

Comparison of Multi-Objective Evolutionary Algorithms to Solve the Modular Cell Design Problem for Novel Biocatalysis

Authors:

Sergio Garcia, Cong T. Trinh

Date Submitted: 2019-08-08

Keywords: MOEA, multi-objective evolutionary algorithms, constraint-based modeling, metabolic network modeling, multi-objective optimization, metabolic engineering, modular cell, modular design, modularity

Abstract:

A large space of chemicals with broad industrial and consumer applications could be synthesized by engineered microbial biocatalysts. However, the current strain optimization process is prohibitively laborious and costly to produce one target chemical and often requires new engineering efforts to produce new molecules. To tackle this challenge, modular cell design based on a chassis strain that can be combined with different product synthesis pathway modules has recently been proposed. This approach seeks to minimize unexpected failure and avoid task repetition, leading to a more robust and faster strain engineering process. In our previous study, we mathematically formulated the modular cell design problem based on the multi-objective optimization framework. In this study, we evaluated a library of state-of-the-art multi-objective evolutionary algorithms (MOEAs) to identify the most effective method to solve the modular cell design problem. Using the best MOEA, we found better solutions for modular cells compatible with many product synthesis modules. Furthermore, the best performing algorithm could provide better and more diverse design options that might help increase the likelihood of successful experimental implementation. We identified key parameter configurations to overcome the difficulty associated with multi-objective optimization problems with many competing design objectives. Interestingly, we found that MOEA performance with a real application problem, e.g., the modular strain design problem, does not always correlate with artificial benchmarks. Overall, MOEAs provide powerful tools to solve the modular cell design problem for novel biocatalysis.

Record Type: Published Article

Submitted To: LAPSE (Living Archive for Process Systems Engineering)

Citation (overall record, always the latest version):

LAPSE:2019.0939

Citation (this specific file, latest version):

LAPSE:2019.0939-1

Citation (this specific file, this version):

LAPSE:2019.0939-1v1

DOI of Published Version: <https://doi.org/10.3390/pr7060361>

License: Creative Commons Attribution 4.0 International (CC BY 4.0)

Article

Comparison of Multi-Objective Evolutionary Algorithms to Solve the Modular Cell Design Problem for Novel Biocatalysis

Sergio Garcia ^{1,2}  and Cong T. Trinh ^{1,2,*} 

¹ Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN 37996, USA; sergiogarciamf@gmail.com

² Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

* Correspondence: ctrinh@utk.edu; Tel.: +1-865-974-2181

Received: 22 April 2019; Accepted: 30 May 2019; Published: 11 June 2019



Abstract: A large space of chemicals with broad industrial and consumer applications could be synthesized by engineered microbial biocatalysts. However, the current strain optimization process is prohibitively laborious and costly to produce one target chemical and often requires new engineering efforts to produce new molecules. To tackle this challenge, modular cell design based on a chassis strain that can be combined with different product synthesis pathway modules has recently been proposed. This approach seeks to minimize unexpected failure and avoid task repetition, leading to a more robust and faster strain engineering process. In our previous study, we mathematically formulated the modular cell design problem based on the multi-objective optimization framework. In this study, we evaluated a library of state-of-the-art multi-objective evolutionary algorithms (MOEAs) to identify the most effective method to solve the modular cell design problem. Using the best MOEA, we found better solutions for modular cells compatible with many product synthesis modules. Furthermore, the best performing algorithm could provide better and more diverse design options that might help increase the likelihood of successful experimental implementation. We identified key parameter configurations to overcome the difficulty associated with multi-objective optimization problems with many competing design objectives. Interestingly, we found that MOEA performance with a real application problem, e.g., the modular strain design problem, does not always correlate with artificial benchmarks. Overall, MOEAs provide powerful tools to solve the modular cell design problem for novel biocatalysis.

Keywords: modularity; modular design; modular cell; metabolic engineering; metabolic network modeling; constraint-based modeling; multi-objective optimization; multi-objective evolutionary algorithms; MOEA

1. Introduction

Multi-objective optimization is a powerful mathematical toolbox widely used in engineering disciplines to solve problems with multiple conflicting design objectives [1]. For example, in the field of chemical engineering, multi-objective optimization has been applied to balance design conflicts in the performance, material and energy requirements, and environmental sustainability of many different chemical processes [2]. In industrial biotechnology, with recent advancements in synthetic biology and metabolic engineering, microorganisms can be genetically modified to produce a large space of molecules with broad applications using renewable lignocellulosic biomass or waste products as feedstocks [3,4]. However, the current strain design process is prohibitively laborious and expensive for broad industrial application [5]. To overcome this challenge, recent studies have

proposed the application of modular design principles commonly used in engineering [6] to microbial biocatalysis [7–10]. This modular cell design approach, known as ModCell, uses multi-objective optimization to account for the competing cellular objectives when cellular metabolism is (re)designed in a modular fashion to produce a diverse class of target chemicals. ModCell has been experimentally demonstrated for biosynthesis of alcohols [7,11,12] and esters [13–17] in *Escherichia coli*.

Despite the broad applicability of multi-objective optimization problems (MOPs) in engineering design, powerful solution algorithms remain elusive. Two approaches can be used to solve MOPs, including multi-objective evolutionary algorithms (MOEAs) and mixed integer linear programming (MILP) algorithms. Unlike MOEA, MILP can ensure that the identified MOP solutions are optimal. Nonetheless, MOEAs are widely used due to the following advantages over MILP: (i) computational scalability for large-scale networks by implementing efficient parallelization algorithms [18], (ii) compatibility with non-linear objectives and constraints, and (iii) unbiased sampling of Pareto optimal solutions without a need to pre-specify objective preference [19]. MOEAs are based on a more general type of optimization method known as evolutionary algorithms, where candidate solutions, that represent individuals of a population, are iteratively modified using heuristic rules to increase their fitness (i.e., objective function values). Recently, much attention has been placed in the development of MOEAs to solve many-objective problems (e.g., problems with 4 or more objectives) that often correspond to real-world applications, but can be very challenging to solve with conventional MOEAs [20]. For the case of ModCell problem, the popular MOEA NSGAI2 [21,22] was used to design a modular cell under 20 different production modules [8]. Due to a large space of molecules that can potentially be synthesized by modular cells, scalability issues are expected to occur when constructing modular cells that are designed to be compatible with tenths or hundreds of products. Furthermore, using the best solver algorithm(s) allows to explore a more diverse design space, resulting in better choices for experimental implementation.

