In Proceedings of the Fourth IEEE Conference on Evolutionary Computation, Piscataway New Jersey : IEEE Service Center , 1997.

Exogenous Parameter Selection in a Real-valued Genetic Algorithm

Charles E. Kaiser, Jr., Gary B. Lamont, Laurence D. Merkle, George H. Gates, Jr., Ruth Pachter

Abstract— **To evaluate the performance of a real-valued genetic algorithm (GA) exploiting domain knowledge, we systematically evaluate the effect of exogenous parameters using analysis of variance. The GA platform used for this study is Genocop-III, a real-valued, co-evolutionary algorithm implementation for numerical optimization. We use the protein structure prediction (PSP) problem as our test domain. Nearly all PSP research assumes the native conformation of a protein corresponds to its global minimum free energy state. Thus, our application integrates Genocop-III with our implementation of the CHARMM energy model as the objective function. Results and conclusions drawn from an extensive experiment set using the polypeptide [Met]-Enkephalin are presented from an exogenous parameter selection perspective.**

Keywords— **Evolutionary Algorithms, Exogenous Parameter Selection, Polypeptide Structure Prediction, Protein Folding, Real-Valued Genetic Algorithms**

I. Introduction

Genetic algorithms, a class of stochastic search and optimization algorithms loosely based on natural evolution and Darwin's theory of "survival of the fittest", are amply described in the literature (e.g. Holland [9], Goldberg [8], Bäck [1], Michalewicz [18]). Although most research has focused primarily on binary-coded GAs, a growing number of practitioners and theorists are examining the performance of real-valued GAs, especially for numerical optimization problems [1], [18]. Unfortunately, implementations of these real-valued GAs have been used on very few test suites or application domains, and most have many more execution parameters than their simpler, binary-valued cousins. We have integrated a real-valued genetic algorithm package, Genocop-III, with our own polypeptide energy model objective function to identify, at a coarse granularity, which exogenous parameters have the greatest impact on the search algorithm's performance (effectiveness and efficiency).

We use our CHARMM polypeptide energy model as the GA's objective function to attempt to identify the native, minimum energy conformation of a small pentapeptide, [Met]-Enkephalin. Given only the amino acid sequence for an arbitrary protein, the prediction of it's native conformation (i.e. molecular structure) is beyond current computational capabilities. This problem is commonly referred to as the *protein structure prediction* (PSP) problem and it's solution has numerous potential applications [3]. More precisely, the PSP problem is a subproblem of the protein folding problem (PFP). The PFP is also concerned with the *path* by which proteins fold to their native conformation(s). Solutions to PSP nearly always assume the native conformation corresponds to the global minimum free energy state of the system. Exploiting this assumption requires the development of efficient global energy minimization techniques. This is a difficult optimization problem because of the non-linear and multi-modal nature of the energy function. For example, [Met]-Enkephalin is estimated to have more than 10^{11} locally optimal conformations. For detailed insight into the protein folding problem and protein structure prediction consult [4], [15], [20], [21], [25]. The polypeptide energy models to which GAs have been applied vary from lattice representations [5], [24] to simplified continuum proteins [22], fixed backbones [23], polypeptide-specific full-atom models, and general full-atom models [7], [17], [14].

We have transformed a genetic algorithm platform into a *stronger* search algorithm by incorporating "natural" real-valued data structures that capture problem specific domain knowledge. Although this limits the algorithm to the specific problem, it enhances its effectiveness in predicting polypeptide structures. To evaluate the Genocop-III GA platform, we suggest and perform a number of PSP experiments that explore the impact of *exogenous* search parameters.

II. Background

This section discusses the three components of our integrated optimization algorithm. First, we elaborate the crucial Genocop-III parameters that are evaluated in II-A. Then, an overview of our CHARMM energy function is presented in II-B. Finally, the domain knowledge that we added to the algorithm is highlighted in II-C.

