Perspective

Research Challenges, Opportunities and Synergism in Systems Engineering and Computational Biology

Christodoulos A. Floudas
Dept. of Chemical Engineering Princeton University, Princeton, NJ 08544

DOI 10.1002/aic.10620
Published online June 14, 2005 in Wiley InterScience (www.interscience.wiley.com).

Introduction

During the last three decades, the research area of systems engineering has emerged as a domain of fundamental importance and major impact within chemical engineering, as well as a cornerstone area in interdisciplinary research initiatives with computer science, operations research, applied mathematics, materials and life sciences. This is attributed to the unique characteristics of systems engineering which are the combination of analysis and synthesis for the design, optimization, and operation of processes and products. The product and process discovery research efforts are founded on fundamental advances in mathematical modeling, optimization theory and algorithms, and insights derived either from existing operations and/or from biology, chemistry, and physics. The fundamental advances are epitomized through new theoretical, algorithmic, and modeling frameworks for (a) mixed-integer linear and nonlinear optimization, (b) deterministic global optimization, and (c) dynamic simulation and optimization. The proposed modeling and optimization approaches have multiscale applications ranging from macroscopic to mesoscopic to microscopic systems, and which provide a natural and fundamental link between systems engineering, computational chemistry, computational biology and systems biology. Approaches based on mixed-integer linear and nonlinear optimization which were typically identified with process synthesis, scheduling and planning applications, have entered the domains of gene regulatory networks, metabolic networks, signal transduction networks, beta-sheet topology prediction in proteins, de novo peptide and protein design, DNA recombination, phase problem in X-ray crystallography, side chain optimization in protein prediction, peptide identification via tandem mass spectroscopy. Approaches based on deterministic global optimization associated with process design, synthesis, scheduling, and pooling/blending applications, are now in the main stream of product design, structure prediction in protein folding, dynamics of protein folding, NMR protein structure refinement, and de novo protein design. Approaches based on dynamic models and large scale optimization which were identified with process models and their applications, are suited for metabolic and signal transduction networks. As a result, a synergism between systems engineering, computational biology, and systems biology has evolved gradually, opened new research avenues and has reached the stage where fascinating research contributions address important questions in computational biology with methods and tools from systems engineering which combine mathematical rigor with key biological insights.

This article provides a perspective on the challenges and opportunities that emerge from the fundamental developments in the research fields of systems engineering and computational biology. The advances in the areas of deterministic global optimization and process scheduling are introduced first, followed by their respective research opportunities. The revolution of genomics is discussed next with a particular focus on the advances and challenges in the areas of structure prediction in protein folding, de novo peptide and protein design, and peptide and protein identification via tandem mass spectroscopy. This article is based on material presented as an invited talk at the session on "The Future of Chemical Engineering Research III" during the 2004 annual AIChE meeting.

Deterministic Global Optimization

Global optimization addresses the computation and characterization of global minima and maxima of a nonconvex objective function subject to a nonconvex set of equality and inequality constraints. There are five primary objectives in deterministic global optimization: (1) determine an epsilon-global minimum with theoretical guarantee; (2) calculate valid and tight lower and upper bounds on the global minimum; (3) identify an ensemble of good quality local solutions close to the global minimum; (4) enclose all solutions of the equality and inequality constraints; and (5) prove whether a constrained nonlinear optimization problem is feasible or infeasible. It is important to emphasize that even though objectives (1), (4), and (5) are the ultimate targets from the mathematical analysis viewpoint, it is objectives (2) and (3) that are of major importance and greater potential impact in a variety of engineering applications.

Email C. A. Floudas at floudas@titan.princeton.edu

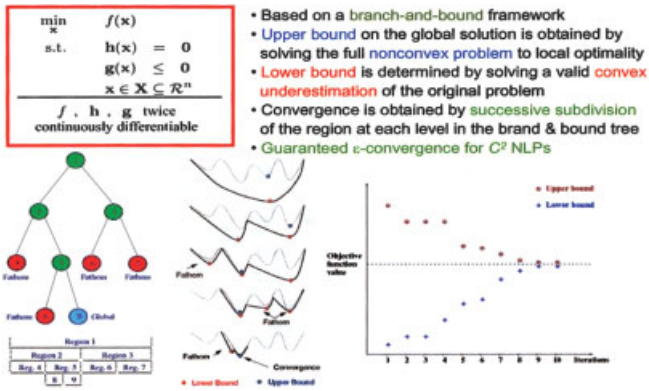

Figure 1. Important elements of deterministic global optimization approaches.

During the last two decades, we have experienced an explosive interest and growth in developing new theoretical and algorithmic frameworks for global optimization, as well as their applications to important scientific problems. From the historical global optimization perspective, there has been a two-order of magnitude increase in the number of publications since the early 1980s. It is now established that global optimization exhibits ubiquitous applications that span all branches of engineering, applied sciences, and sciences.[1] In this century, several textbooks addressing a diverse set of topics in global optimization were published (e.g., Floudas[1]; Horst, Pardalos, and Thoai[2]; Tawarmalani and Sahinidis[3]; Zabinsky[4]). Neumaier[5] surveyed constrained global optimization and continuous constraint satisfaction problems with an emphasis on interval arithmetic. In a recent review article, Floudas et al.[6] provided a detailed account of the research progress in deterministic global optimization.

Figure 1 presents the fundamental components of deterministic global optimization methods. These include (1) the generation of convex underestimators, (2) the partitioning of the continuous domain into subdomains, based on the principles of the divide and conquer (i.e., Branch and Bound) approach, and (3) the generation of lower bounding and upper bounding sequences which converge within epsilon in a finite number of steps. In the following section, a summary of important advances in deterministic global optimization along with representative references is presented first, followed by an assessment of the current status along with the posed challenges and opportunities.

Advances in Deterministic Global Optimization. The important advances belong to the following six categories: (a) convex envelopes and convex under-estimators; (b) twice continuously differentiable constrained nonlinear optimization problems; (c) mixed-integer nonlinear optimization problems; (d) bilevel nonlinear optimization problems; (e) optimization problems with differential-algebraic equations; (f) grey-box and factorable models; and (g) enclosure of all solutions. In (a), convex envelopes were proposed for odd degree univariate monomials,[7] for trilinear monomials with positive or negative or mixed domains,[8,9] for fractional terms over a unit hypercube,[10] and for edge-concave functions.[11] Convex extensions were introduced for lower semicontinuous functions.[12] Convexification techniques were introduced for general twice continuously differentiable functions,[13-16] for fractional terms,[17-19] for trigonometric functions,[20] generalized polynomials,[21] and through the reformulation-linearization technique.[22] In (b), new classes of global optimization algorithms were introduced, such as the αBB,[14,15] p-αBB,[23] generalized-αBB,[16,24] interval analysis techniques,[25-28] and terrain methods.[29,30] In (c), a variety of theoretical and algorithmic approaches for mixed-integer nonlinear optimization problems were proposed. These include disjunctive programming techniques,[31-33] the extended cutting plane approach,[34-36] the smin-αBB and gmin-αBB approaches,[37] decomposition-based approaches,[38,39] and the branch and reduce optimization navigator, Baron.[40] In (d), global optimization methods were introduced for bilevel nonlinear models,[41,42] for bilevel linear-quadratic models,[43] and for bilevel mixed-integer optimization models.[44] In (e), global optimization methods were proposed for dynamic parameter estimation models and optimal control problems[45-52], and for hybrid discrete/continuous dynamic models.[53-56] In (f), approaches were introduced based on interval analysis,[25,57] response surface methods,[58] radial basis functions,[59] and nonfactorable constraints.[60] In (g), methods for the enclosure of all solutions were introduced using interval analysis[61-63] and using the αBB approach.[64] As result of these advances, the current status in deterministic global optimization can be regarded as having great success for the development of new theories and algorithms, but with applications restricted to small and medium size problems.

Challenges in Deterministic Global Optimization. The research opportunities and challenges in global optimization include (a) developing improved convex underestimation techniques for general functions; (b) introducing new theoretical results for the derivation of convex envelopes or approximate convex envelopes for general multi-linear functions, general twice continuously differentiable functions, and trigonometric functions; (c) addressing medium to large-scale twice continuously differentiable nonlinear optimization models, such as pooling problems; (d) developing methods for medium and large-scale mixed-integer nonlinear optimization models which arise in process synthesis, design, planning and scheduling, and generalized pooling problems; (e) introducing new theoretical results and algorithms for dynamic optimization models and semiinfinite programming problems; (f) developing new approaches for grey-box optimization, (g) introducing new theories and algorithms for multi-level nonlinear optimization, and (h) developing new global optimization frameworks for nondifferentiable optimization models.