Many MOEAs have been proposed over the past two decades since the inception of landmark algorithms such as NSGAI2 [23] and SPEA2 [24]. New MOEAs are benchmarked against libraries of artificial problems with known solutions [25,26], and are expected to show enhanced performance for a subset of these problems in terms of scalability, identification of Pareto optimal solutions, and number of simulation generations needed to converge. This benchmarking methodology does not always reflect MOEA performance for general problems, since specialized parameter configurations or heuristics are often used and can lead to drastically different performance towards a specific problem of interest. Thus, the best MOEA for a certain application problem needs to be determined empirically. In this study, we evaluated a library of state-of-the-art MOEAs to solve the multi-objective ModCell problem, with the focus on many-objectives methods. Several cases study of increasing difficulty were examined using common performance indicators of solution optimality and diversity, and critical algorithm parameters that determine solution quality were also investigated.

2. Methods

2.1. Multi-Objective Modular Cell Design

Modular cell design enables rapid generation of optimal production strains with desirable phenotypes from a modular (chassis) cell [8], requiring minimal strain optimization cycles. These production strains are assembled from a modular cell and various compatible pathway modules. A modular cell is constructed by eliminating genes from a parent strain to maintain only core metabolic pathways shared across all pathway modules. Each module enables an optimized target product synthesis phenotype that leads to high yields, titers, and production rates. The different biochemical nature of each target metabolite can make the objectives compete with each other, turning the modular cell design problem into a multi-objective optimization problem known as ModCell2 [8]:

$$\max_{y_j, z_{jk}} (f_1, f_2, \dots, f_{|\mathcal{K}|})^T \quad \text{s.t.} \quad (1)$$

$$f_k \in \arg \max \left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right. \quad (2)$$

$$\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0 \quad \text{for all } i \in \mathcal{I}_k \quad (3)$$

$$l_{jk} \leq v_{jk} \leq u_{jk} \quad \text{for all } j \in \mathcal{J}_k \quad (4)$$

$$l_{jk} d_{jk} \leq v_{jk} \leq u_{jk} d_{jk} \quad \text{for all } j \in \mathcal{C} \quad (5)$$

$$\left. \text{where } d_{jk} = y_j \vee z_{jk} \right\} \quad \text{for all } k \in \mathcal{K}$$

$$z_{jk} \leq (1 - y_j) \quad \text{for all } j \in \mathcal{C}, k \in \mathcal{K} \quad (6)$$

$$\sum_{j \in \mathcal{C}} z_{jk} \leq \beta_k \quad \text{for all } k \in \mathcal{K} \quad (7)$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \leq \alpha \quad (8)$$

where \mathcal{I}_k , \mathcal{J}_k , and \mathcal{K} are the sets of metabolites, reactions, and associated production metabolic networks (i.e., the combination of the chassis organism with a specific product synthesis pathway), respectively. The optimization problem seeks to simultaneously maximize all objectives f_k (1). The desirable phenotype f_k for production module k is determined based on key metabolic fluxes v_{jk} (mmol/gDCW/h) predicted by the constraint-based metabolic model (2)–(5) [27]. For example, the weak growth coupled to product formation (*wGCP*), a common design objective, requires a high minimum product synthesis rate at the maximum growth-rate, enabling growth selection of optimal production strains. Thus, in *wGCP* design, the inner optimization problem seeks to maximize growth rate while calculating the minimum product synthesis rate through the linear objective function (2) (where c_{jk} is 1 and -0.0001 for j corresponding to the biomass and product reactions across all networks k , respectively, and 0 otherwise) subject to: (i) mass-balance constraints (3), where S_{ijk} represents the stoichiometric coefficient of metabolite i in reaction j of production network k , (ii) flux bound constraints (4) that determine reaction reversibility and available substrates, where l_{jk} and u_{jk} are lower and upper bounds, respectively, and (iii) genetic manipulation constraints (5), i.e., deletion of a reaction j in the chassis through the binary indicator y_j , or insertion of a reaction j in a specific production network k through the binary indicator z_{jk} . The maximum product synthesis rate of each production network k , f_k^{max} , is determined by maximizing the product synthesis reaction subject to (3)–(4), allowing to bound f_k in *wGCP* between 0 and 1. Only a subset of all metabolic reactions, \mathcal{C} , are considered as candidates for deletion, since many of the reactions in the metabolic model cannot be manipulated to enhance the target phenotype. Certain reactions can be deleted in the chassis but inserted back to specific production modules, enabling the chassis to be compatible with a broader number of modules (6). The numbers of module-reaction additions and reaction deletions in the chassis are constrained by the parameters β_k (7) and α (8), respectively, to avoid unnecessary genetic manipulations that are generally time-consuming to implement and can lead to unforeseen phenotypes.

2.2. Optimal Solutions for a Multi-Objective Optimization Problem

Optimal solutions for a MOP (1)–(8) are defined based on the concept of domination: A vector $a = (a_1, \dots, a_K)^T$ dominates another vector $b = (b_1, \dots, b_K)^T$, denoted as $a \prec b$ if and only if $a_i \geq b_i \forall i \in \{1, 2, \dots, K\}$ and $a_i \neq b_i$ for at least one i . Letting x be the design variables (i.e., y_j and z_{jk}) and X be the feasible set determined by the problem constraints (2)–(8), a feasible solution $x^* \in X$ of the

MOP is called a Pareto optimal solution if and only if there does not exist a vector $x' \in X$ such that $F(x') \prec F(x^*)$. The set of all Pareto optimal solutions is called Pareto set:

$$PS := \{x \in X : \nexists x' \in X, F(x') \prec F(x)\} \quad (9)$$

The projection of the Pareto set in the objective space is denoted as Pareto front:

$$PF := \{F(x) : x \in PS\} \quad (10)$$

2.3. MOEA Selection

To find the best MOEAs for ModCell2, we evaluated a recent and comprehensive set of MOEAs implemented in the PlatEMO platform [28]. From over 50 algorithms available in PlatEMO, we selected 2 methods for benchmark study, including NSGAII/gamultiobj and MOEAIGDNS, and 8 methods that have been specifically developed to tackle many-objective problems with discrete variables like ModCell2, including ARMOEA, EFRRR, MaOEADDFC, SPEAR, tDEA, BiGE, NSGAIII, and SPEA2SDE (Table 1). It should be noted that gamultiobj is an alternative implementation of the NSGAII algorithm available in Matlab.