A. Genocop-III Parameters

The input parameters for Genocop-III are divided into four categories: static, domain constraints [1], linear constraints, and operator probability distributions. Additionally, the static constraints and operator probability distributions are related to both the algorithm domain and the application domain, while the domain and linear constraints are characteristics of the application domain only. Then there is the class of parameters we call *exogenous*. These are the static

Charles E. Kaiser, Jr., is with the Air Force Global Weather Center. E-mail: kaiserc@afgwc.offutt.af.mil

Gary B. Lamont is a professor in the Department of Electrical and Computer Engineering, Graduate School of Engineering, Air Force Institute of Technology. E-mail: lamont@afit.af.mil

Laurence D. Merkle is with the Center for Plasma Theory and Computation, Phillips Laboratory, Kirtland AFB, NM 87117. E-mail: lmerkle@arthur.plk.af.mil

George H. Gates, Jr., and Ruth Pachter are with Wright Laboratory, 3005 P St., Ste. 1, Wright-Patterson AFB, OH 45433-7702. E-mail:{gatesgh, pachterr}@ml.wpafb.af.mil

[1] Here the phrase "Domain Constraints" is used in a context more limited than normally used in computer science, specifically, the allowable range of specified variables. If a domain constraint is not defined for a variable, it defaults to the architecture dependent range.

parameters and operator probabilities that are strictly in the algorithm domain. A clear grasp of exogenous parameters can be had when it is understood that if non-exogenous parameters are changed, a new problem is solved, but if exogenous parameters are changed, the performance of the algorithm will change, but the same problem is solved. It is the effects of these exogenous parameters that we wish to evaluate.

The static input parameters for Genocop-III are presented in Table I. Only those below the double line are exogenous parameters. Descriptions of these parameters can be found in [19]. The table also shows the value(s) used for each parameter during the test runs.

TABLE I
STATIC INPUT PARAMETERS

Description	Values
Total number of variables	24
Number of nonlinear equality constraints	0
Number of nonlinear inequality constraints	18
Number of linear inequality constraints	0
Number of variable constraints	24
Number of operators	10
Number of total evaluations	50,000
Init method for reference population	multiple
Init method for search population	multiple
Objective function type	min
Test case number	15
EPSILON for equalities	0.001
Random number seed 1	0 - 31328
Random number seed 2	0 - 30081
Operator frequency control	adaptive
Size of reference population	20, 50
Size of search population	20, 50, 70
Period of evaluation of reference population	50, 100, 150
Number of offspring for each ref population eval	10, 20, 30
Selection of ref point to repair search point	random, ordered
Selection of repair method for search population	random, deterministic
Prob. of replacement for search population	0.05, 0.2, 0.5

The operator probability distribution is a sequence of real numbers indicating the relative frequency for each operator. If operator frequency control is *fixed*, these numbers are normalized by the algorithm. However, if operator frequency control is *adaptive*, operators with nonzero values are assigned a starting relative frequency midway between the upper and lower bounds. Those with a zero value are assigned a relative frequency of zero. To minimize the combinatoric number of experiments, adaptive operator frequency control is used.

B. CHARMM Energy Model

This section discusses the objective function used in our energy minimization application and the real-valued encoding scheme. The objective function we seek to minimize is based on the CHARMM [2] force field energy model which is more general than other available energy models including AMBER and ECEPP/2 which we have employed elsewhere [16].

$$
\begin{aligned}
E = \quad & \sum_{(i,j)\in\mathcal{B}} K_{r_{ij}}(r_{ij}-r_{eq})^2 + \\
& \sum_{(i,j,k)\in\mathcal{A}} K_{\Theta_{ijk}}(\Theta_{ijk}-\Theta_{eq})^2 + \\
& \sum_{(i,j,k,l)\in\mathcal{D}} K_{\Phi_{ijkl}}[1+\cos(n_{ijkl}\Phi_{ijkl}-\gamma_{ijkl})] + \\
& \sum_{(i,j)\in\mathcal{N}} \left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^{6} + \frac{q_i q_j}{4\pi\varepsilon r_{ij}}\right] + \\
& \frac{1}{2}\sum_{(i,j)\in\mathcal{N}'} \left[\left(\frac{A_{ij}}{r_{ij}}\right)^{12} - \left(\frac{B_{ij}}{r_{ij}}\right)^{6} + \frac{q_i q_j}{4\pi\varepsilon r_{ij}}\right]
\end{aligned}
\tag{1}
$$

Where

- $\mathcal{B}$ is the set of bonded atom pairs,
- $\mathcal{A}$ is the set of atom triples defining bond angles,
- $\mathcal{D}$ is the set of atom 4-tuples defining dihedral angles,
- $\mathcal{N}$ is the set of non-bonded atom pairs,
- $\mathcal{N}'$ is the set of position 1-4 interaction pairs,
- r_{ij} is the distance between atoms i and j,
- Θ_{ijk} is the angle formed by atoms i, j, and k,
- Φ_{ijkl} is the dihedral angle formed by atoms i, j, k, and l,
- q_i is the partial atomic charges of atom i,
- the $K_{r_{ij}}$'s, r_{eq}'s, $K_{\Theta_{ijk}}$'s, Θ_{eq}'s, $K_{\Phi_{ijkl}}$'s, γ_{ijkl}'s, A_{ij}'s, B_{ij}'s, and ε are empirically determined constants (taken from the QUANTA parameter files).

