

LONG TERM PROGRAM PROPOSAL  
Navigating Chemical Compound Space for  
Materials and Bio Design

at the  
Institute for Pure and Applied Mathematics  
University of California Los Angeles  
in  
Academic Year 2010/2011

Main organizing team (all confirmed)

Jean-Loup Faulon, University of Evry/Genopol Paris, France  
William Hart, Sandia National Laboratories  
Kendall Houk, University of California Los Angeles  
Peter Jones, Yale University  
Steven Lustig, DuPont Central R&D  
Tamar Seideman, Northwestern University  
Mark Tuckerman, New York University  
Anatole von Lilienfeld (Chair), Sandia National Laboratories

**Abstract**

The design of chemical compounds with specific physical, chemical, or biological properties is a central goal of many fundamental as well as industrially relevant research fields. Due to the combinatorial nature of chemical compound space, however, even *in silico* a systematic screening for interesting properties is beyond any current capacity. This long-term program aims to unite more experienced as well as younger scientists from various fields in order to establish a wide and interdisciplinary scientific community which will represent the emerging field of computational compound design.

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	The fundamental problem . . . . .	2
1.2	Outside funding . . . . .	3
<b>2</b>	<b>Optimization, search, and graph-theoretical algorithms for CCS</b>	<b>4</b>
2.1	Motivation . . . . .	4
2.2	Organizers (all confirmed) . . . . .	4
2.3	Possible Additional Speakers or Participants . . . . .	5
<b>3</b>	<b>Physical frameworks for sampling CCS</b>	<b>5</b>
3.1	Motivation . . . . .	5
3.2	Organizers (all confirmed) . . . . .	5
3.3	Possible Additional Speakers or Participants . . . . .	6
<b>4</b>	<b>Drugs and bio-molecular design in CCS</b>	<b>6</b>
4.1	Motivation . . . . .	6
4.2	Organizers (all confirmed) . . . . .	6
4.3	Possible Additional Speakers or Participants . . . . .	7
<b>5</b>	<b>Materials design in CCS</b>	<b>7</b>
5.1	Motivation . . . . .	7
5.2	Organizers (all confirmed) . . . . .	7
5.3	Possible Additional Speakers or Participants . . . . .	7

## 1 Introduction

### 1.1 The fundamental problem

Chemical compound space (CCS) is a combinatorial set which encompasses all stable chemical compounds. One can view CCS as the high dimensional space spanned by all the possible combinations (stoichiometries) and configurations (isomers) of electrons and atomic nuclei. It provides a natural framework in which to construct rigorous mathematical tools for the development of direct and inverse quantitative structure-property relationships (QSPR). Systematically generating high quality QSPRs will provide huge advances in the virtual design and purposeful engineering of new materials, chemicals, drugs, proteins, crystals, liquids, and even organisms. In very general terms, compound design efforts attempt to invert the direct mapping of a given compound (defined by its Hamiltonian  $\hat{\mathcal{H}}$ ) to the observable of interest ( $\mathcal{O}$ ),

$$\left(\hat{\mathcal{H}} \mapsto \mathcal{O}\right) \longleftrightarrow \left(\mathcal{O}_0 \mapsto \hat{\mathcal{H}}_0\right). \quad (1)$$

Usually, once some knowledge, experience, and information is acquired of how  $\hat{\mathcal{H}}$  needs to be modified such that the observable of interest approaches a desired

property value,  $\mathcal{O}_0$ , one attempts to invert the initial direct mapping correspondingly in order to identify the target system,  $\mathcal{H}_0$ . Consequently, the underlying interest in the mapping of a given system ( $\mathcal{H}$ ) to its properties is frequently motivated by their possible tuning through the compounds' compositional variation in its CCS, the CCS being defined by the details of the mapping procedure. Various scientific communities, such as materials design or drug discovery programs, deal with this inverse question on a daily basis, leveraging historically grown experimental insights as well as the advances made in theoretical and computational sciences which involve diverse areas such as statistical mechanics, liquid and solid state physics, quantum chemistry, graph theory, molecular physics, condensed matter physics, optimization algorithms, data mining, and others.

Ideally, analytic structure property relationships can be identified and inverted, usually in terms of a numerical screen or optimization, aiming to answer the inverse question namely, under such and such conditions, which compound will exhibit a certain desired property? Albeit successful in some cases, this approach is not yet generally viable. Also, it is unclear how to generally extract simplified relationships from the attempts to perform the direct mapping of system to property, as they are present in the various research fields at various levels of theory and accuracy.

We therefore propose to bring together diverse scientific communities which are involved with the design question within a strongly mathematical environment where they can communicate current bottlenecks among each other and in particular towards the applied mathematics community. Ideally, this program will lead to fruitful collaborations where all participants benefit largely from mathematical insights on their specific optimization and design problems, leading to new and unconventional solutions for old and outstanding problems.