Table 1. Summary of MOEAs used in this study.

Abbreviation	Name	Notes	Reference
NSGAII	Non-dominated sorting genetic algorithm 2	Highly applied MOEA	[23]
gamultiobj	Matlab implementation of NSGAII	Used in the original ModCell2 study [8]	[21]
MOEAIGDNS	Multi-objective evolutionary algorithm based on an enhanced inverted generational distance metric	General MOEA with an implementation that works well with discrete variables	[29]
ARMOEA	Adaptation to reference points multi-objective evolutionary algorithm	Many-objective EA based on MOEAIGDNS	[30]
EFRRR	Ensemble fitness ranking with ranking restriction	Many-objective EA	[31]
MaOEADDFC	Many-objective evolutionary algorithm based on directional diversity and favorable convergence	Many-objective EA	[32]
SPEAR	Strength Pareto evolutionary algorithm based on reference direction	Many-objective EA	[33]
tDEA	θ -dominance evolutionary algorithm	Many-objective EA	[34]
BiGE	Bi-goal evolution	Many-objective EA	[35]
NSGAIII	Non-dominated sorting genetic algorithm 3	Many-objective EA	[36]
SPEA2SDE	Strength Pareto evolutionary algorithm 2 with shift-based density estimation	Many-objective EA	[37]

2.4. Performance Metrics

To evaluate the performance of different MOEAs for a given problem, each algorithm was ran for the same number of generations, and the resulting solutions, known as Pareto front approximations, are compared using functions that measure two qualities: (i) solution accuracy, i.e., to determine how similar the solution is to the true Pareto front and (ii) solution diversity, i.e., to evaluate how well distributed are the points in the solution. We selected the top 5 most used metrics according to a recent literature survey [38]. These include, in order of popularity, hypervolume (*HV*), generational distance (*GD*), epsilon indicator (ϵ), inverted generational distance (*IGD*), and coverage (*C*). Based on a recent study [39], we considered the average Hausdorff distance (Δ_p), that combines *GD* and *IGD*, and hence simplified the number of performance metrics to 4 in our study. These metrics are defined as follows:

HV: This metric measures the volume occupied by the union of the smallest hyperboxes formed by each point in the Pareto front approximation and the reference point. This Pareto front approximation

corresponds to the solution of a specific MOEA (denoted as PF) and the reference point is selected to be greater or equal to the maximum value attainable by any objective, which in our case is $\vec{1}$ (Figure 1a):

$$HV = \bigcup_{i \in I} \text{Volume}(\text{Box}(PF_i, \vec{1})) \quad (11)$$

where I is the index set of PF points.

GD : This metric measures the distance between the solution PF and the best Pareto front approximation determined by combining non-dominated points from all MOEA solutions of a specific case study, denoted PF^* . More specifically, GD corresponds to the average Euclidean distance between each point in PF and the nearest point in PF^* , denoted as $d_i = \min_{k \in K} \left(\sum_{j \in J} (PF_{ij} - PF_{kj}^*)^2 \right)^{\frac{1}{2}}$, where I ($i \in I$), K ($k \in K$), and J ($j \in J$) correspond to the index sets of PF points, PF^* points, and problem objectives, respectively (Figure 1b):

$$GD = \frac{\sum_{i \in I} d_i}{|I|} \quad (12)$$

IGD : This metric measures the distance between PF and PF^* . It is determined by the average Euclidean distance between each point in PF^* and the nearest point in PF denoted as $\hat{d}_k = \min_{i \in I} \left(\sum_{j \in J} (PF_{kj}^* - PF_{ij})^2 \right)^{\frac{1}{2}}$ (Figure 1b):

$$IGD = \frac{\sum_{k \in K} \hat{d}_k}{|K|} \quad (13)$$

Δ_p : This metric combines GD and IGD metric and thus has superior properties [39]:

$$\Delta_p = \max(GD, IGD) \quad (14)$$

C : This metric determines the fraction of PF^* captured by the solution PF (Figure 1c):

$$C = \frac{|PF \cap PF^*|}{|PF^*|} = \frac{|\{k \in K : \exists i \in I \text{ such that } PF_{kj}^* = PF_{ij} \text{ for all } j \in J\}|}{|K|} \quad (15)$$

ϵ : This metric is the additive epsilon indicator [40] that measures the smallest value to be added to any point in PF to make it non-dominated with respect to some point in PF^* . In other words, it is the smallest value ϵ such that for any solution in PF^* there is at least one solution in PF that is not worse by a difference of ϵ (Figure 1d):

$$\epsilon = \inf\{\epsilon \in \mathbb{R} : \text{for all } i \in I \exists k \in K \text{ such that } PF_{ij} + \epsilon \geq PF_{kj}^* \text{ for all } j \in J\} \quad (16)$$

Use of these metrics can be illustrated with a two-objectives design example with 4 generations of improving Pareto front approximations, where the final Pareto front is used as a reference (i.e., PF^*) (Figure 1e). As the Pareto fronts contain points that dominate the previous generations, all metrics decrease monotonically with the exception of C that increases to a value of 1 when both Pareto front approximation and reference are the same (Figure 1f).