Process Operations: Scheduling

Multiproduct and multipurpose plants that operate in batch, semicontinuous, and continuous mode manufacture a variety of products through several sequences of operations, denoted as recipes. At the same time, the products share the available pieces of production equipment, inventory and storage units, intermediate materials and raw materials. Process scheduling addresses the optimal assignment of tasks to units over the allotted time horizon in such complex operations. The typical data provided in process scheduling problems include data on the production (i.e., tasks and sequences for each product); the available production units and their capacities; the initial, intermediate, and final storage capacities; the cleanup require-

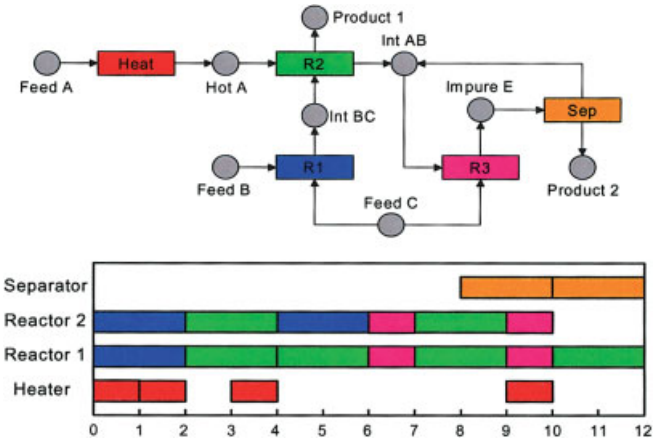

Figure 2. State task network and optimal schedule.

ments for transition between different products; the product demands and their due dates; and the time horizon of interest. The primary objectives are to determine (1) the optimal sequence of tasks that take place in each unit over time; (2) the optimal amount of material produced at each task in each unit and each time; and (3) the processing time of each task in each unit over time. Typical performance criteria introduced for optimality include the maximization of production, the minimization of makespan, the minimization of cost, and the maximization of the contribution margin. Figure 2 depicts an instance of a state task network and the gantt chart of the optimal schedule.

During the last three decades, the research area of process scheduling has received great attention from both academia and industry. A number of reviews in process scheduling were published. Reklaitis[65] provided an overview of scheduling and planning for batch operations. Rippin[66] outlined the batch process systems engineering area. Bassett et al.[67] reviewed the techniques with a focus on model integration. Shah[68] reviewed optimization-based approaches for process scheduling of individual and multiple sites. Pekny and Reklaitis[69] pointed out the implications of the solution methods for scheduling and planning problems. Pinto and Grossmann[70] reviewed the assignment and sequencing models for process scheduling. Floudas and Lin[71] provided an overview and assessment of the continuous-time formulations vs. the discrete-time models for process scheduling problems, and discussed the role of scheduling at the design stage and in the presence of uncertainty. Floudas and Lin[72] reviewed the advances of mixed-integer linear optimization approaches for the scheduling of chemical process systems with a focus on short-term scheduling. In the sequel, a summary of the key advances in process scheduling is discussed, and the research challenges are outlined.

Advances in Process Scheduling. A key advance in the late 1980s was the development of a general discrete-time formulation for scheduling[73,74,75] which introduced a unified framework based on the state-task-network (STN) representation. During the last decade, the most important advances in process scheduling correspond to the transition from discrete-time formulations to continuous-time representations and their formulations (see review by Floudas and Lin[71]). These advances can address (a) sequential processes, and (b) general processes represented as networks. For sequential processes, the continuous-time contributions can be categorized as slot-based approaches,[76-82] and nonslot-based methods.[83-91] For general processes, the continuous time advances can be classified into global event based models,[92-104] and unit-specific event-specific based representations.[105-114] The advances based on the continuous-time representations resulted in significant reduction of the combinatorial complexity, provided improved solutions and reduced integrality gaps, and allowed for the effective treatment of short-term scheduling in large-scale industrial case studies. Advances were also introduced for medium-term scheduling,[110,115] reactive scheduling,[116-121] and scheduling under uncertainty.[122-132] As a result of these advances, the current status of process scheduling reflected through the novel continuous-time formulations can be considered as leading toward bridging the gap between theory and applications, especially since proposed approaches can be applied to large-scale industrial case studies.

Challenges in Process Scheduling. The research opportunities in the general area of scheduling include (a) new modeling and algorithmic approaches for reducing/closing the integrality gap for short-term scheduling problems; (b) improved methods for medium-term scheduling; (c) new approaches for multi-site production scheduling; (d) new reactive scheduling methods; (e) the theoretical and algorithmic studies of methods for scheduling under uncertainty which can address a large number of uncertain parameters; (f) new methods for design, synthesis and scheduling under uncertainty; (g) new approaches for planning under uncertainty, and (h) new unified frameworks for planning and scheduling under uncertainty.

The Genomics Revolution

The genomics revolution has elevated the importance of challenges and opportunities in bioinformatics and computational biology, and opened new avenues for the development of fundamental methods, which share the systems engineering viewpoint, and their applications to important systems biology problems. Figure 3 depicts a number of such challenges. The completion of several genome projects provided a detailed map from the gene sequences to the protein sequences. The gene sequences can be used to assist and/or infer the connectivity within or among the pathways. The large number of protein sequences makes protein structure prediction from the amino acid sequence of paramount importance. The elucidation of the

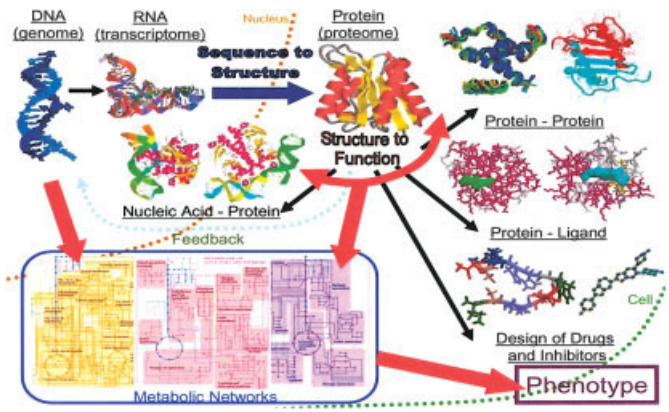

Figure 3. The genomics revolution.

protein structures through novel computational frameworks and established experimental protocols provides key elements for the structure-based prediction of protein function. These include the identification of the type of fold, the type of packing, the residues that are exposed to solvent, the highly conserved residues, the candidate residues for mutations, as well as the shape and electrostatic properties of the fold. Such elements provide the basis for the development of approaches for the location of active sites, the determination of structural and functional motifs, the study of protein-protein, protein-ligand complexes and protein-DNA interactions, the design of new inhibitors, and drug discovery through target selection, lead discovery and optimization. Better understanding of these interactions will assist in addressing key topology related questions in both the cellular metabolic and signal transduction networks. In the sequel, three major components of the genomics revolution roadmap will be addressed: structure prediction in protein folding; de novo protein design; and peptide and protein identification via tandem mass spectroscopy.

Structure Prediction in Protein Folding

Proteins are polymers consisting of 20 amino acids joined by peptide bonds and fold into a unique three-dimensional (3-D) structure which according to the thermodynamic hypothesis,[133] corresponds to the global minimum free energy of the system (i.e., monomeric globular protein in solution at physiological temperatures).

Protein structure prediction is a fundamental scientific problem and it is regarded as a holy grail in computational chemistry, molecular and structural biology. Given an amino acid sequence (i.e., the primary structure) which represents a monomeric globular protein in aqueous solution and at physiological temperatures, the objectives are to determine (1) all helical segments and all beta-strands (i.e., the secondary structure elements), (2) all pairs of beta-strands which form beta-sheets (i.e., the beta-sheet topology), (3) all disulfide bridges if cysteines are present, and (4) the 3-D folded protein structure (i.e., the tertiary structure) which also includes loops that connect secondary structure elements and links that have no secondary structure.

During the last six decades, the protein structure prediction problem and the question of how proteins fold have attracted the interest of and tantalized many researchers across several disciplines. Two viewpoints provide competing explanations to the protein folding question. The classical opinion regards folding as a hierarchical process, implying that the process is initiated by rapid formation of secondary structural elements, followed by the slower arrangement of the actual three-dimensional structure of the tertiary fold. The opposing perspective is based on the idea of a hydrophobic collapse, and suggests that the tertiary and secondary features form concurrently. Contributions for protein structure prediction are classified into four major categories: (a) homology/comparative modeling, (b) fold recognition/threading, (c) first principles methods which use database information, and (d) first principles methods without database information. In a recent review article, Floudas et al.[134] provide a detailed description of the research progress in protein structure prediction and de novo protein design. In the remainder of this section, an outline of the key advances is presented along with the current status and the research challenges.