The five terms represent the energy due to bond stretching, bond angle deformation, dihedral angle deformation, non-bonded interactions, and 1-4 interactions, respectively. The bond lengths and bond angles are relatively rigid compared to the dihedral angles which account for the overall flexibility of most molecular systems. Thus, in general, bond lengths and bond angles are treated as fixed and their energy contributions are a constant. Although a set of independent dihedral angles are the only free variables in the system given the above assumptions, the energy contribution of the dihedral term is typically insignificant compared to the magnitude of the non-bonded energy terms. In the CHARMM energy model, these terms, 4 and 5, are approximated using the Lennard-Jones potential and an electrostatic term. The Lennard-Jones potential includes both an attractive and a repulsive component to simulate the

Vanderwal's forces in the system. The electrostatic term treats all pairs of atoms as point charges and includes a small dielectric constant.

The molecule used for this case study is the pentapeptide [Met]-enkephalin. This small molecule is chosen because it has been used as a test problem for many other energy minimization investigations, and its minimum energy conformation is known (with respect to the ECEPP/2, and ECEPP/3 energy models). [Met]-enkephalin has 75 atoms and 24 independent dihedral angles. In our real-valued representation each GA individual is a vector of real variables, $\vec{x} = (x_1, \ldots, x_n) \in R^n intherange\text{-}\pi \leq x_i \leq \pi$.

C. Domain Knowledge and Constraint Development

While no general algorithmic solutions to the *protein folding problem* exist today (in spite of more than 30 years effort) a considerable body of knowledge has been amassed. A few examples are:

- ω angles typically assume either a *cis* or *trans* orientation; i.e., a unique isomerization conformation for each residue $(\pm 0, \pm 180)$ [20]
- χ_1 angles are usually -60, 60, 180 degrees $\pm$ some deviation. One can also use data from *rotamer* libraries; i.e., libraries of known side-chain structures
- Certain values for ϕ and ψ angle pair are frequently or rarely observed. These constraints can be visualized with *Ramachandran* plots [21]

Assuming bond lengths and bond angles are held constant, the search space for the fixed geometry model is $[-\pi, \pi]^n$ where n in the number of independent dihedral angles. Knowledge about the problem space can be used to constrain this search space. Most constraints can be expressed as nonlinear inequalities in one of the following generalized forms developed previously[13]:

$$0 \leq \cos(\theta - \frac{\theta_{min} + \theta_{max}}{2}) - \cos(\frac{\theta_{min} - \theta_{max}}{2}) \qquad (2)$$

$$0 \leq \cos(3\theta - \frac{\theta_{min} + \theta_{max}}{2}) - \cos(\frac{\theta_{min} - \theta_{max}}{2}) \qquad (3)$$

Inequality 2 defines the constraints for the backbone dihedral angles, $\{\phi, \psi, \omega\}$ and inequality 3 defines the constraints for the first sidechain dihedral angle, $\{\chi_1\}$.

For purposes of algorithm development, our constraint sets are adequate. An initial *loose* constraint set for [Met]-enkephalin was developed from Ramachandran plots of observed values for ϕ and ψ angles in the residues alanine and glycine [4]. Of the twenty naturally occurring amino acids, only proline and glycine have unique $\phi - \psi$ distributions. The other residues are similar to alanine. The *tight* constraints used in this set of experiments (Table II) consider the previous discussed relationships and infer additional insights from homologous molecules.

The experiment set consist of a full factorial design using treatment levels shown in Table III. Some combinations are not practical, such as generating more than the reference population size number of offspring during a reference population evaluation. The overall combinatoric analysis results in a total of 540 different experiments.