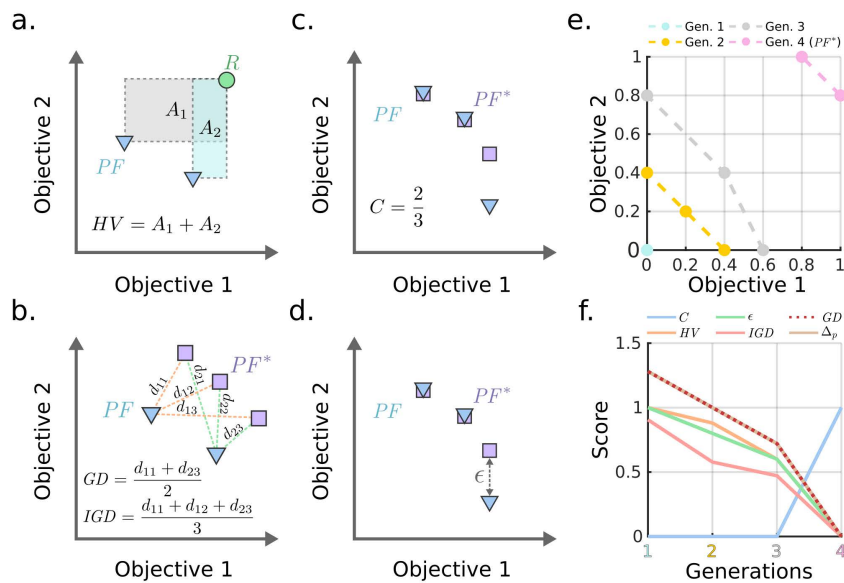


Figure 1. (a–d) Conceptual illustration of performance metrics of MOEAs for a two-objectives design problem. PF and PF^* correspond to the Pareto front approximation and the best Pareto front available, respectively. The reference point R must always dominate all solutions in PF . (e–f) An example of Pareto fronts with 2 dimensions and associated metrics. The 4th generation corresponds to PF^* used as a reference for comparison.

2.5. Algorithm Parameters

All parameters used in the simulations of this study were left as default except the following ones. The total number of generations was set to be 200, which was sufficient to reach high quality solutions for the problems of this study. In addition, the population size was set to be 100 for all algorithms unless noted otherwise. All problems were solved in triplicates with unique random number generator seeds.

2.6. Metabolic Models

For all simulations, we used a core *E. coli* model, downloaded from the BiGG database (<https://bigg.ucsd.edu>) [41], that captures the most important metabolic pathways [27]. The product synthesis pathways for each module correspond to native *E. coli* pathways together with well-characterized heterologous pathways for the synthesis of propanol [42], butanol [43], isobutanol [44], and pentanol [42]. The metabolic reactions associated with these pathways are described in the software implementation (Supplementary Material 1).

2.7. Implementation

The simulations were performed using the ModCell2 software framework [8]. The MOEAs are implemented in the PlatEMO Matlab library [28], except *gamultiobj* which is implemented as part of the Matlab Optimization Toolbox. HV was calculated using the *hv* package [45]. All computations were executed in a computer with the Arch Linux operative system, Intel Core i7-3770 processor, and 32 GB of random-access memory. The Matlab 2018b code used to generate the results of this manuscript is available in Supplementary Material 1 and <https://github.com/trinhlab/compare-moea>.

3. Results and Discussion

3.1. Case 1: A 3-Objectives Design Problem

We first formulated a design problem that considers an *E. coli* core model and 3 production modules based on the endogenous acetate, D-lactate, and ethanol biosynthesis pathways (Figure 2a).

We used all MOEAs to solve for the problem by setting the following design parameters: $wGCP$ design objective, a maximum number of reaction deletions $\alpha = 3$, and no module reactions $\beta = 0$. These design parameters were sufficiently restrictive to generate conflicting objectives. A total coverage of PF^* ($C = 1$) was reached within 20 generations by several algorithms (Figure 2b,e,h,i) and by *gamultiobj* after 150 generations (Figure 2k), while the remaining algorithms could not attain C values above 0.8 (Figure 2c,d,f,g,j,l). In particular, MaOEADDFC and BiGE obtained the worst C , ϵ , and Δ_p values (Figure 2m). Although C , ϵ , and Δ_p values of BiGE indicated inferior performance, this algorithm had the lowest HV since it generated only one point with a high objective value (Figure 2o). Due to the simplicity of the problem, every algorithm except MaOEADDFC, tDEA, and BiGE converged to very similar Pareto fronts (Figure 2n–x), and 5 of them reached $C = 1$, indicating convergence to the reference Pareto front (Figure 2y).

3.2. Case 2: A 10-Objectives Design Problem

Using the same model and design parameters as in Case 1, we expanded the number of objectives to represent a more realistic scenario. These objectives correspond to 6 endogenous pathways for biosynthesis of D-lactate, acetate, ethanol, formate, pyruvate and L-glutamate and 4 heterologous pathways for biosynthesis of propanol, butanol, isobutanol, and pentanol. The additional objectives increased the difficulty of the problem, leading to more notable difference among algorithm performances (Figure 3a–k). The SPEA2SDE algorithm displayed consistent improvement of C as generations progressed, and quickly reached the smallest values of ϵ and Δ_p (Figure 3h). Other algorithms, including ARMOEA and MOEAIGDNS, also improved their ϵ values with the increasing number of generations and reached the same final values of ϵ and Δ_p as SPEA2SDE (Figure 3a,d). However, SPEA2SDE approached $C \cong 0.6$, which is twice the value reached by the next best-performing methods (Figure 3l). Remarkably, SPEA2SDE outperformed every other algorithm in all metrics, except HV . The HV metric continues to show bias towards algorithms that generated a small number of points and scored poorly in other metrics.

3.3. Case 3: Use of Large Population Size Overcomes Poor MOEA Performance

Increasing the number of objectives often leads to a combinatorial explosion of the number of feasible Pareto optimal points and consequently causes poor MOEA performance. This problem can be alleviated by using a larger population size to sample a broader volume of solution space [46]. To test this strategy for the 10-objectives design problem above, we increased the population size from 100 to 1000 individuals while all other parameters remained unchanged. The result showed that ARMOEA, MOEAIGDNS, NSGAI, SPEA2SDE (the best performer in Case 2), and *gamultiobj*, could reach C of 0.7, ϵ of 0, and Δ_p of 0 in fewer than 50 generations (Figure 4a,d,g,h,j). These 5 algorithms also yielded very similar final values across all metrics (Figure 4l). The remaining algorithms converged to considerably lower C values (Figure 4b,c,e,f,i,k). Remarkably, NSGAI/*gamultiobj*, that is not considered a many-objective solver, performed better than more recent many-objective algorithms such as NSGAIII.