Advances in Protein Structure Prediction The important advances in protein structure prediction will be discussed based on the aforementioned four classes. In (a), the probe and template sequences are evolutionarily related and the main hypothesis is that sequence similarity implies structural similarity.[135-140] The methods in (b) rely on the better evolutionary conservation of structure than sequence, and the query sequence is matched to a structure from a library of known folds using a variety of scoring functions.[141-151] In (c), information from databases and/or statistical methods is used for secondary structure elements, certain tertiary contacts, and fragments (i.e., short amino acid sequences) which are assembled using scoring functions.[149,150,152-159] The methods in (d) are first principles approaches that do not make use of the database information, but instead seek the minimum of the free energy of the protein in an aqueous solution using physics based atomistic potential force fields.[160-179] An example of the methods of (4) is the Astro-Fold framework[176] which combines the aforementioned classical and new views of protein folding. Astro-Fold identifies first the helical segments through detailed free energy calculations of an overlapping set of oligopeptides and the introduction of deterministic global optimization. In the second stage, Astro-Fold predicts the beta strands and beta sheet topologies via a mixed-integer linear optimization model that maximizes the hydrophobic interactions. In the third stage, free energy calculations for the loops provide tighter bounds for the backbone angles of the loop residues. Finally, the fourth stage combines the secondary structure predictions, develops restraints, formulates a constrained global optimization model, and predicts the tertiary structure through a novel class of hybrid global optimization methods. As evidenced from the CASP experiments,[180-183] the current status of the research in protein structure prediction is that significant progress has taken place and low/medium resolution structures can be predicted for proteins of about 150–200 amino acids with different degrees of success.[134]

Challenges in Protein Structure Prediction. The research challenges and opportunities in protein structure prediction include (a) improved methods for the prediction of helices; (b) improved methods for the prediction of beta-sheet topologies; (c) new methods for loop structure prediction with fixed stems, and flexible stems; (d) new methods for the prediction of disulfide bridges; (e) improved force fields for fold recognition; (f) improved force fields for high-resolution structure prediction; (g) new methods for fold recognition; (h) new approaches for treating uncertainty in force fields and their predictions; (i) new methods for the packing of helices in globular proteins; (j) new methods for the packing of helices in membrane proteins; and (k) methods for structure prediction of helical membrane proteins.

De Novo Protein Design

The major aim in the research area of de novo protein design is to determine the amino acid sequences, which are compatible with specific template backbone structures that may be rigid or flexible. In the early 1980s, it was denoted as the "inverse protein folding" problem[184] primarily because of the screening of sequences for a fixed structure. The de novo protein design

problem is of fundamental importance since it aims at improving our understanding on the mapping of the space of amino acid sequences to known protein folds or postulated/putative protein folds. It is also of significant practical importance on the grounds that it can lead to the improved design of inhibitors, design of novel sequences with better stability, design of catalytic sites of enzymes, and drug discovery.

There are three important components in the de novo protein design problem: (a) the definition of the template backbone structure; (b) the sequence selection; and (c) the validation of the fold specificity and fold stability. The template backbone structure can be (1) a single rigid backbone (e.g., the average NMR structure for a protein); (2) a set of rigid backbone structures (e.g., all NMR structures for a protein or a discrete number of randomly selected rigid structures, based on some algorithmic procedure or a discrete set of rigid structures, based on a parameterization of the backbone); or (3) a flexible backbone structure defined by lower and upper bounds on the distances between the alpha carbon atoms and the backbone dihedral angles. It is apparent that true backbone template flexibility is reflected in (3) since it allows for all possible combinations of distances and backbone dihedral angles within their specified ranges, while (2) considers only a small subset of flexible structures, and (1) is restricted to a single structure only. Recent studies discussed the degree and modes of flexibility via principal component analysis applied to a database of protein structures.[185,186] The sequence selection component faces enormous combinatorial complexity since the search space is m n**,** where m are the amino acids, and n is the number of positions (e.g., for all 20 amino acids in a protein with 100 positions, there are 20 100 sequences). The validation of the fold specificity requires structure prediction calculations for the sequences, while the fold stability requires appropriate free energy calculations. Experimental validation may also be needed for the fold specificity and fold stability.

During the last two decades, a lot of academic and industrial attention focused on the de novo protein design problem. This is evidenced by several recent reviews.[124,187-194] Most of the contributions assume a single rigid backbone template,[195-197] or they allowed for slight overlaps of atoms by reducing the atomic radii.[198,199] A number of contributions considered a set of rigid backbone template structures.[200-206] In contrast, the flexibility of the backbone template structure expressed as ranges for the distances, and the backbone dihedral angles is considered in one recent approach.[207,208]

Advances in De Novo Protein Design The important advances in computational methods for de novo protein design can be divided into three groups: (a) deterministic approaches; (b) stochastic approaches; and (c) combinatorial library centered methods. The deterministic approaches in (a) can be classified into those based on the dead end elimination (DEE) criterion and/or its variants,[196,197,209-218] those based on the self-consistent mean field (SCMF) framework,[219,220] and those based on quadratic assignment-like models coupled with distance dependent force fields for the sequence selection and deterministic global optimization with atomistic level force fields for the validation of fold specificity.[207,208,221] The DEE and SCMF approaches assume rigid backbone templates and a discrete set of rotamers. The stochastic approaches in (b) can be classified into those based on genetic algorithms and/or combination with Monte Carlo sampling,[198,201,222,223] those based on iterating between sequence optimization for a fixed backbone conformation and gradient based optimization of the backbone coordinates for a fixed sequence using a Monte Carlo protocol and the Rosetta program,[206,224,225] and those combining Monte Carlo searches for the sequence selection and the self-consistent mean field for structure generation and energy evaluation.[204,205] The combinatorial library methods formulate the de novo protein design problem as the maximization of entropy subject to a set of constraints.[193,226,227] The current status of research in computational methods for de novo protein design has several validated successes and offers an optimistic view for the future. These include the design of the active site on alpha-lytic protease,[228] the zinc finger,[197] the catalytic site of superoxide dismutase,[229] the WW motif,[223] the protein Top7 which exhibits a new fold,[206] novel inhibitors for complement three,[207,208] novel receptors and sensor proteins and a calcium binding protein,[230,231] and a biologically active enzyme.[232]

Challenges in De Novo Protein Design. The research challenges and opportunities in de novo protein design include (a) improved methods for in silico sequence selection with flexible backbone templates; (b) improved force fields for de novo design; (c) simultaneous sequence and structure selection with flexible templates; (d) design of inhibitors of components of the complement three such as C3a; (e) design of novel antibacterial peptides based on human beta-defensins; (f) discovery of novel G protein coupled receptors; (g) de novo design of small and medium size proteins with known and postulated folds; and (h) mapping the sequences to known folds.

Proteomics: Peptide and Protein Identification via Tandem Mass Spectroscopy

Peptide and protein identification are the central and most fundamental problems in proteomics. Tandem mass spectroscopy (MS/MS) coupled with high-performance liquid chromatography (HPLC), has emerged as a powerful experimental technique which can be used to reliably identify and analyze peptides and proteins within a complex mixture of proteins. A mixture of proteins is initially digested into peptides by enzymes, such as trypsin, and the peptides are subsequently separated via HPLC, ionized and measured for mass/charge ratios via a mass spectrometer (e.g., Finnigan LCQ ESI-MS/MS). Peptides with a specific mass/charge ratio are subsequently fragmented via collision-induced dissociation (CID), and the resulting ions mass/charge ratios are measured by the mass spectrometer. Several types of ions are generated with the most typical being b-ions and y-ions. The primary objective is to identify the peptides and proteins that exist in the complex mixture from the ion peaks in the spectra produced using tandem MS/MS, and develop novel in silico methods for high-throughput proteomics (see Figure 4).

The extensive amount of sequence information embedded in spectra from tandem MS/MS has served as an impetus for the recent development of numerous computational approaches to sequence peptides robustly and efficiently, with particular emphasis on the integration of these algorithms into a high-throughput computational framework for proteomics. The two most frequent computational approaches reported in the literature are database search methods and de novo graph theory based methods, both of which can utilize deterministic, prob-

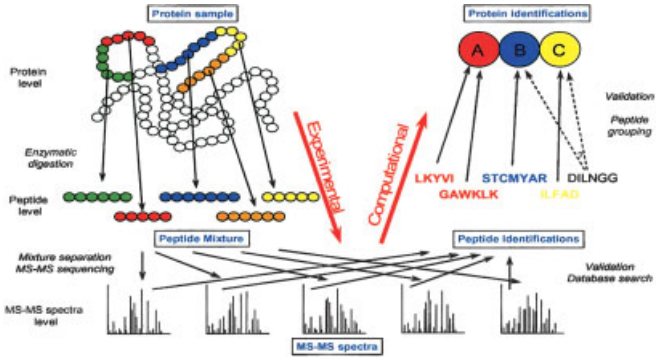

Figure 4. Peptide and protein identification through tandem mass spectroscopy.

abilistic and/or stochastic solution techniques. The database-based approaches have been used more extensively than the de novo approaches especially for high-throughput proteomics. At the same time, however, they exhibit several limitations, which include that (1) they may not include the spectrum for the sought peptide; (2) the protein database may not have the correct protein due to errors in gene-finding programs; (3) sequence databases do not exist for proteomes whose genomes have not been resolved; (4) databases can not capture new protein variants that result from gene splicing; and (5) they cannot address amino acid mutations and protein modifications, such as phosphorylation. Several review articles presented the state of the art in the field from the database and statistical analysis perspective,[233,234] and the de novo viewpoint.[235]

Advances in Peptide and Protein Identification. The key advances in computational methods for peptide and protein identification are discussed, based on the classification into (a) database-based approaches, and (b) de novo approaches. In (a), most of the approaches rely on the use of probabilistic or statistical models for scoring the tandem mass spectra with those available in the databases.[234,236-238] Another class of methods which are stochastic and use genetic algorithms were recently proposed.[239,240] The methods in (b) employ a graph theoretical framework coupled with a penalty/reward function which is correlated by empirical observations and/or heuristic methods.[235,241-245] An alternative technique to the graph-based approaches postulates hypothetical spectra and uses an empirical best-fit objective which tries to match the experimental spectra.[246]

Challenges in Peptide and Protein Identification. The research challenges and opportunities in peptide and protein identification through tandem mass spectroscopy include (a) new first principles methods, based on combinatorial optimization for peptide identification using only information of the ion peaks in the spectrum; (b) new methods for peptide identification under uncertainty in the experimental data; (c) new in silico methods for peptide identification which address the missing peaks from the spectra; (d) efficient hybrid methods which combine first principles methods with database driven approaches for robust peptide identification; and (e) new approaches for protein identification.