Four one week workshops are being proposed for this long-term program. The four workshops would deal with the Mathematics and computer science of CCS (Section 2), the physics modeling background for sampling CCS (Section 3), bio-molecular and drug discovery in CCS (Section 4), and materials design in CCS (Section 5). Eventually, the culminating Workshop at Lake Arrowhead, as well as reunions in later years, will revive collaborations between long-term participants as well as selected attendants of individual workshops.

## 1.2 Outside funding

Additional funding resources from Industry, National Laboratories, or 'Computational Design Centers' are being sought. So far, the following has been obtained.

### *Sandia National Laboratories*

H. Eliot Fang, Chair of the symposium on "Computational Materials Design via Multiscale Modeling" at the Materials Research Society meeting in December 2008 in Boston, and Deputy and Technical Assistant to the Office of Science & Technology and Research Foundations at Sandia, believes that the idea of having Sandia personnel to spend time at IPAM is possible and doable, and that Sandia

has an interest in positioning itself in this emerging area. In addition to joining an organizing team he is also advertising Sandia's co-sponsoring of this event among higher management.

Lower level managers of the Computer Science Research Institute (CSRI) at Sandia National Laboratories are all supportive of the efforts to develop an IPAM long term program. Unfortunately, its support can only be intangible; at the moment there are no CSRI funds available to directly contribute since CSRI has zero formal budget in fiscal year 2009. CSRI will, however, help with enabling interested staff to participate. This includes adjusting schedules and work assignments, as needed, and recognizing their extra efforts in engaging in the IPAM program at merit review.

*DuPont*

Steven Lustig, Research Scientist at DuPont, did not only offer to join the main organizing team but also to ask management at DuPont for sponsorship of this program.

## 2 Optimization, search, and graph-theoretical algorithms for CCS

### 2.1 Motivation

This workshop focuses on the computational science aspects when navigating, exploring, or optimizing object functions in CCS. Data-mining, search-, organization-, optimization-, and classification-algorithms in the group of chemical compounds populating CCS require particular computational science efforts. The computational cost of evaluating the object function, for example, is strongly dependent on the instantaneous combination of compositional variables which define a compound. Also, depending on the definition of the object function, CCS can be rendered continuous through arbitrary paths – offering additional and unexplored degrees of freedom for improving the efficiency of optimization algorithms. This workshop would also gather experts in the field of mathematical chemistry to report on their recent research efforts regarding molecular graph-theory and topologies, cheminformatics, quantitative structure-property relationships, and related topics. Mathematical chemistry deals with topics such as the mathematical study of isomerism and the development of topological descriptors or indices which find application in quantitative structure-property relationships; chemical aspects of group theory, which finds applications in stereochemistry, quantum chemistry, and crystallography; and topological aspects of chemistry.

### 2.2 Organizers (all confirmed)

Mauro Maggioni, Duke University

Cynthia Phillips, Sandia National Laboratories

Jean-Louis Raymond, University of Bern, Switzerland

Jean-Paul Watson, Sandia National Laboratories

### **2.3 Possible Additional Speakers or Participants**

Bruce Hendrickson Sandia National Laboratories (confirmed)  
Krishna Rajan, Rensselaer Polytechnic Institute  
Yannis G. Kevrekidis, Princeton University  
Pierre Baldi, University of California Irvine  
Tudor Oprea, University of New Mexico  
Alex Tropsha, University of North Carolina at Chapel Hill  
Ovidiu I. Ivanciuc, University of Texas Medical Branch, Galveston  
Alexandru T. Balaban, Texas A&M University, Galveston  
Rajarshi Guha, Indiana University  
Dick Cramer and Bob Clark, Tripos, St Louis & Santa Fe  
Jean-Philippe Vert, Polytechnique Paris, France  
Chris Steinbeck, European Bioinformatics Institute  
Andreas Bender, University of Amsterdam/Leiden, Netherlands

## **3 Physical frameworks for sampling CCS**

### **3.1 Motivation**

Generation of the grand-canonical ensemble distribution using classical statistical mechanical approaches such as grand-canonical Monte Carlo is relatively straightforward. This workshop, however, will focus on the grand-canonical aspect related to chemical “transformations” (or “alchemical transformations”) corresponding to walking through CCS. Depending on the level of theory, such moves can be performed within coarse-grained, classical force field, or quantum mechanics based models, and correspond either to the realistic chemical transformation of one compound into each other, or rely on an arbitrary “ $\lambda$ ”-paths, such as in thermodynamic integration methods. Experts will meet, report, and discuss progress of various research fields including redox processes, isomolar Monte Carlo or molecular dynamics simulation in the grand canonical ensemble, generation of the grand-canonical ensemble distribution using classical statistical mechanical approaches such as grand-canonical Monte Carlo, conceptual density functional theory, and others.