TABLE II
TIGHT CONSTRAINTS FOR [MET]-ENKEPHALIN

Dihedral	Midpoint	Radius
$\Phi_{Non-glycine}$	-120	60
$\Phi_{Glycine}$	130	70
Ψ	150	140
Ω	180	12.5
χ_1	$-60 \mid 60 \mid 180$	7.5

TABLE III
EXPERIMENT VALUES

Parameter (Short Name)			
Reference Population Size (Ref Pop)	20	50	
Search Population Size (Search Pop)	20	50	70
Periodicity of Reference Population Evaluation (Periodicity)	50	100	150
Offspring per Reference Population Evaluation (Offsprings)	10	20	30
Probability of Replacement for Search Population (Probability)	0.05	0.20	0.50
Selection of Reference Point for Repair (Ref Point)	0	1	
Repair Method (Repair)	0	1	

$0 \equiv random$, $1 \equiv deternimistic\ or\ ordered$

The preliminary experimental results are subjected to the twelve hypothesis tests detailed in Table V to determine which, if any, of the factors are significant sources of variation.

TABLE IV
HYPOTHESIS TESTS

Test Number	H_0 Not Significant H_1 Significant
1	Ref Pop
2	Search Pop
3	Periodicity
4	Offsprings
5	Probability
6	Ref Point
7	Repair
8	Ref Pop x Search Pop
9	Ref Pop x Repair
10	Ref Point vs Repair
11	Offsprings x Ref Pop
12	Probability x Repair

III. METHODOLOGY

The performance of a real-valued GA is dependent on choosing appropriate execution parameters. The objective

of this experiment set is the preliminary characterization of the effects of exogenous parameters on the performance of Genocop-III applied to the PSP problem.

Michalewicz has published results using earlier variants of Genocop [10]. Frequently reported parameters include 70 for search population size, 0.20 for probability of search population replacement, and 28 offspring per generation. Most experiments were run for 5,000 iterations in order to observe convergence. Given that Genocop is a steady state GA, it wasn't clear whether *iteration* meant evaluations or generations. Personal communications with a co-creator of Genocop-III, Girish Nazhiyath, validated that an iteration is approximately equal to a generation. Therefore, 5,000 iterations equates to 350,000 function evaluations with a population size of 70. Our previous experiments had produced good results with 20,000 to 40,000 function evaluations, therefore, 50,000 evaluations were used in this experiment set.

It is highly likely the effects of the static parameters and the operator probability distribution are cross-linked. In an attempt to decouple them, the static parameters are tested while the operator frequency control is set to *adaptive*. This allow the frequency of the operators to float in response to the needs of the algorithm. After the other parameters have been analyzed, specific operator frequencies can be studied.

The experimental results are subjected to the twelve hypothesis tests detailed in Table V to determine which, if any, of the factors are significant sources of variation.

TABLE V
HYPOTHESIS TESTS

Test Number	H_0 Not Significant or H_1 Significant
1	Reference Pop Size
2	Search Pop Size
3	Period of Ref Pop Eval
4	Offspring per Ref Pop Eval
5	Prob. of Search Pop Replacement
6	Selection of Ref Point for Repair
7	Repair Method
8	1 x 2
9	1 x 7
10	6 x 7
11	4 x 1
12	5 x 7

IV. RESULTS

Summarized data is presented in Table VI. Analysis of variance (ANOVA) results are presented in Table VII for $\alpha = 0.05$, the source of variation being tested for significant at the α level. The parameters for the experiments yielding the 5 lowest energies are shown in Table VIII. The plot of the *current best trajectory* of these 5 experiments are shown in Figure 1.

The single factor analysis is presented first. Test 1 considers the reference population size (Ref Pop) parameter.

TABLE VI
SUMMARY DATA

#	Parameter	Value	Avg Fitness	Std Dev
	Overall		-24.607	1.99
1	Ref Pop	20	-24.29	2.01
		50	-24.65	1.96
2	Search Pop	20	-24.46	1.94
		50	-24.49	1.97
		70	-24.57	2.06
3	Periodicity	50	-24.72	1.97
		100	-24.36	2.07
		150	-24.44	1.90
4	Offspring	10	-24.43	1.90
		20	-24.44	2.03
		30	-24.79	2.06
5	Probability	0.05	-23.80	1.73
		0.2	-24.78	1.92
		0.5	-24.95	2.10
6	Ref Point	random	-24.42	2.04
		ordered	-24.60	1.93
7	Repair	random	-25.21	1.87
		deterministic	-23.81	1.84