Summary

This article has highlighted the advances and challenges in the multidisciplinary research domains of deterministic global optimization, process scheduling, structure prediction in protein folding, de novo protein design, and peptide/protein identification via tandem mass spectroscopy. As evidenced from the contributions, there exists a synergism for systems engineering approaches with formal mathematical analysis frameworks and computational biology. From the challenges perspective, there is a plethora of research opportunities at both the macroscopic and microscopic levels for chemical engineering and computational biology.

Acknowledgments

Support from the National Science Foundation, the National Institutes of Health (R01 GM52032, R24 GM069736), Arkema Corp., AspenTech Corp., BASF Corp., and Siemens Corp., is gratefully acknowledged.

Literature Cited

1. Floudas CA. Deterministic global optimization: Theory, methods and applications. *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers; 2000.
2. Horst R, Pardalos PM, Thoai NV. Introduction to global optimization. *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers; 2000.
3. Tawarmalani M, Sahinidis NV. Convexification and global optimization in continuous and mixed-integer programming. *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers; 2002.
4. Zabinsky ZB. Stochastic adaptive search for global optimization. *Nonconvex Optimization and Its Applications*.Kluwer Academic Publishers; 2003.
5. Neumaier A. Complete search in continuous global optimization and constraint satisfaction. In: Iserles A. *Acta numerica.* Cambridge University Press; 2004:271-369.
6. Floudas CA, Akrotirianakis IG, Caratzoulas S, Meyer CA, Kallrath J. Global optimization in the 21st century: Advances and challenges. *Comp Chem Eng.* 2005;29, 6, 1185-1202.
7. Liberti L, Pantelides CC. Convex envelopes of monomials of odd degree. *J Global Optim*. 2003;25:157-168.
8. Meyer CA, Floudas CA. Trilinear monomials with positive or negative domains: Facets of convex and concave envelopes. In: Floudas CA, Pardalos PM. *Santorini, Greece: Frontiers in Global Optimization.* Kluwer Academic Publishers; 2003:327-352.
9. Meyer CA, Floudas CA. Convex hull of trilinear monomials with mixed-sign domains. *J. Global Optim.* 2004; 29:125-155.
10. Tawarmalani M, Sahinidis NV. Semidefinite relaxations of fractional programs via novel convexification techniques. *J Global Optim.* 2001;20:137-158.
11. Meyer CA, Floudas CA. Convex envelopes for edge-concave functions. *Math. Program.* 2005a (in press).
12. Tawarmalani M, Sahinidis NV. Convex extensions and envelopes of lower semi-continuous functions. *Math Program.* 2002;93:247-263.

13. Adjiman CS, Floudas CA. Rigorous convex underestimators for general twice-differentiable problems. *J Global Optim.* 1996;9:23-40.
14. Adjiman CS, Dallwig S, Floudas CA, Neumaier A. A global optimization method, αBB, for general twice-differentiable NLPs - I. Theoretical advances. *Comp Chem Eng.* 1998a;22:1137-1158.
15. Adjiman CS, Androulakis IP, Floudas CA. A global optimization method, αBB, for general twice-differentiable NLPs - II. Implementation and computational results. *Comp Chem Eng.* 1998b;22:1159-1179.
16. Akrotirianakis IG, Floudas CA. A new class of improved convex underestimators for twice continuously differentiable constrained NLPs. *J Global Optim.* 2004a;30:367-390.
17. Zamora JM, Grossmann IE. A global MINLP optimization algorithm for the synthesis of heat exchanger networks with no stream splits. *Comp Chem Eng.* 1998a;22: 367-384.
18. Zamora JM, Grossmann IE. Continuous global optimization of structured process systems models. *Comp Chem Eng.* 1998b;22:1749-1770.
19. Zamora JM, Grossmann IE. A branch and contract algorithm for problems with concave univariate, bilinear and linear fractional terms. *J Global Optim.* 1999;14:217-219.
20. Caratzoulas S, Floudas CA. A trigonometric convex underestimator for the base functions in Fourier space. *J Optimiz Theory Appl.* 2005;124:339-362.
21. Björk KJ, Lindberg PO, Westerlund T. Some convexifications in global optimization of problems containing signomial terms. *Comp Chem Eng.* 2003;27:669-679.
22. Sherali HD, Adams WP. A reformulation-linearization technique for solving discrete and coninuous nonconvex problems. *Nonconvex Optimization and Its Applications.* Kluwer Academic Publishers; 1999.
23. Meyer CA, Floudas CA. Convex underestimation of twice continuously differentiable functions by piecewise quadratic perturbation: Spline αBB underestimators. *J Global Optim.* 2005b (in press).
24. Akrotirianakis IG, Floudas CA. Computational experience with a new class of convex underestimators: Box-constrained NLP problems. *J Global Optim.* 2004b;29: 249-264.
25. Gau CY, Stadtherr MA. Dynamic load balancing for parallel interval-Newton using message passing. *Comp Chem Eng.* 2002a;26:811-825.
26. Gau CY, Stadtherr MA. New interval methodologies for reliable chemical modeling. *Comp Chem Eng.* 2002b;26: 827-840.
27. Byrne RP, Bogle IDL. Global optimization of constrained nonconvex programs using reformulation and interval analysis. *Comp Chem Eng.* 1999;23:1341-1350.
28. Zilinskas J, Bogle IDL. Evaluation ranges of functions using balanced random interval arithmetic. *Informatica Lithuan.* 2003;14:403-416.
29. Lucia A, Feng Y. Global terrain methods. *Comp Chem Eng.* 2002;26:529-546.
30. Lucia A, Feng Y. Multivariable terrain methods. *AIChE J.* 2003;49:2553-2563.
31. Vecchietti A, Grossmann IE. LOGMIP: A disjunctive 0-1 nonlinear optimizer for process systems models. *Comp Chem Eng.* 1999;23:555-565.
32. Lee A, Grossmann IE. A global optimization algorithm for nonconvex generalized disjunctive programming and applications to process systems. *Comp Chem Engng.* 2001;25:1675-1697.
33. Lee A, Grossmann IE. Global optimization of nonlinear generalized disjunctive programming with bilinear equality constraints: Applications to process networks. *Comp Chem Eng.* 2003;27:1557-1575.
34. Westerlund T, Skrifvars H, Harjunkoski I, Pörn R. An extended cutting plane method for a class of non-convex MINLP problems. *Comp Chem Eng.* 1998;22:357-365.
35. Pörn R, Harjunkoski I, Westerlund T. Convexification of different classes of non-convex MINLP problems. *Comp Chem Eng.* 1999;23:439-448.
36. Pörn R, Westerlund T. A cutting plane method for minimizing pseudo-convex functions in mixed integer case. *Comp. Chem. Eng.* 2000;24:2655-2665.
37. Adjiman CS, Androulakis IP, Floudas CA. Global optimization of mixed-integer nonlinear problems. *AIChE J.* 2000;46:1769-1797.
38. Kesavan P, Barton P. Generalized branch and cut framework for mixed-integer nonlinear optimization problems. *Comp Chem Eng.* 2000;24:1361-1366.
39. Kesavan P, Allgor RL, Gadzke EP, Barton P. Outer approximation algorithms for separable nonconvex mixed-integer nonlinear problems. *Math Program.* 2004; 100:517-535.
40. Tawarmalani M, Sahinidis NV. Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. *Math Program.* 2004;99:563-591.
41. Gümüş ZH, Floudas CA. Global optimization of nonlinear bilevel programming problems. *J Global Optim.* 2001;20:1-31.
42. Floudas CA, Gümüş ZH, Ierapetritou MG. Global optimization in design under uncertainty: feasibility test and flexibility index problems. *Ind Eng Chem Res.* 2001;40: 4267-4282.
43. Pistikopoulos EN, Dua V, Ryu J. Global optimization of bilevel programming problems via parametric programming. In: Floudas CA, Pardalos PM. *Santorini, Greece: Frontiers in Global Optimization*. Kluwer Academic Publishers; 2003:457-476.
44. Gümüş ZH, Floudas CA. Global optimization of mixed-integer bilevel programming problems. *Comput Manage Sci.* 2005 (in press).
45. Esposito WR, Floudas CA. Global optimization for the parameter estimation of differential-algebraic systems. *Ind Eng Chem Res.* 2000a;39:1291-1310.
46. Esposito WR, Floudas CA. Deterministic global optimization in nonlinear optimal control problems. *J Global Optim.* 2000b;17:97-126.
47. Esposito WR, Floudas CA. Deterministic global optimization in isothermal reactor network synthesis. *J Global Optim.* 2002;22:59-95.
48. Papamichail I, Adjiman CS. A rigorous global optimization algorithm for problems with ordinary differential equations. *J Global Optim.* 2002;24:1-33.
49. Adjiman CS, Papamichail I. A deterministic global optimization algorithm for problems with nonlinear dynam-