### **3.2 Organizers (all confirmed)**

Paul Ayers, McMaster University  
David Beratan and Weitao Yang, Duke University  
Ed Maginn, University of Notre Dame  
Peter Politzer, University of New Orleans  
Markus Reiher, ETH Zuerich, Switzerland

Maria-Lore Sulpizi, Cambridge University, United Kingdom  
Aidan Thompson, Sandia National Laboratories

### **3.3 Possible Additional Speakers or Participants**

Paul Geerlings, Free University of Brussels  
Christoph A. Haselwandter, Massachusetts Institute of Technology  
Joachim Blumberger, Cambridge University, United Kingdom (confirmed)  
Marcus G. Martin Useful Bias Incorporated  
Rodolphe Vuilleumier, Université Pierre et Marie Curie Paris, France  
Michiel Sprik, Cambridge University, United Kingdom  
Bruce Tidor, Massachusetts Institute of Technology  
Paul Mezey, Memorial University Newfoundland & Labrador, Canada

## **4 Drugs and bio-molecular design in CCS**

### **4.1 Motivation**

Biological space can be thought of as the unabridged complement of gene products. Since it is defined by an organisms genomic inventory it is, although large, finite. By contrast the combinatorial manifold of molecular compounds means that that CCS is essentially boundless. Molecular biology and modern genomics offer unforeseen opportunities to explore CCS genetically and through enzyme catalysis, usually for the purpose of drug discovery. The intersection of biological space and chemical space, however, forms the backdrop to the arena of drug discovery. From an atomistic point of view, Molecular dynamics and Monte Carlo schemes, frequently also combined, through multi-scaling methods, with continuous and quantum chemical models, have also evolved to become powerful tools for tackling biochemical problems which are at the focus of research fields such as rational drug design, bio-mimetics, enzymatic catalysis, bio-inspired ligands in organo- and bio-inorganic metal complexes. This workshop will bring together experts from all these fields.

### **4.2 Organizers (all confirmed)**

David Baker, University of Washington Washington  
Cecilia Clementi, Rice University  
William Jorgensen, Yale University  
Michael Klein, University of Pennsylvania  
Ursula Röthlisberger, EPF Lausanne, Switzerland  
Jeffery G. Saven, University of Pennsylvania

### 4.3 Possible Additional Speakers or Participants

Christopher A. Lipinski Melior Discovery, Inc.  
Hugo Kubinyi, Germany  
Gerd Folkers, ETH Zuerich, Switzerland  
Ute Roehrig, Ludwig Cancer Research Centre  
Paolo Carloni, International School for Advanced Studies (SISSA), Trieste  
Chris Voight, University of California San Francisco  
VanGunsteren, ETH Zuerich, Switzerland  
Walter Thiel, Max-Planck-Institut für Kohlenforschung, Germany  
Arieh Warshel, University of California Los Angeles  
Yingkai Zhang, New York University  
Jiali Gao, University of Minnesota  
Sameer Varma, Sandia National Laboratories

## 5 Materials design in CCS

### 5.1 Motivation

Investigating and predicting materials behavior can be achieved through various ways reaching from numerically intensive high performance computing applications of multi-scale methods to simple analytical effective models. How do materials simulation methods need to be altered in order to address the question of not only describing but also designing materials? A large manifold of materials' properties could be addressed that way. For this workshop materials researchers will meet to discuss progress on catalyst design, meta-materials, heat-transfer fluid design, ionic liquid design, designer materials, crystal engineering etc.

### 5.2 Organizers (all confirmed)

Ralf Drautz, Interdisciplinary Centre for Advanced Materials Simulation, Germany  
Eliot Fang, Sandia National Laboratories  
Kirsten Fichthorn and Vincent Crespi, Center for Computational Materials Design, PennState/GeorgiaTech  
Graeme Henkelman, University of Texas Austin  
Artem Oganov, ETH Zuerich, Switzerland

### 5.3 Possible Additional Speakers or Participants

Clemence Corminboef, EPF Lausanne, Switzerland  
Jens Norskov, Center for Atomic-scale Materials Design, Denmark  
Koichi Yamashita, Department of Chemical System Engineering, University of Tokyo, Japan

Dennis Andrienko, Max-Planck-Institute for Polymer Research Mainz, Germany  
Sharon Glotzer, University of Michigan  
Gerbrand Ceder, Department of Materials Science and Engineering, Massachusetts  
Institute of Technology  
Joachim Sauer, Humboldt University, Germany  
Robin Hayes, New York University