Fitness is in kcal/mol

On the average, lower energies are found in experiments with a reference population size of 50. The average runtime is greater with the reference population size of 50, which was expected due to the sorting time of the larger population[2]. The reference population size parameter is significant at the 95% confidence level but not at higher confidence levels. Test 5 considers the probability (Probability) that a repaired candidate solution will replace its parent in the search population. Of all factors, this is the second most significant. Test 7 considers the method (Repair) used to repair the candidate solution from the search population. Repair is attempted by generating a linear combination of the candidate solution and the reference point with a value γ. The repair method parameter controls how γ is determined, either randomly generate values $[0.0, 1.0]$ for γ (Random), or repeatedly generate a bisection, $\gamma = 2^{-i}$ $1 \leq i \leq 20$, until a fully feasible candidate is generated (Deterministic). This parameter is the most significant source of variance, with the random option producing the best results.

Test 2 considers the search population size (Search Pop) which is not significant at the 95% confidence level. Test 3 considers the periodicity (Periodicity) of the reference population evaluation. That is, the reference population is evaluated [3] every time this many evaluations have been

[2]Hypothesis testing was not done on run times because system loading in the multi-user environment could not be controlled. They are provided for reference only. However, the large number of experiments tends to dampen out cases were the platform was heavily loaded. Thus, the data are insightful.

[3]This nomenclature is from the Genocop documentation. It would

performed, regardless of which population those evaluations were performed on. Within the parameters of this experiment set, this input parameter is not significant. Test 4 considers the number of offspring (Offspring) to be generated for the reference population every time the reference population is evaluated. Remember that members of the reference population must be fully feasible, thus there may be no new chromosomes added to the reference population. Like the periodicity, this parameter by itself, is statistically insignificant. Test 6 considers the method (Ref Point) used to select the point, or individual, in the reference population that is used to repair an infeasible candidate solution from the search population. The options are to either select an individual based on a randomly generated number $[0.0, 1.0)$ (Random), or use the probability distribution of the reference population (Ordered). This parameter is not a significant source of variation either.

TABLE VII
Analysis of Variance Table, $\alpha = 0.05$

Variation	SS	DOF	MS	F_0	α
Ref Pop	16.28	1	16.28	5.02	3.92
Search Pop	1.21	2	0.61	0.19	3.07
Periodicity	13.09	2	6.55	2.05	3.07
Offspring	10.98	2	5.49	1.72	3.07
Prob	137.72	2	68.86	21.58	3.07
Ref Point	4.41	1	4.41	1.38	3.92
Repair	262.54	1	262.54	82.27	3.90
Ref Pop x Search Pop	1.78	2	0.89	0.28	3.07
Ref Pop x Repair	15.01	1	15.01	4.70	3.92
Ref Point x Repair	0.52	1	0.52	0.16	3.92
Offspring x Prob	6.04	4	1.51	0.47	2.45
Prob x Repair	3.93	2	1.97	0.62	3.07
Error	1653.01	518	3.19		
Total	2126.52	539			

Now we consider selected second order interactions. Test 8 considers the interaction between the size of the reference (Ref Pop) and search populations (Search Pop). Individually, the reference size is significant, but not the search size. As a source variation, the interaction between these two factors is insignificant. Test 9 considers the interaction between the reference population size (Ref Pop) and the choice of repair method (Repair). Individually, both of these factors are significant, with the choice of repair method highly significant. Their interaction is also significant, but only at the $\alpha = 0.05$ level. Test 10 considers the interaction between the method (Ref Point) used to select the point, or individual, in the reference population that is used to repair a infeasible candidate solution from the search population and the choice of repair method (Repair). Individually, the first factor was not significant while the second was extremely significant. Their interaction is an insignificant source of variation. Test 11 considers the number of offspring generated (Offspring) generated per reference population evaluations and the probability (Probability) that a repaired candidate solution will replace its parent in the search population. Individually, the first factor was not significant while the second was extremely significant. The interaction between these two factors is also insignificant. Test 12 considers the probability of replacement (Probability) and the choice of repair method (Repair). Individually, both of these factors are extremely significant. Surprisingly, the interaction between the two factors is insignificant.