ics. In: Floudas CA, Pardalos PM. *Santorini, Greece: Frontiers in Global Optimization.* Kluwer Academic Publishers; 2003:1-24.
50. Singer AB, Barton PI. Global solution of optimization problems with dynamic systems embedded. In: Floudas CA, Pardalos PM. *Santorini, Greece: Frontiers in Global Optimization.* Kluwer Academic Publishers; 2003:477-498.
51. Singer AB, Barton PI. Global solution of linear dynamic embedded optimization problems. *J Optimiz Theory Appl.* 2004;121:613-646.
52. Chachuat B, Latifi MA. A new approach in deterministic global optimization of problems with ordinary differential equations. In: Floudas CA, Pardalos PM. *Santorini, Greece: Frontiers in Global Optimization*. Kluwer Academic Publishers; 2003:83-108.
53. Barton PI, Banga JR, Galan S. Optimization of hybrid discrete/continuous dynamic systems. *Comp Chem Eng.* 2000;24:2171-2182.
54. Lee CK, Barton PI. Global dynamic optimization of linear hybrid systems. In: Floudas CA, Pardalos PM. *Santorini, Greece: Frontiers in Global Optimization.* Kluwer Academic Publishers; 2003:289-312.
55. Barton PI, Lee CK. Global dynamic optimization of linear time varying hybrid systems. *Dynamics of continuous discrete and impulsive systems.* 2003:153-158.
56. Lee CK, Singer AB, Barton PI. Global optimization of linear hybrid systems with explicit transitions. *Syst Control Lett.* 2004;51:363-375.
57. Byrne RP, Bogle IDL. Global optimization of modular process flowsheets. *Ind Eng Chem Res.* 2000;39:4296-4301.
58. Jones DR. A taxonomy of global optimization methods based on response surfaces. *J Global Optim.* 2001;21: 345-383.
59. Gutmann HM. A radial basis function method for global optimization. *J Global Optim.* 2001;19:201-227.
60. Meyer CA, Floudas CA, Neumaier A. Global optimization with nonfactorable constraints. *Ind Eng Chem Res.* 2002;41:6413-6424.
61. Hua JZ, Brennecke JF, Stadtherr MA. Reliable computation for phase stability using interval analysis: Cubic equation of state models. *Comp Chem Eng.* 1998a;22: 1207-1214.
62. Hua JZ, Brennecke JF, Stadtherr MA. Enhanced interval analysis for phase stability: Cubic equation of state models. *Ind Eng Chem Res.* 1998b;37:1519-1527.
63. Tessier SR, Brennecke JF, Stadtherr MA. Reliable phase stability analysis for excess Gibbs energy models. *Chem Eng Sci.* 2000;55:1785-1796.
64. Harding ST, Floudas CA. Locating heterogeneous and reactive azeotropes. *Ind Eng. Chem Res.* 2000;39:1576-1595.
65. Reklaitis GV. Overview of scheduling and planning of batch process operations., 1992. NATO Advanced Study Institute - Batch Process Systems Engineering, Antalya, Turkey.
66. Rippin DWT. Batch process systems engineering: A retrospective and prospective review. *Comp Chem Eng.* 1993;17:S1–S13.
67. Bassett MH, Pekny JF, Reklaitis GV. Decomposition techniques for the solution of large-scale scheduling probems. *AIChE J.* 1996;42:3373-3387.
68. Shah N. Single- and multisite planning and scheduling: Current status and future challenges. In: Pekny JF, Blau GE. *Proceedings of the 3rd International Conference on Foundations of Computer-Aided Process Operations.* Snowbird, Utah; July 5-10 1998;75-90.
69. Pekny JF, Reklaitis GV. Towards the convergence of theory and practice: A technology guide for scheduling/planning methodology. In: Pekny JF, Blau GE. *Proceedings of the 3rd International Conference on Foundations of Computer-Aided Process Operations.* Snowbird, Utah; July 5-10 1998;91-111.
70. Pinto JM, Grossmann IE. Assignment and sequencing models for the scheduling of process systems. *Ann Oper Res.* 1998;81:433-466.
71. Floudas CA, Lin X. Continuous-time versus discrete-time approaches for scheduling of chemical processes: A review. *Comp Chem Eng.* 2004; 28:2109-2129.
72. Floudas CA, Lin X. Mixed integer linear programming in process scheduling: Modeling, algorithms, and applications. *Ann Oper Res.* 2005 (in press).
73. Kondili E, Pantelides CC, Sargent RWH. A general algorithm for scheduling batch operations. In *Proceedings of the 3rd International Symposium on Process Systems Engineering.* Sydney, Australia, August 28 - September 2 1988;62-75.
74. Kondili E, Pantelides CC, Sargent RWH. A general algorithm for short-term scheduling of batch operations - I. MILP formulation. *Comp Chem Eng.* 1993;17:211-227.
75. Shah N, Pantelides CC, Sargent RWH. A general algorithm for short-term scheduling of batch operations - II. Computational issues. *Comp Chem Eng.* 1993;17:229-244.
76. Pinto JM, Grossmann IE. A continuous time mixed integer linear programming model for short term scheduling of multistage batch plants. *Ind Eng Chem Res.* 1995;34: 3037-3051.
77. Pinto JM, Grossmann IE. An alternate MILP model for short-term scheduling of batch plants with preordering constraints. *Ind Eng Chem Res.* 1996;35:338-342.
78. Pinto JM, Türkay A, Bolio B, Grossmann IE. STBS: A continuous-time MILP optimization for short-term scheduling of batch plants. *Comp Chem Eng.* 1998;22:1297-1308.
79. Karimi IA, McDonald CM. Planning and scheduling of parallel semicontinuous processes. 2. Short-term scheduling. *Ind Eng Chem Res.* 1997; 36:2701-2714.
80. Bok J, Park S. Continuous-time modeling for short-term scheduling of multipurpose pipeless plants. *Ind Eng Chem Res.* 1998;37:3652-3659.
81. Lamba N, Karimi IA. Scheduling parallel production lines with resource constraints. 1. Model formulation. *Ind Eng Chem Res.* 2002a;41:779-789.
82. Lamba N, Karimi IA. Scheduling parallel production lines with resource constraints. 2. Decomposition algorithm. *Ind Eng Chem Res.* 2002b;41:790-800.
83. Moon S, Park S, LeeWK. New MILP models for scheduling of multiproduct batch plants under zero-wait policy. *Ind Eng Chem Res.* 1996;35:3458-3469.
84. Cerdá J, Henning GP, Grossmann IE. A mixed-integer