One limitation of using larger populations is an increased cost in computational time. We observed that a 10-fold increase in population sizes resulted in a 10-fold increase in the run times (Figure 5). Nonetheless, all metrics reached a stable value in the top performing algorithms after 50 generations (out of 200 total), suggesting that fewer generations were needed by using a larger population size. Among the best performing algorithms with large population sizes, *gamultiobj*, implemented in the Matlab Optimization Toolbox, required the shortest run time, followed by NSGAI and SPEA2SDE implemented in PlatEMO.

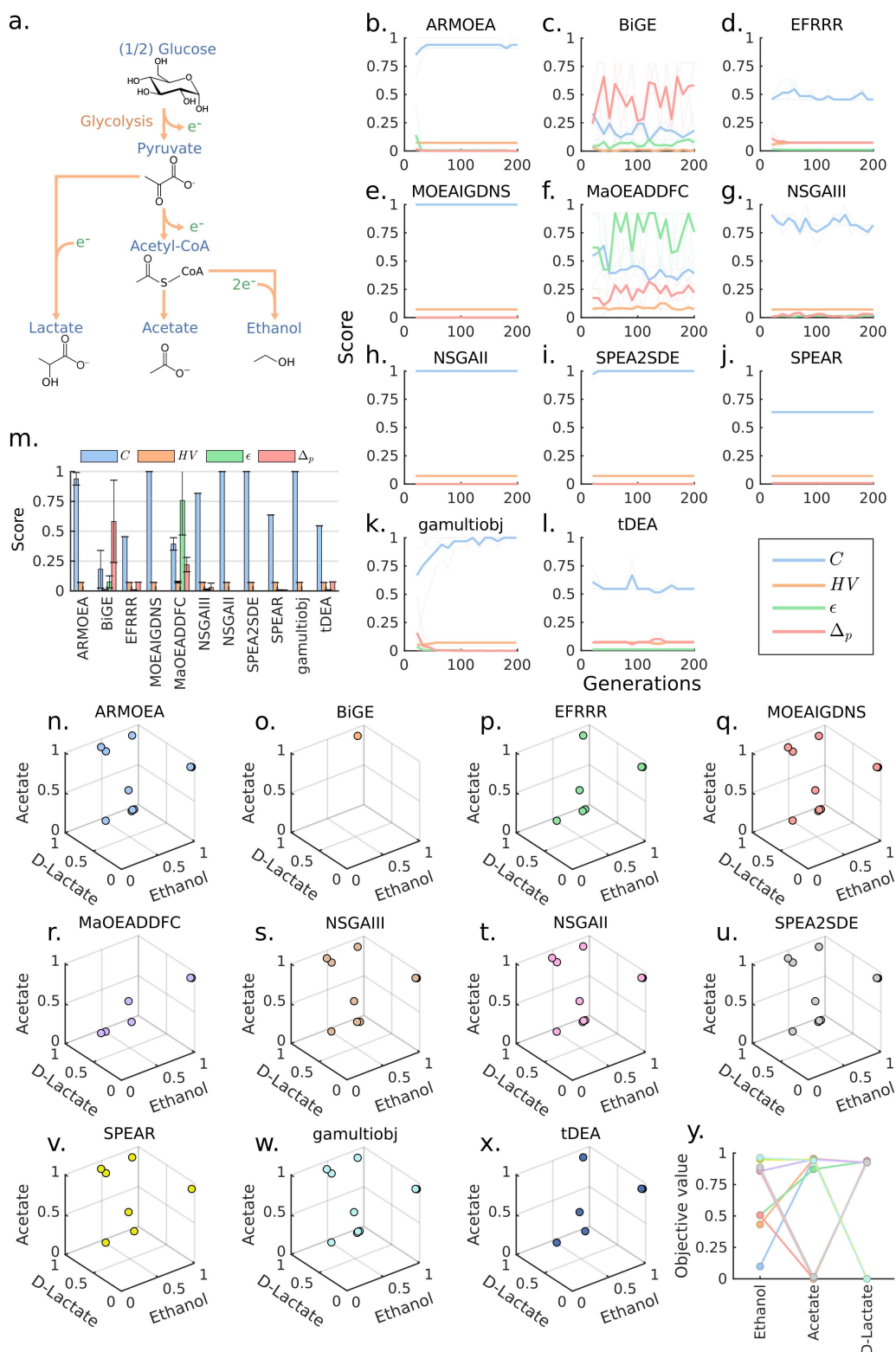


Figure 2. Comparison of MOEAs for a 3-objectives design problem. **(a)** The simplified metabolic pathways for conversion of glucose to the target products. Reducing equivalents are presented with e^- . **(b–l)** Generation-dependent performance metrics for various MOEAs. **(m)** Performance metrics for various MOEAs at the last generation. **(n–x)** Pareto fronts of various MOEAs at the last generation. It should be noted that only the first replicate is plotted for clear illustration. **(y)** Reference Pareto front (PF^*). Each line represents a solution.