The analysis of variance test results as reflected in Table VII support the above analysis. Refering to the Table, SS refers to the sum of the squares - test variance, DOF refers to degrees of freedom associated with test, MS refers to mean square and is normally equal to the SS divided by the DOF, f_0 is the MS of the test divided by the MS of the error and thus is the threshold value for determining if the hypothesis is significant or is not significant. If f_0 is relatively low, then the associated hypothesis is not significant. Thus, Table VII indicates that tests 5 and 7 reflect significant hypotheses. The Kruskal-Wallis H test was used throughout as a means of verifying ANOVA results[13].

Michalewicz reports best results using 0.20 probability of replacement with Genocop-III for nonlinear optimization problems[19]. Of the three values tested, $\{0.05, 0.2, 0.5\}$, our best results are with 0.5 on average. Granted, we are discussing two different problems. Furthermore, these results are only applicable to the specific experiments run. Additional study is needed to fully characterize this parameter. Table VIII shows the parameters and results for the 5 best experiments. The experimental trajectories are shown in Figure 1. Notice the path of the best result, # 95, has the steepest initial decent, then continues to make modest gains.

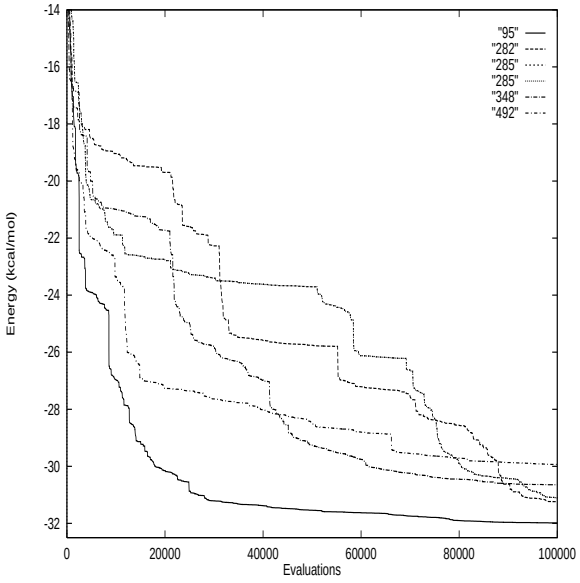
TABLE VIII
5 Best Results from Input Parameter Analysis

	95	282	285	348	492
Ref Pop	20	50	50	50	50
Search	50	20	20	50	70
Period	50	100	100	50	100
Offspring	20	30	30	20	20
Prob	0.5	0.2	0.5	0.5	0.5
Ref Pnt	1	0	0	1	1
Repair	0	1	0	1	1
Fitness	-31.98	-31.25	-31.1	-30.	-29.94
Run Time	4.56	5.28	3.59	5.11	4.72

$0 \equiv$ *Random*, $1 \equiv$ *Ordered or Derministic*, Run Time is in hours

be more accurate to say the reference population is operated upon.

Fig. 1. Trajectory Plot, 5 Best Phase I Experiments



V. Conclusions

The GA community has limited experience with domain specific evolutionary algorithms, such as those based on Genocop-III. We feel these results shed significant insight into our specific problem domain for selection of exogenous parameters and provide a starting point for more in-depth studies of parameter selection for real-valued GAs.

While analysis of this experiment set provides insight into exogenous parameters, judgement is limited. Stronger conclusions can be made for parameters with binary values, (Ref Point and Repair) than for others whose values are heuristically selected. In particular, the optimization effectiveness seems to be directly influenced by the practioners choice of reference population size, search population replacement probability, and the choice of repair method.

Our optimization results using a real-valued GA are promising. They improve when domain constraints are exploited. The domain constraints we've used are of low fidelity. An experienced biochemist could certainly develop more precise sets that may (or may not) help convergence. This work provides a small base of knowledge from which to build a more complete investigation of exogenous parameter selection for real-valued GAs. We are also investigating the addition of local minimization to Genocop-III as another operator [11]. Finally, we are developing a parallel version that exploits island model GAs with different constraints on each island and farming models to increase efficiency [12].

References

[1] Thomas Bäck. *Evolutionary Algorithms in Theory and Practice.* Oxford Press, New York, 1996.

[2] Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[3] H. Chan and K. Dill. The protein folding problem. *Physics Today*, pages 24–32, February 1993.

[4] Thomas E. Creighton, editor. *Protein Folding.* W. H. Freeman, New York, 1992.