linear programming model for short-term scheduling of single-stage multiproduct batch plants with parallel lines. *Ind Eng Chem Res.* 1997;36:1695-1707.
85. Ku H, Karimi IA. Scheduling in serial multiproduct batch processes with finite interstage storage: A Mixed integer linear programming formulation. *Ind Eng Chem Res.* 1988;27:1840-1848.
86. Méndez CA, Henning GP, Cerdá J. Optimal scheduling of batch plants satisfying multiple product orders with different due-dates. *Comp Chem Eng.* 2000b;24:2223-2245.
87. Méndez CA, Henning GP, Cerdá J. An MILP continuous-time approach to short-term scheduling of resource-constrained multistage flowshop batch facilities. *Comp Chem Eng.* 2001;25:701-711.
88. Hui C, Gupta A, van der Meulen HAJ. A novel MILP formulation for short-term scheduling of multi-stage multi-product batch plants with sequence-dependent constraints. *Comp Chem Eng.* 2000;24:2705-2717.
89. Hui C, Gupta A. A bi-index continuous-time mixed-integer linear programming model for single-stage batch scheduling with parallel units. *Ind Eng Chem Res.* 2001; 40:5960-5967.
90. Orçun S, Altinel IK, Hortaçsu Ö. General continuous time models for production planning and scheduling of batch processing plants: Mixed integer linear program formulations and computational issues. *Comp Chem Eng.* 2001;25:371-389.
91. Lee K, Heo S, Lee H, Lee I. Scheduling of single-stage and continuous processes on parallel lines with intermediate due dates. *Ind Eng Chem Res.* 2002;41:58-66.
92. Zhang X, Sargent RWH. The optimal operation of mixed production facilities - A general formulation and some solution approaches for the solution. *Comp Chem Eng.* 1996;20:897-904.
93. Zhang X, Sargent RWH. The optimal operation of mixed production facilities - Extensions and improvements. *Comp Chem Eng.* 1998;22:1287-1295.
94. Schilling G, Pantelides CC. A simple continuous-time process scheduling formulation and a novel solution algorithm. *Comp Chem Eng.* 1996; 20:S1221–S1226.
95. Schilling G, Pantelides CC. Optimal periodic scheduling of multipurpose plants. *Comp Che Engng.* 1999;23:635-655.
96. Mockus L, Reklaitis GV. Mathematical programming formulation for scheduling of batch operations based on nonuniform time discretization. *Comp Chem Eng.* 1997; 21:1147-1156.
97. Mockus L, Reklaitis GV. Continuous time representation approach to batch and continuous process scheduling. 1. MINLP formulation. *Ind Eng Chem Res.* 1999a;38:197-203.
98. Mockus L, Reklaitis GV. Continuous time representation approach to batch and continuous process scheduling. 2. Computational issues. *Ind Eng Chem Res.* 1999b;38:204-210.
99. Castro P, Barbosa-Póvoa APFD, Matos H. An improved RTN continuous-time formulation for the short-term scheduling of multipurpose batch plants. *Ind Eng Chem Res.* 2001;40:2059-2068.
100. Majozi T, Zhu XX. A novel continuous-time MILP formulation for multipurpose batch plants. 1. Short-term scheduling. *Ind Eng Chem Res.* 2001;40:5935-5949.
101. Burkard RE, Fortuna T, Hurkens CAJ. Makespan minimization for chemical batch processes using non-uniform time grids. *Comp Chem Eng.* 2002;26:1321-1332.
102. Wang S, Guignard M. Redefining event variables for efficient modeling of continuous-time batch processing. *Ann Oper Res.* 2002;116:113-126.
103. Maravelias CT, Grossmann IE. New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants. *Ind Eng Chem Res.* 2003;42:3056-3074.
104. Maravelias CT, Grossmann IE. A hybrid MILP/CP decomposition approach for the continuous time scheduling of multipurpose batch plants. *Comp Chem Eng.* 2004;28: 1921-1949.
105. Ierapetritou MG, Floudas CA. Effective continuous-time formulation for short-term scheduling: 1. Multipurpose batch processes. *Ind Eng Chem Res.* 1998a;37:4341-4359.
106. Ierapetritou MG, Floudas CA. Effective continuous-time formulation for short-term scheduling: 2. Continuous and semi-continuous processes. *Ind Eng Chem Res.* 1998b; 37:4360-4374.
107. Ierapetritou MG, Floudas CA. Comments on "An improved RTN continuous-time formulation for the short-term scheduling of multipurpose batch plants". *Ind Eng Chem Res.* 2001;40:5040-5041.
108. Ierapetritou MG, Hené TS, Floudas CA. Effective continuous-time formulation for short-term scheduling: 3. Multiple intermediate due dates. *Ind Eng Chem Res.* 1999;38:3446-3461.
109. Lin X, Floudas CA. Design, synthesis and scheduling of multipurpose batch plants via an effective continuous-time formulation. *Comp Chem Eng.* 2001;25:665-674.
110. Lin X, Floudas CA, Modi S, Juhasz NM. Continuous-time optimization approach for medium-range production scheduling of a multiproduct batch plant. *Ind Eng Chem Res.* 2002;41:3884-3906.
111. Lin X, Chajakis ED, Floudas CA. Scheduling of tanker lightering via a novel continuous-time optimization framework. *Ind Eng Chem Res.* 2003;42:4441-4451.
112. Janak SL, Lin X, Floudas CA. Enhanced continuous-time unit-specific event-based formulation for short-term scheduling of multipurpose batch processes: Resource constraints and mixed storage policies. *Ind Eng Chem Res.* 2004;43:2516-2533.
113. Jia X, Ierapetritou MG. Efficient short-term scheduling of refinery operations based on a continuous time formulation. *Comp Chem Eng.* 2004a;28:1001-1019.
114. Floudas CA, Janak SL. Comments on "New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants" by Christos T. Maravelias and Ignacio E. Grossmann and on "Enhanced continuous-time unit-specific event-based formulation for short-term scheduling of multipurpose batch processes: Resource constraints and mixed storage policies" by Stacy L. Janak, Xiaoxia Lin, and Christodoulos A. Floudas. *Ind Eng Chem Res.* 2005;44:1985-1986.
115. Dimitriadis AD, Shah N, Pantelides CC. RTN-based rolling horizon algorithms for medium term scheduling of

multipurpose plants. *Comp Chem Eng.* 1997;21:S1061–S1066.
116. Sanmartí E, Huercio A, Espuña A, Puigjaner L. A combined scheduling/reactive scheduling strategy to minimize the effect of process operations uncertainty in batch plants. *Comp Chem Eng.* 1996;20:S1263–S1268.
117. Rodrigues MTM, Gimeno L, Passos CAS, Campos MD. Reactive scheduling approach for multipurpose chemical batch plants. *Comp Chem Eng.* 1996;20:S1215–S1220.
118. Honkomp SJ, Mockus L, Reklaitis GV. A framework for schedule evaluation with processing uncertainty. *Comp Chem Eng.* 1999;23:595-609.
119. Vin JP, Ierapetritou MG. A new approach for efficient rescheduling of multiproduct batch plants. *Ind Eng Chem Res.* 2000;39:4228-4238.
120. Roslöf J, Harjunkoski I, Björkqvist J, Karlsson S, Westerlund T. An MILP-based reordering algorithm for complex industrial scheduling and rescheduling. *Comp Chem Eng.* 2001;25:821-828.
121. Méndez CA, Cerdá J. An MILP framework for batch reactive scheduling with limited discrete resources. *Comp Chem Eng.* 2004;28:1059-1068.
122. Bassett MH, Pekny JF, Reklaitis GV. Using detailed scheduling to obtain realistic operating policies for a batch processing facility. *Ind Eng Chem Res.* 1997;36: 1717-1726.
123. Sanmartí E, Espuña A, Puigjaner L. Batch production and preventive maintenance scheduling under equipment failure uncertainty. *Comp Chem Eng.* 1997;21:1157-1168.
124. Vin JP, Ierapetritou MG. Robust short-term scheduling of multiproduct batch plants under demand uncertainty. *Ind Eng Chem Res.* 2001; 40:4543-4554.
125. Balasubramanian J, Grossmann IE. A novel branch and bound algorithm for scheduling flowshop plants with uncertain processing times. *Comp Chem Eng.* 2002;26: 41-57.
126. Balasubramanian J, Grossmann IE. Scheduling optimization under uncertainty - An alternative approach. *Comp Chem Eng.* 2003;27:469-490.
127. Balasubramanian J, Grossmann IE. Approximation to multistage stochastic optimization in multiperiod batch plant scheduling under demand uncertainty. *Ind Eng Chem Res.* 2004;43:3695-3713.
128. Jia Z, Ierapetritou MG. Short-term scheduling under uncertainty using MILP sensitivity analysis. *Ind Eng Chem Res.* 2004b;43:3782-3791.
129. Lin X, Janak SL, Floudas CA. A new robust optimization approach for scheduling under uncertainty: I. Bounded uncertainty. *Comp Chem Eng.* 2004;28:1069-1085.
130. Bonfill A, Bagajewicz M, Espuña A, Puigjaner L. Risk management in the scheduling of batch plants under uncertain market demand. *Ind Eng Chem Res.* 2004;43: 741-750.
131. Bonfill A, Espuña A, Puigjaner L. Addressing robustness in scheduling batch processes with uncertain operation times. *Ind. Eng. Chem. Res.* 2005;44:1524-1534.
132. Janak SL, Lin X, Floudas CA. A new robust optimization approach for scheduling under uncertainty: II. Uncertainty with known probability distribution. 2005 (submitted for publication).
133. Anfinsen CB. Principles that govern the folding of protein chains. *Science.* 973;181:223-230.
134. Floudas CA, Fung HK, McAllister SR, Mönningmann M, Rajgaria R. Advances in protein structure prediction and de novo protein design: A review. *Chem Eng Sci* 2005b (accepted for publication).
135. Karplus K, Barret C, Hughey R. Hidden Markov models for detecting remote protein homologies. *Bioinformatics.* 1998;14:846-856.
136. Vitkup D, Melomud E, Moult J, Sander C. Completeness in structural genomics. *Nat Struct Biol.* 2001;8:559-566.
137. Al-Lazikani B, Sheinerman FB, Honig B. Combining multiple structure and sequence alignments to improve sequence detection and alignment: Application to the SH2 domains of Janus kinases. *P Natl Acad Sci.USA.* 2001;98:14796-14801.
138. Notredame C. Recent progress in multiple sequence alignment: A survey. *Pharmacogenomics J.* 2002;3:131-144.
139. Fiser A, Feig M, Brooks-III CL, Sali A. Evolution and physics in comparative protein structure modeling. *Accounts Chem Res.* 2002;35:413-421.
140. Tramontano A, Morea V. Assessment of homology-based predictions in CASP5. *Prot Struct Funct Bioinf.* 2003;53: 352-368.
141. Jones DT. GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol.* 1999a;287:797-815.
142. Jones DT. Protein secondary structure prediction based on position specific scoring matrices. *J Mol Biol.* 1999b; 292:195-202.
143. Xu Y, Xu D. Protein threading using PROSPECT: Design and evolution. *Prot Struct Funct Bioinf.* 2000;40: 343-354.
144. An Y, Friesner RA. A novel fold recognition method using composite predicted secondary structures. *Prot Struct Funct Bioinf.* 2002;48:352-366.
145. Kihara D, Skolnick J. The PDB is a covering set of small protein structures. *J Mol Biol.* 2003;334:793-802.
146. Kim D, Xu D, Guo J, Ellrott K, Xu Y. PROSPECT II: Protein structure prediction program for genome-scale applications. *Protein Eng.* 2003;16:641-650.
147. Xu J, Li M, Kim D, Xu Y. RAPTOR: Optimial protein threading by linear programming. *J Bioinf Comput Biol.* 2003;1:95-117.
148. Przybylski D, Rost B. Improving fold recognition without folds. *J Mol Biol.* 2004;341:255-269.
149. Zhang Y, Skolnick J. Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. *Biophys J.* 2004a;87:2647-2655.
150. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. *P Natl Acad Sci USA.* 2004b;101:7594-7599.
151. Skolnick J, Kihara D, Zhang Y. Development and large scale benchmark testing of the PROSPECTOR 3 threading algorithm. *Prot Struct Funct Bioinf.* 2004;56:502-518.
152. Simons KT, Kooperberg C, Huang C, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol.* 1997;268:209-225.