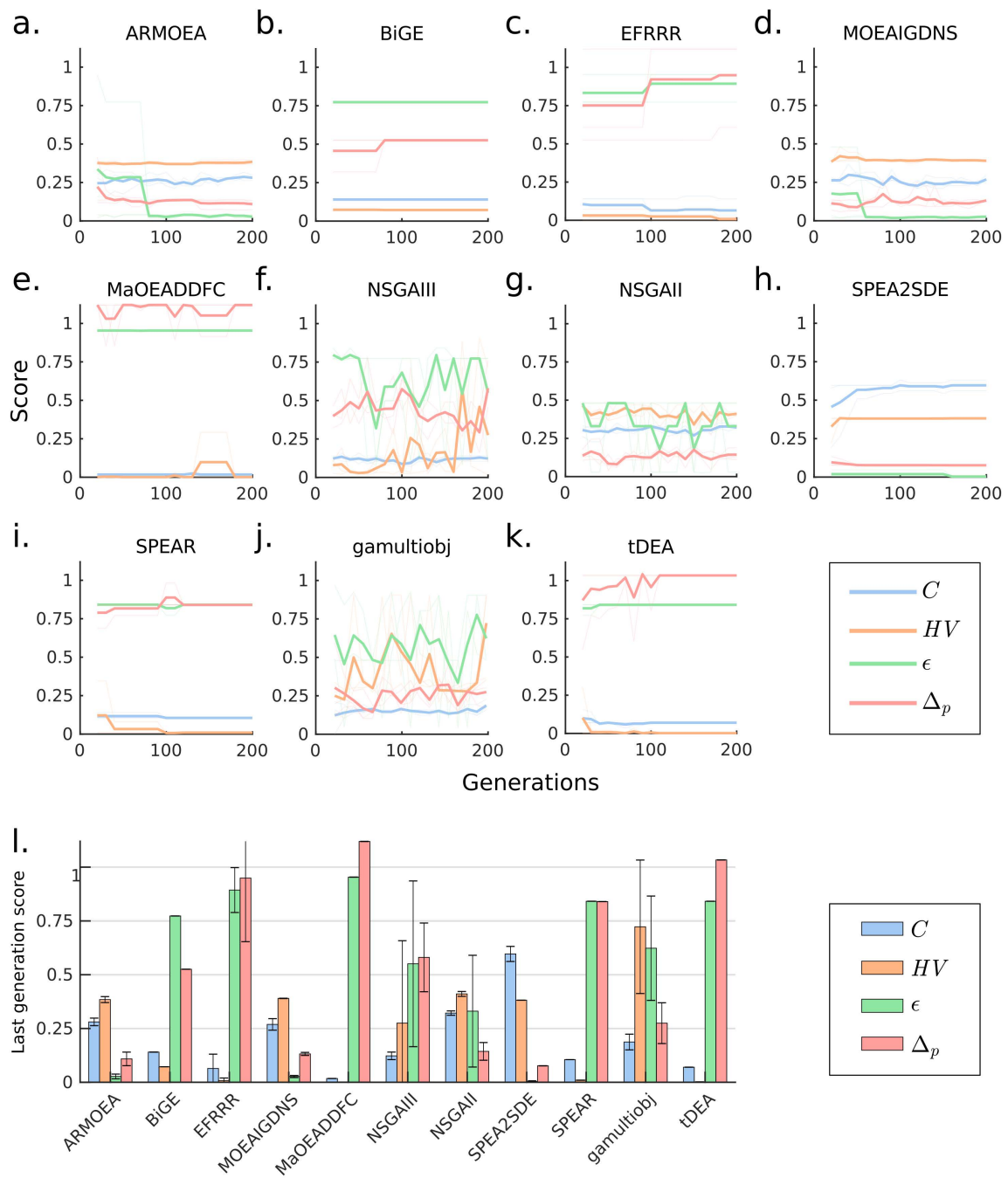


Figure 3. Comparison of MOEAs for a 10-objectives design problem. (a–k) Generation-dependent performance metrics for various MOEAs. (l) Performance metrics for various MOEAs at the last generation.

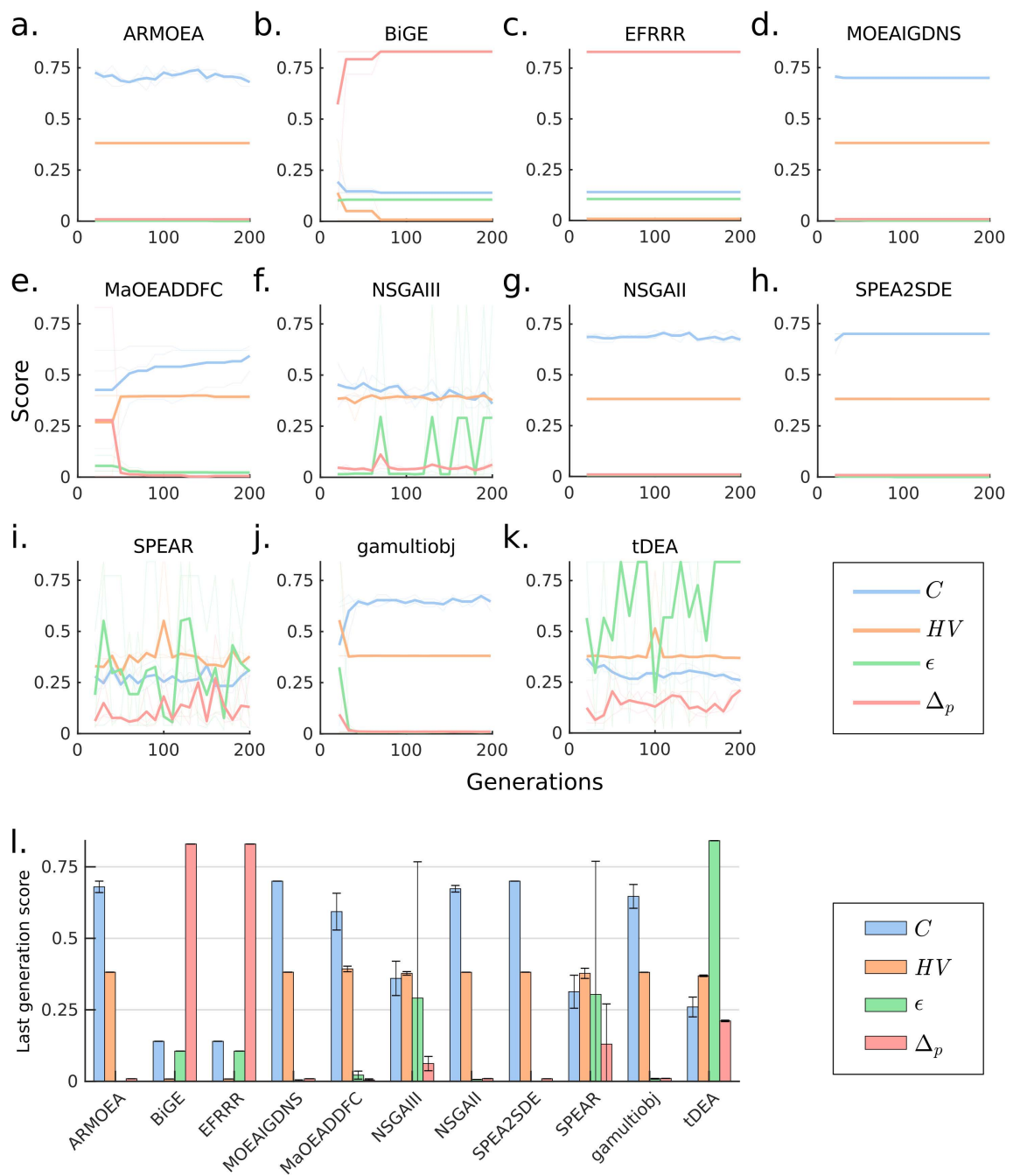


Figure 4. Comparison of MOEAs for a 10-objectives design problem with larger population sizes (a–k) Generation-dependent performance metrics for various MOEAs. (l) Performance metrics for various MOEAs at the last generation.

