

Dead-End Elimination with Backbone Flexibility

Ivelin Georgiev¹ and Bruce R. Donald^{1,2,*}

¹Department of Computer Science, Duke University and ²Department of Biochemistry, Duke University Medical Center, Durham, NC 27708, USA

ABSTRACT

Motivation: Dead-End Elimination (DEE) is a powerful algorithm capable of reducing the search space for structure-based protein design by a combinatorial factor. By using a fixed backbone template, a rotamer library, and a potential energy function, DEE identifies and prunes rotamer choices that are provably not part of the Global Minimum Energy Conformation (GMEC), effectively eliminating the majority of the conformations that must be subsequently enumerated to obtain the GMEC. Since a fixed-backbone model biases the algorithm predictions against protein sequences for which even small backbone movements may result in a significantly enhanced stability, the incorporation of backbone flexibility can improve the accuracy of the design predictions. If explicit backbone flexibility is incorporated into the model, however, the traditional DEE criteria can no longer guarantee that the *flexible-backbone* GMEC, the lowest-energy conformation when the backbone is allowed to flex, will not be pruned.

Results: We derive a novel DEE pruning criterion, *flexible-backbone DEE (BD)*, that is provably accurate with backbone flexibility, guaranteeing that no rotamers belonging to the *flexible-backbone* GMEC are pruned; we also present further enhancements to BD for improved pruning efficiency. The results from applying our novel algorithms to redesign the $\beta 1$ domain of protein G and to switch the substrate specificity of the NRPS enzyme GrsA-PheA are then compared against the results from previous fixed-backbone DEE algorithms. We confirm experimentally that traditional-DEE is indeed not provably-accurate with backbone flexibility and that BD is capable of generating conformations with significantly lower energies, thus confirming the feasibility of our novel algorithms.

Availability: Contact authors for source code.

Contact: brd+ismb07@cs.duke.edu

1 INTRODUCTION

The main computational challenge for structure-based protein design algorithms is the combinatorial nature of the problem search space. In the standard formulation of the protein design problem, by allowing sets of amino acids at each mutable position in a protein, a potential energy function (Gordon *et al.*, 1999) is used to redesign the initial protein structure and sequence. Some form of protein flexibility is further included in most design algorithms, in order to improve the accuracy of the model (Dahiyat and Mayo, 1997a; Desjarlais and Handel, 1999; Jaramillo *et al.*, 2001; Jin *et al.*, 2003; Lilien *et al.*, 2005;

Looger *et al.*, 2003; Street and Mayo, 1999). Due to the significantly increased computational complexity when explicit backbone flexibility is incorporated into the model, the majority of the design algorithms incorporate only side-chain flexibility using a discrete rotamer library of low-energy side-chain conformations (Dunbrack, 2002; Lovell *et al.*, 2000; Ponder and Richards, 1987) and aim at stabilizing the fixed backbone template. Recent improvements have shown that allowing side-chain dihedral energy minimization from the initial rotamer conformations can result in the generation of significantly lower-energy conformations and sequences (Georgiev *et al.*, 2006b). Some successful efforts of applying different levels of backbone flexibility have also been reported (Desjarlais and Handel, 1999; Kuhlman *et al.*, 2003).

Protein design is NP-hard (Chazelle *et al.*, 2004; Pierce and Winfree, 2002). Hence, some provably-accurate algorithms, the dominant of which is the Dead-End Elimination (DEE) algorithm (Desmet *et al.*, 1992; Gordon *et al.*, 2003), apply sophisticated techniques to significantly reduce the search space of candidate conformations before the subsequent generation of the optimal solution, the Global Minimum Energy Conformation (GMEC). To increase the computational efficiency, some DEE-like algorithms use heuristic pruning techniques that no longer have guarantees about producing the GMEC (Desmet *et al.*, 2002; Shah *et al.*, 2004). Alternatively, different heuristic approaches, such as Monte Carlo sampling and genetic algorithms, have been applied (Desjarlais and Handel, 1999; Hellinga and Richards, 1991; Jin *et al.*, 2003; Kuhlman and Baker, 2000; Street and Mayo, 1999). Although such heuristics can perform reasonably well in practice, they have no guarantees about the accuracy of the design predictions: in a recent study (Voigt *et al.*, 2000), a set of heuristic approaches (including Monte Carlo) generated significantly incorrect solutions in some of the test cases.

In contrast to the GMEC-based approaches discussed above, partition functions for *ensembles* of low-energy conformations are used in (Lilien *et al.*, 2005; Stevens *et al.*, 2006) to compute a provably-accurate approximation, K^* , to the binding constant for a target protein-ligand complex; the best-scoring mutation predictions are generated based on their K^* scores.

1.1 Fixed-backbone Dead-End Elimination

The traditional DEE criterion (Desmet *et al.*, 1992; Lasters and Desmet, 1993) is applied to a model with a fixed backbone and a set of *rigid rotamers* (rotamers for which the side-chain dihedrals are fixed); we will refer to this type of

*To whom correspondence should be addressed.

DEE as *traditional-DEE*. Using rotameric energy interactions, the goal of traditional-DEE is to prune rotamers that are provably not part of the *fixed-backbone rigid-rotamer GMEC* (the GMEC when a fixed backbone and rigid rotamers are used).