153. Simons KT, Ruczinki I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like structures using a combination of sequence-dependent and sequence-independent features of proteins. *Prot Struct Funct Bioinf.* 1999;34:82-95.
154. Eyrich VA, Standley DM, Felts AK, Friesner RA. Protein tertiary structure prediction using a branch and bound algorithm. *Prot Struct Funct Bioinf.* 1999a;35:41-57.
155. Eyrich VA, Standley DM, Friesner RA. Prediction of protein tertiary structure to low resolution: Performace for a large and structurally diverse test set. *J Mol Biol.* 1999b;288:725-742.
156. Skolnick J, Kolinski A, Kihara D, Betancourt M, Rotkiewicz P, Boniecki M. *Ab initio* protein structure prediction via a combination of threading, lattice folding, clustering and structure refinement. *Prot Struct Funct Bioinf.* 2001;5:S149–S156.
157. Skolnick J, Zhang Y, Arakaki AK, Kolinski A, Boniecki M, Szilágyi A, Kihara D. TOUCHSTONE: A unified approach to protein structure prediction. *Prot Struct Funct Bioinf.* 2003;53:469-479.
158. Rohl CA, Strauss CEM, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with Rosetta. *Prot Struct Funct Bioinf.* 2004;55:656-677.
159. Lee J, Kim SY, Joo K, Kim I, Lee J. Prediction of protein tertiary strucure using PROFESY, a novel method based on fragment assembly and conformational space annealing. *Prot Struct Funct Bioinf.* 2004;56:704-714.
160. Srinivasan R, Rose GD. LINUS: A hierarchic procedure to predict the fold of a protein. *Prot Struct Funct Bioinf.* 1995;22:81-89.
161. Lee J, Scheraga HA, Rackovsky S. New optimization method for conformational energy calculations on polypeptides : Conformational space annealing. *J Comput Chem.* 1997;18:1222-1232.
162. Lee J, Scheraga HA, Rackovsky S. Conformational analysis of the 20-residue membrane-bound portion of melittin by conformational space annealing. *Biopolymers.* 1998;46:103-115.
163. Lee J, Pillardy J, Czaplewski C, Arnautova Y, Ripoll DR, Liwo A, Gibson KD, Wawak RJ, Scheraga HA. Efficient parallel algorithms in global optimization of potential energy functions for peptides, proteins and crystals. *Comput Phys Commun.* 2000;128:399-411.
164. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. A united-residue force field for off-lattice protein structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem.* 1997a;18:849-873.
165. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA. A united-residue force field for off-lattice protein structure simulations. II. Parameterization of short-range interactions and determination of weights of energy terms by z-score optimization. *J Comput Chem.* 1997b;18:874-887.
166. Liwo A, Czaplewski C, Pillardy J, Scheraga HA. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J Chem. Phys.* 2001;115:2323-2347.
167. Liwo A, Arlukowicz P, Czaplewski C, Oldziej S, Pillardy J, Scheraga HA. A method for optimizing potential-energy functions by hierarchical design of the potential-energy landscape: Application to the UNRES force field. *P Natl Acad Sci USA.* 2002;99:1937-1942.
168. Ripoll D, Liwo A, Scheraga HA. New developments of the electrostatically driven Monte Carlo method: Tests on the membrane-bound portion of melittin. *Biopolymers.* 1998;46:117-126.
169. Lee J, Scheraga HA. Conformational space annealing by parallel computations: Extensive conformational search of met-enkephalin and the 20-residue membrane-bound portion of melittin. *Int J Quantum Chem.* 1999;75:255-265.
170. Pillardy J, Czaplewski C, Liwo A, Wedemeyer WJ, Lee J, Ripoll DR, Arlukowicz P, Oldziej S, Arnautova EA, Scheraga HA. Development of physics-based energy functions that predict medium resolution structure for proteins of α, β and α/β structural classes. *J Phys Chem B.* 2001;105:7299-7311.
171. Czaplewski C, Liwo A, Pillardy J, Oldziej S, Scheraga HA. Improved conformational space annealing method to treat beta-structure with the UNRES force-field and to enhance scalability of parallel implementation. *Polymer.* 2004a;45:677-686.
172. Czaplewski C, Oldziej S, Liwo A, Scheraga HA. Prediction of the structures of proteins with the UNRES force field, including dynamic formation and breaking of disulfide bonds. *Protein Eng Des Sel.* 2004b;17:29-36.
173. Klepeis JL, Floudas CA. Ab initio prediction of helical segments in polypeptides. *J Comput Chem.* 2002;23:245-266.
174. Klepeis JL, Floudas CA. Prediction of beta-sheet topology and disulfide bridges in polypeptides. *J Comput Chem.* 2003a;24:191-208.
175. Klepeis JL, Floudas CA. Ab initio tertiary structure prediction of proteins. *J Global Optim.* 2003b;25:113-140.
176. Klepeis JL, Floudas CA. ASTRO-FOLD: A combinatorial and global optimization framework for *ab initio* prediction of three-dimensional structures of proteins from the amino acid sequence. *Biophys. J* 2003c;85:2119-2146.
177. Klepeis JL, Pieja MT, Floudas CA. A new class of hybrid global optimization algorithms for peptide structure prediction: Integrated hybrids. *Comput Phys Commun.* 2003a;151:121-140.
178. Klepeis JL, Pieja MT, Floudas CA. Hybrid global optimization algorithms for protein structure prediction : Alternating hybrids. *Biophys J.* 2003b;84:869-882.
179. Klepeis JL, Wei Y, Hecht MH, Floudas CA. Ab initio prediction of the three-dimensional structure of a de novo designed protein: A double-blind case study. *Pro .Struct Funct Bioinf.* 2005;58:560-570.
180. Moult J, Hubbard T, Bryant SH, Fidelis K, Pedersen JT. Critical assessment of methods of protein structure prediction (CASP): Round II. *Prot Struct Funct Bioinf.* 1997;1:S2–S6.
181. Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction CASP - Round 4. *Prot Struct Funct Bioinf.* 2001;5:S2–S7.
182. Moult J, Fidelis K, Zemla A, Hubbard T. Critical assess-