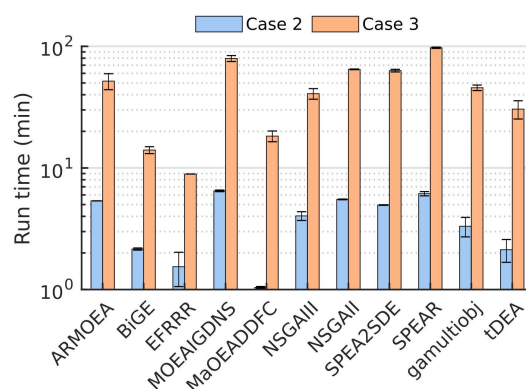


Figure 5. Wall-clock run times for the 10-objectives design problem with population sizes of 100 (Case 2) and 1000 (Case 3).

4. Conclusions

In this study, we evaluated the performance of several MOEAs to solve the modular cell design problem. SPEA2SDE, the recently developed many-objectives method, was the best performing MOEA for limited population sizes in our study. However, for sufficiently large populations, several algorithms attained the best results, including the well-established NSGAI, which performed better than more recently developed many-objective MOEAs. We used the most popular performance metrics to compare MOEAs and found that the coverage (C) metric is the most valuable indicator. This metric can provide an intuitive quantitative meaning and tends to increase monotonically with the number of generations simulated. In contrast, hypervolume (HV) generally did not differentiate algorithm performance and was misleading in some scenarios where an algorithm generated very few solutions. Overall, these results highlight the need for empirical testing of MOEAs towards specific problems and the population size as a more important factor in performance than the unique heuristics commonly used by different algorithms. For the application of modular cell engineering, efficient MOEAs will enable the design of modular cell(s) compatible with many product synthesis modules for large-scale metabolic networks and the identification of more diverse and better solutions that will provide more viable options for practical implementation.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2227-9717/7/6/361/s1>.

Author Contributions: C.T. initiated and supervised the study. S.G. and C.T. designed experiments. S.G. performed simulation experiments and analyzed data. S.G. and C.T. wrote and approved the manuscript.

Funding: This research was funded by the NSF CAREER Award (NSF#1553250) and the Center of Bioenergy Innovation (CBI), U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the funding agencies.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Coello, C.A.C.; Lamont, G.B. *Applications of Multi-Objective Evolutionary Algorithms*; World Scientific: Singapore, 2004; Volume 1.
2. Rangaiah, G.P. *Multi-Objective Optimization: Techniques and Applications In Chemical Engineering*; World Scientific: Singapore, 2009; Volume 1.
3. Trinh, C.T.; Mendoza, B. Modular cell design for rapid, efficient strain engineering toward industrialization of biology. *Curr. Opin. Chem. Eng.* **2016**, *14*, 18–25. [[CrossRef](#)]
4. Lee, S.Y.; Kim, H.U.; Chae, T.U.; Cho, J.S.; Kim, J.W.; Shin, J.H.; Kim, D.I.; Ko, Y.S.; Jang, W.D.; Jang, Y.S. A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2019**, *2*, 18. [[CrossRef](#)]
5. Nielsen, J.; Keasling, J. Engineering Cellular Metabolism. *Cell* **2016**, *164*, 1185–1197. [[CrossRef](#)] [[PubMed](#)]