[5] Thomas Dandekar and Patrick Argos. Potential of genetic algorithms in protein folding and protein engineering simulations. *Protein Engineering*, 5(7):637–645, 1992.

[6] David Fogel, editor. *Proceedings of the Second IEEE Conference on Evolutionary Computation*, Rundle Mall, South Australia, 1995. Causal Productions Pty Ltd. Published on CD-Rom.

[7] George H. Gates, Jr., Ruth Pachter, Laurence D. Merkle, and Gary B. Lamont. Parallel simple and fast messy GAs for protein structure prediction. In *Proceedings of the Intel Supercomputer Users' Group 1995 Annual North America Users Conference*, Beaverton, Oregon, 1995. Intel Supercomputer Systems Division.

[8] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison–Wesley Publishing Company, Reading MA, 1989.

[9] John H. Holland. *Adaptation in Natural and Artificial Systems.* MIT Press, Cambridge, MA, First MIT Press edition, 1992.

[10] Cezary Z. Janikow and Zbigniew Michalewicz. An experimental comparison of binary and floating point representations in genetic algorithms. In *International Conference on Tools for Artificial Intelligence*, pages 31–36, 1990.

[11] Charles E. Kaiser, Gary B. Lamont, Laurence D. Merkle, Jr. George H. Gates, and Ruth Pachter. Polypeptide structure prediction: Real-valued versus binary hybrid genetic algorithms. In *Proceedings of the 1997 ACM Symposium on Applied Computing*, New York, 1997. The Association for Computing Machinery.

[12] Charles E. Kaiser, Gary B. Lamont, Laurence D. Merkle, Jr. George H. Gates, and Ruth Pachter. Case studies in protein structure prediction with real-valued genetic algorithms. In *Proceedings of the 1997 8th SIAM Conference on Parallel Processing for Scientific Computing*, New York, 1997. SIAM.

[13] Charles E. Kaiser, Jr. Refined genetic algorithms for polypeptide structure prediction. Ms thesis, afit/gce/eng/96d-13, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, December 96. In Process.

[14] Charles E. Kaiser, Jr., Laurence D. Merkle, George H. Gates, Jr., Gary B. Lamont, and Ruth Pachter. Real-valued and hybrid genetic algorithms for polypeptide structure prediction. In *Applied Computing 1997: Proceedings of the 1997 Symposium on Applied Computing*, New York, 1997. The Association for Computing Machinery. Accepted for Publication.

[15] Jr. Kenneth M. Merz and Scot M. LeGrand. *The Protein Folding Problem and Tertiary Structure Prediction.* Birhäuser, Boston, 1994.

[16] Laurence D. Merkle, George H. Gates, Jr., Gary B. Lamont, and Ruth Pachter. Hybrid genetic algorithms for minimization of a polypeptide specific energy model. In *Proceedings of the Third IEEE Conference on Evolutionary Computation*, 1996.

[17] Laurence D. Merkle, Robert L. Gaulke, George H. Gates, Jr., Gary B. Lamont, and Ruth Pachter. Hybrid genetic algorithms for polypeptide energy minimization. In *Proceedings of the 1996 Symposium on Applied Computing*, New York, 1996. The Association for Computing Machinery.

[18] Zbigniew Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs.* Springer-Verlag, New York, 3rd edition, 1996.

[19] Zbigniew Michalewicz and Girish Nazhiyath. Genocop III: a coevolutionary algorithm for numerical optimization. In *In [6]*, volume 2, pages 647–651, 1995.

[20] Roger H. Pain, editor. *Mechanisms of Protein Folding.* IRL Press, New York, 1994.

[21] Lubert Stryer. *Biochemistry.* W.H. Freeman, New York, 4rd edition, 1995.

[22] Shaojian Sun. Reduced representation model of protein structure prediction: Statistical potential and genetic algorithms. *Protein Science*, 2:762–785, 1993.

[23] P. Tuffery, C. Etchebest, S. Hazout, and R. Lavery. A new approach to the rapid determination of protein side chain conformations. *Journal of Biomolecular Structure & Dynamics*, 8(6):1267, 1991.

[24] Ron Unger and John Moult. Genetic algorithms for protein folding simulations. *Journal of Molecular Biology*, 231:75–81, 1993.

[25] Maximiliano Vásquez, George Némethy, and Harold A Scheraga. Conformational energy calculations on polypeptides and proteins. *Chemical Reviews*, 94:2183–2239, 1994.