ment of methods of protein structure prediction (CASP)-Round V. *Prot Struct Funct Bioinf.* 2003;53:334-339.
183. Moult J. Predicting protein three-dimensional structure. *Curr Opin Biotech.* 1999;10:583-588.
184. Pabo C. Molecular technology - Designing proteins and peptides. *Nature*. 1983;301:200-200.
185. Emberly EG, Mukhopadhyay R, Tang C, Wingreen NS. Flexibility of α-helices: results of a statistical analysis of database protein structures. *J Mol Biol.* 2003;327:229-237.
186. Emberly EG, Mukhopadhyay R, Tang C, Wingreen NS. Flexibility of β-sheets: principal component analysis of database protein structures. *Proteins.* 2004;55:91-98.
187. Street AG, Mayo SL. Computational protein design. *Struct Fold Des.* 1999;7:R105–R109.
188. Saven JG. Designing protein energy landscapes. *Chem Rev* 2001;101:3113-3130.
189. Saven JG. Combinatorial protein design. *Curr Opin Struc Biol.* 2002;12:453-458.
190. Pokala N, Handel TM. Review: Protein design - where we were, where we are, where we are going. *J Struct Biol.* 2001;134:269-281.
191. Kraemer-Pecore CM, Wollacott AM, Desjarlais JR. Computational protein design. *Curr Opin Chem Biol.* 2001;5:690-695.
192. Kuhlman B, Baker D. Exploring folding free energy landscapes using computational protein design. *Curr Opin Struc Biol.* 2004;14:89-95.
193. Park S, Yang X, Saven JG. Advances in computational protein design. *Curr Opin Struc Biol.* 2004;14:487-494.
194. Hecht MH, Das A, Go A, Bradley LH, Wei Y. De novo proteins from designed combinatorial libraries. *Protein Sci.* 2004;13:1711-1723.
195. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol.* 1987;193:775-791.
196. Dahiyat BI, Mayo SL. Protein design automation. *Protein Sci.* 1996;5:895-903.
197. Dahiyat BI, Mayo SL. De novo protein design: Fully automated sequence selection. *Science.* 1997;278:82-87.
198. Desjarlais JR, Handel TM. De novo design of the hydrophobic cores of proteins. *Protein Sci.* 1995;4:2006-2018.
199. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *P Natl Acad Sci USA.* 2000;97:10383-10388.
200. Su A, Mayo SL. Coupling backbone flexibility and amino acid sequence selection in protein design. *Protein Sci.* 1997;6:1701-1707.
201. Desjarlais JR, Handel TM. Side-chain and backbone flexibility in protein core design. *J Mol Biol.* 1999;290:305-318.
202. Farinas E, Regan L. The de novo design of a rubredoxin-like Fe site. *Protein Sci.* 1998;7:1939-1946.
203. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. *Science.* 1998;282:1462-1467.
204. Koehl P, Levitt M. De novo protein design. I. In search of stability and specificity. *J Mol Biol* 1999a;293:1161-1181.
205. Koehl P, Levitt M. De novo protein design. II. Plasticity in sequence space. *J Mol Biol.* 1999b;293:1183-1193.
206. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science.* 2003;302:1364-1368.
207. Klepeis JL, Floudas CA, Morikis D, Tsokos CG, Argyropoulos E, Spruce L, Lambris JD. Integrated computational and experimental approach for lead optimization and design of compstatin variants with improved activity. *J Am Chem Soc.* 2003c;125:8422-8423.
208. Klepeis JL, Floudas CA, Morikis D, Tsokos CG, Lambris JD. Design of peptide analogs with improved activity using a novel de novo protein design approach. *Ind Eng Chem Res.* 2004;43:3817-3826.
209. Desmet J, Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature.* 1992;356:539-542.
210. Goldstein RF. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys J.* 1994;66:1335-1340.
211. Malakauskas SM, Mayo SL. Design, structure, and stability of a hyperthermophilic protein variant. *Nat Struct Biol.* 1998;5:470-475.
212. Strop P, Mayo SL. Rubredoxin variant folds without irons. *J Am Chem Soc.* 1999;121:2341-2345.
213. Pierce NA, Spriet JA, Desmet J, Mayo SL. Conformational splitting: A more powerful criterion for dead-end elimination. *J Comput Chem.* 2000;21:999-1009.
214. Wernisch L, Hery S, Wodak SJ. Automatic protein design with all atom force-fields by exact and heuristic optimization. *J Mol Biol.* 2000;301:713-736.
215. Voigt CA, Mayo SL, Arnold FH, Wang ZG. Computational method to reduce the search space for directed protein evolution. *P Natl Acad Sci USA.* 2001;98:3778-3783.
216. Bolon DN, Mayo SL. Enzyme-like proteins by computational design. *P Natl Acad Sci USA* 2001;98:14274-14279.
217. Looger LL, Hellinga HW. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: Implications for protein design and structural genomics. *J Mol Biol.* 2001;307:429-445.
218. Gordon DB, Hom GK, Mayo SL, Pierce NA. Exact rotamer optimization for protein design. *J Comput Chem.* 2003;24:232-243.
219. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J Mol Biol* 1994;239:249-275.
220. Lee C. Predicting protein mutant energetics by self-consistent ensemble optimization. *J Mol Biol.* 1994;236:918-939.
221. Fung HK, Floudas CA. Computational comparison studies of quadratic assignment like formulations for the in silico sequence selection problem in de novo protein design. *J Global Optim.* 2005 (in press).
222. Tuffery P, Etchebest C, Hazout S, Lavery R. A new approach to the rapid determination of protein side chain conformations. *J Biomol Struct Dyn.* 1991;8:1267-1289.
223. Kraemer-Pecore CM, Lecomte JT, Desjarlais JR. A de

novo redesign of the WW Domain. *Protein Sci.* 2003;12: 2194-2205.
224. Kuhlman B, O'Neill JW, Kim DE, Zhang KYJ, Baker D. Accurate computer-based design of a new backbone conformation in the second turn of protein L. *J Mol Biol.* 2002;315:471-477.
225. Dantas G, Kuhlman B, Callender D, Wong M, Baker D. A large scale test of computational protein design: Folding and stability of nine completely redesigned globular proteins. *J Mol Biol.* 2003;332:449-460.
226. Zou J, Saven JG. Statistical theory of combinatorial libraries of folding proteins: Energetic discrimination of a target structure. *J Mol Biol.* 2000;296:281-294.
227. Kono H, Saven JG. Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *J Mol Biol.* 2001;306:607-628.
228. Wilson C, Mace JE, Agard DA. Computational method for the design of enzymes with altered substrate specificity. *J Mol Biol.* 1991;220:495-506.
229. Benson DE, Wisz MS, Hellinga HW. Rational design of nascent metalloenzymes. *P Natl Acad Sci USA.* 2000;97: 6292-6297.
230. Looger LL, Dwyer MW, Smith JJ, Hellinga HW. Computational design of receptor and sensor proteins with novel functions. *Nature.* 2003;423:185-190.
231. Yang W, Jones LM, Isley L, Ye Y, Lee H-W, Wilkins A, Liu ZR, Hellinga HW, Malchow R, Ghazi M, Yang JJ. Rational design of a calcium-binding protein. *J Am Chem Soc.* 2003;125:6165-6171.
232. Dwyer MA, Looger LL, Hellinga HW. Computational design of a biologically active enzyme. *Science.* 2004; 304:1967-1971.
233. Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem Rev.* 2001;101:269-295.
234. Nesvizhskii AI, Aebersold R. Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *DDT.* 2004;9:173-181.
235. Lu B, Chen C. Algorithms for de novo peptide sequencing using tandem mass spectrometry. *DDT: BIOSILICO.* 2004;2:85-90.
236. Bafna V, Edwards N. SCOPE: A probabilistic model for scoring tandem mass spectra against a peptide database. *Bioinformatics.* 2001;17:S13–S21.
237. Pevzner PA, Mulyukov Z, Dancik V, Tang CL. Efficiency of database search for identification of mutated and modified proteins via mass spectrometry. *Genome Res.* 2001;11:290-299.
238. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem.* 2002;74:5383-5392.
239. Heredia-Langner A, Cannon W.R., Jarman KD, Jarman KH. Sequence optimization as an alternative to de novo analysis of tandem mass spectrometry data. *Bioinformatics.* 2004;20:2296-2304.
240. Malard JM, Heredia-Langner A, Baxter DJ, Jarman KH, Cannon WR. Constrained de novo peptide sequencing via multi-objective optimization. In: *Proceedings of the Eighteenth International Parallel and Distributed Processing Symposium.* Santa Fe, New Mexico; April 26-30 2004.
241. Dancik V, Addona TA, Clauser KR, Vath JE, Pevzner PA. De novo peptide sequencing via tandem mass spectrometry. *J Comp Biol.* 1999;6:327-342.
242. Chen T, Kao M, Tepel M, Rush J, Church GM. A dynamic programming approach to de novo peptide sequencing via tandem mass spectrometry. *J Comp Biol.* 2001;10:325-337.
243. Taylor JA, Johnson RS. Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal Chem.* 2001;73:2594-2604.
244. Lubeck O, Sewell C, Gu S, Chen X, Cai DM. New computational approaches for de novo peptide sequencing from MS/MS experiments. *Proc IEEE.* 2002;90: 1868-1874.
245. Lu B, Chen T. A suboptimal algorithm for de novo peptide sequencing via tandem mass spectrometry. *J Comp Biol.* 2003;10:1-12.
246. Bin M, Kaizhong Z, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2003;17:2337-2342.