6. Bonvoisin, J.; Halstenberg, F.; Buchert, T.; Stark, R. A systematic literature review on modular product design. *J. Eng. Des.* **2016**, *27*, 488–514. [[CrossRef](#)]
7. Trinh, C.T. Elucidating and reprogramming *Escherichia coli* metabolisms for obligate anaerobic n-butanol and isobutanol production. *Appl. Microbiol. Biotechnol.* **2012**, *95*, 1083–1094. [[CrossRef](#)] [[PubMed](#)]
8. Garcia, S.; Trinh, C. Multiobjective strain design: A framework for modular cell engineering. *Metab. Eng.* **2019**, *51*. [[CrossRef](#)] [[PubMed](#)]
9. Trinh, C.T.; Liu, Y.; Conner, D.J. Rational design of efficient modular cells. *Metab. Eng.* **2015**, *32*, 220–231. [[CrossRef](#)]
10. Garcia, S.; Trinh, C. Modular design: Applying proven engineering principles to biotechnology. **2019**, under review.
11. Trinh, C.T.; Li, J.; Blanch, H.W.; Clark, D.S. Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. *Appl. Environ. Microbiol.* **2011**, *77*, 4894–4904. [[CrossRef](#)]
12. Wilbanks, B.; Layton, D.; Garcia, S.; Trinh, C. A Prototype for Modular Cell Engineering. *ACS Synthetic Biol.* **2017**. [[CrossRef](#)]
13. Layton, D.S.; Trinh, C.T. Engineering modular ester fermentative pathways in *Escherichia coli*. *Metab. Eng.* **2014**, *26*, 77–88. [[CrossRef](#)] [[PubMed](#)]
14. Layton, D.S.; Trinh, C.T. Expanding the modular ester fermentative pathways for combinatorial biosynthesis of esters from volatile organic acids. *Biotechnol. Bioeng.* **2016**, *113*, 1764–1776. [[CrossRef](#)] [[PubMed](#)]
15. Layton, D.S.; Trinh, C.T. Microbial synthesis of a branched-chain ester platform from organic waste carboxylates. *Metab. Eng. Commun.* **2016**, *3*, 245–251. [[CrossRef](#)] [[PubMed](#)]
16. Wierzbicki, M.; Niraula, N.; Yarrabothula, A.; Layton, D.S.; Trinh, C.T. Engineering an *Escherichia coli* platform to synthesize designer biodiesels. *J. Biotechnol.* **2016**, *224*, 27–34. [[CrossRef](#)] [[PubMed](#)]
17. Lee, J.; Trinh, C.T. De novo Microbial Biosynthesis of a Lactate Ester Platform. *bioRxiv* **2018**. [[CrossRef](#)]
18. Coello, C.A.C.; Lamont, G.B.; Van Veldhuizen, D.A. *Evolutionary Algorithms for Solving Multi-Objective Problems*; Volume 5, Springer: Berlin, Germany, 2007.
19. Marler, R.T.; Arora, J.S. Survey of multi-objective optimization methods for engineering. *Struct. Multidisc. Optim.* **2004**, *26*, 369–395. [[CrossRef](#)]
20. Li, B.; Li, J.; Tang, K.; Yao, X. Many-objective evolutionary algorithms: A survey. *ACM Comput. Surv. (CSUR)* **2015**, *48*, 13. [[CrossRef](#)]
21. Matlab Documentation Gamultiobj Algorithm. Available online: <https://www.mathworks.com/help/gads/gamultiobj-algorithm.html>. (accessed on 4 February 2019).
22. Kalyanmoy, D. *Multi Objective Optimization Using Evolutionary Algorithms*; John Wiley and Sons: Chichester, UK, 2001.
23. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [[CrossRef](#)]
24. Zitzler, E.; Laumanns, M.; Thiele, L. *SPEA2: Improving the Strength Pareto Evolutionary Algorithm*; TIK-Report; ETH Zurich: Zurich, Switzerland, 2001; Volume 103.
25. Zitzler, E.; Deb, K.; Thiele, L. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evol. Comput.* **2000**, *8*, 173–195. [[CrossRef](#)]
26. Deb, K.; Thiele, L.; Laumanns, M.; Zitzler, E. Scalable multi-objective optimization test problems. In Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600), Honolulu, HI, USA, 12–17 May 2002; Volume 1, pp. 825–830.
27. Palsson, B.Ø. *Systems Biology: Constraint-Based Reconstruction and Analysis*; Cambridge University Press: Cambridge, UK, 2015.
28. Tian, Y.; Cheng, R.; Zhang, X.; Jin, Y. PlatEMO: A MATLAB platform for evolutionary multi-objective optimization. *IEEE Comput. Intell. Mag.* **2017**, *12*, 73–87. [[CrossRef](#)]
29. Tian, Y.; Zhang, X.; Cheng, R.; Jin, Y. A multi-objective evolutionary algorithm based on an enhanced inverted generational distance metric. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; pp. 5222–5229.
30. Tian, Y.; Cheng, R.; Zhang, X.; Cheng, F.; Jin, Y. An indicator-based multiobjective evolutionary algorithm with reference point adaptation for better versatility. *IEEE Trans. Evol. Comput.* **2018**, *22*, 609–622. [[CrossRef](#)]
31. Yuan, Y.; Xu, H.; Wang, B.; Zhang, B.; Yao, X. Balancing convergence and diversity in decomposition-based many-objective optimizers. *IEEE Trans. Evol. Comput.* **2016**, *20*, 180–198. [[CrossRef](#)]

32. Cheng, J.; Yen, G.G.; Zhang, G. A many-objective evolutionary algorithm with enhanced mating and environmental selections. *IEEE Trans. Evol. Comput.* **2015**, *19*, 592–605. [[CrossRef](#)]
33. Jiang, S.; Yang, S. A strength Pareto evolutionary algorithm based on reference direction for multiobjective and many-objective optimization. *IEEE Trans. Evol. Comput.* **2017**, *21*, 329–346. [[CrossRef](#)]
34. Yuan, Y.; Xu, H.; Wang, B.; Yao, X. A new dominance relation-based evolutionary algorithm for many-objective optimization. *IEEE Trans. Evol. Comput.* **2016**, *20*, 16–37. [[CrossRef](#)]
35. Li, M.; Yang, S.; Liu, X. Bi-goal evolution for many-objective optimization problems. *Artif. Intell.* **2015**, *228*, 45–65. [[CrossRef](#)]
36. Deb, K.; Jain, H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Trans. Evol. Comput.* **2014**, *18*, 577–601. [[CrossRef](#)]
37. Li, M.; Yang, S.; Liu, X. Shift-based density estimation for Pareto-based algorithms in many-objective optimization. *IEEE Trans. Evol. Comput.* **2014**, *18*, 348–365. [[CrossRef](#)]
38. Riquelme, N.; Von Lüken, C.; Baran, B. Performance metrics in multi-objective optimization. In Proceedings of the 2015 Latin American Computing Conference (CLEI), Arequipa, Peru, 19–23 October 2015; pp. 1–11.
39. Schutze, O.; Esquivel, X.; Lara, A.; Coello, C.A.C. Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Trans. Evol. Comput.* **2012**, *16*, 504–522. [[CrossRef](#)]
40. Zitzler, E.; Thiele, L.; Laumanns, M.; Fonseca, C.M.; Da Fonseca Grunert, V. *Performance Assessment of Multiobjective Optimizers: An Analysis And Review*; TIK-Report; ETH Zurich: Zurich, Switzerland, 2002; Volume 139.
41. King, Z.A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J.A.; Ebrahim, A.; Palsson, B.O.; Lewis, N.E. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **2015**, *44*, D515–D522. [[CrossRef](#)]
42. Tseng, H.C.; Prather, K.L. Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17925–17930. [[CrossRef](#)] [[PubMed](#)]
43. Shen, C.R.; Lan, E.I.; Dekishima, Y.; Baez, A.; Cho, K.M.; Liao, J.C. High titer anaerobic 1-butanol synthesis in *Escherichia coli* enabled by driving forces. *Appl. Environ. Microbiol.* **2011**. . [[CrossRef](#)] [[PubMed](#)]
44. Atsumi, S.; Hanai, T.; Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008**, *451*, 86. [[CrossRef](#)] [[PubMed](#)]
45. Fonseca, C.M.; Paquete, L.; López-Ibáñez, M. An improved dimension-sweep algorithm for the hypervolume indicator. In Proceedings of the 2006 IEEE International Conference on Evolutionary Computation, Vancouver, BC, Canada, 16–21 July 2006; pp. 1157–1163.
46. Ishibuchi, H.; Sakane, Y.; Tsukamoto, N.; Nojima, Y. Evolutionary many-objective optimization by NSGA-II and MOEA/D with large populations. In Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 11–14 October 2009; pp. 1758–1763.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).