MA Cuendet, B Hess and E Lindahl. **Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models.** *J. Chem. Theory Comput.* (2010) 6:459–466
2. B Hess, C Kutzner, D van der Spoel, E Lindahl. **GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.** *J. Chem. Theory Comput.* (2008) 4:435–447
3. B Hess. **P-LINCS: A parallel linear constraint solver for molecular simulation.** *J.* Chem. Theory Comput. (2008) 4:116–122
4. B Hess, D van der Spoel. **GROMACS–The road ahead.** *C*omp. Molec. Sci. (2011) 1:710–715
5. B Hess, H Bekker, HJC Berendsen, JGEM Fraaije. **LINCS: A linear constraint solver for molecular simulations.** *J. Comp. Chem.* (1997) 18:1463–1472