The total energy of a given conformation can be represented as a sum of pairwise interactions:

$$E_T = E_r + \sum_i E(i_r) + \sum_i \sum_{j>i} E(i_r, j_s), \quad (1)$$

where E_r is the fixed template energy of the system; i_r represents rotamer identity r at position i ; $E(i_r)$ is the self-energy of i_r (combining the intra-residue and residue-to-template energies for i_r); and $E(i_r, j_s)$ is the pairwise energy between rotamers i_r and j_s . Using Eq. (1), the initial traditional-DEE criterion (Desmet *et al.*, 1992) for rotamer i_r is defined as:

$$E(i_r) + \sum_{j \neq i} \min_s E(i_r, j_s) > E(i_t) + \sum_{j \neq i} \max_s E(i_t, j_s). \quad (2)$$

When Equation (2) holds, the lowest conformational energy achievable when position i has the particular *target* rotamer identity r is higher than the worst conformational energy achievable when the same position i has the *competitor* rotamer identity t . Thus, rotamer i_r cannot be part of the fixed-backbone rigid-rotamer GMEC and can therefore be pruned from further consideration for position i .

A single *DEE cycle* loops through all positions and all target and competitor rotamers, identifying subsets of rotamers for each position that can be provably pruned from consideration in the following DEE cycles. Thus, for a system with n residue positions and at most q rotamers per position, the complexity of applying Equation (2) for a single DEE cycle is $O(q^2 n^2)$. DEE cycles are repeated until no more rotamers can be pruned, at which point the remaining conformations can be enumerated to obtain the fixed-backbone rigid-rotamer GMEC. Several provably accurate extensions to the original traditional-DEE criterion have been shown to significantly increase the pruning efficiency of the algorithm (Georgiev *et al.*, 2006a; Goldstein, 1994; Lasters and Desmet, 1993; Pierce *et al.*, 2000).

To increase the accuracy of the model while still keeping the protein backbone fixed, rotamers may be allowed to energy-minimize from their initial conformation in order to accommodate changes in the rotamer identities of surrounding positions. Such a minimization can be achieved by allowing conformation space voxel-constrained movement of the side-chain dihedrals (Lilien *et al.*, 2005). However, when rotameric energy minimization is performed, the traditional-DEE criterion can no longer guarantee that rotamers belonging to the optimal solution, the *fixed-backbone energy-minimized GMEC* (*fixed-backbone minGMEC*), will not be pruned. In Georgiev *et al.*, (2006b), MinDEE, a DEE-based algorithm that is provably accurate with respect to the fixed-backbone minGMEC, is derived. Georgiev *et al.*, (2006b) showed that traditional-DEE is indeed not provably accurate with rotameric energy minimization; when MinDEE was

used instead, conformations and sequences with considerably lower energies were generated, at the expense of slower running times.

1.2 Backbone flexibility for protein design

Due to the greatly increased computational complexity when backbone flexibility is incorporated into the protein model, many algorithms keep the backbone fixed and aim at optimizing the protein sequence and side-chain placement for the fixed template. It has been shown, however, that backbone movements can be significant for some systems (Lim *et al.*, 1994), although in other studies backbone movements are deemed to be less significant than side-chain flexibility (Najmanovich *et al.*, 2000). In general, a fixed backbone model biases the design predictions against sequences for which even small backbone movements may result in significantly improved conformational energies (Desjarlais and Handel, 1999). Hence, some sequences that are ranked low for a fixed-backbone model may become top-ranked when a model with a flexible backbone is used.

Some models implicitly incorporate a notion of backbone flexibility by scaling the atomic van der Waals (vdW) radii, thus allowing some overpacking (Dahiyat and Mayo, 1997b; Looger and Hellinga, 2001). Several efforts of explicitly incorporating different levels of backbone flexibility have also been reported (Desjarlais and Handel, 1997; Fung *et al.*, 2007; Harbury *et al.*, 1998; Kuhlman *et al.*, 2003; Su and Mayo, 1997; Zanghellini *et al.*, 2006). Harbury *et al.*, (1995, 1998) used algebraic backbone parameterization for coiled coils with exhaustive rotamer enumeration but without provable rotamer pruning. Backbone parameterization is also used in Su and Mayo, (1997) to generate a discrete set of backbones; DEE is applied to each backbone in that set to determine the respective optimal structure. In Desjarlais and Handel, (1999), a Monte Carlo/genetic algorithm is used to sample the backbone dihedral space around an initial backbone structure and to optimize the template, amino acid sequence, and side-chain placements. A successful design of a novel protein fold is reported in Kuhlman *et al.*, (2003). The approach in that work alternates between sequence optimization for a fixed backbone and backbone optimization for a fixed sequence. The backbone structures are obtained through random sampling of the backbone dihedrals [similarly to Desjarlais and Handel (1999)] and through dihedral substitution using dihedral values from the PDB. Although the approaches above have been successfully applied in practice, they can give no guarantees about the identification of the optimal solution over the backbone/sequence search space. In Fung *et al.*, (2007), a promising new algorithm that incorporates backbone flexibility by setting upper and lower bounds on pairs of $C_\alpha - C_\alpha$ distances and on the backbone dihedral angles is described. This approach does not require explicit sampling of the (ϕ, ψ) space and guarantees the identification of the optimal solution. However, the energy function used is quite simplified: it is a function mainly of $C_\alpha - C_\alpha$ distances and there is no explicit side-chain flexibility incorporated into the model. Results in Kuhlman *et al.* (2003), however, showed that backbone design without consideration of side-chain packing

can lead to considerably higher energy states of the designed structures. Moreover, a more accurate energy function and the incorporation of explicit side-chain flexibility can significantly increase the computational requirements of the design problems.

There is no previous algorithm for protein design with backbone flexibility admitting provable properties similar to traditional-DEE's for a fixed backbone. The two most important characteristics of such an algorithm should be: (1) efficient elimination of the majority of candidate sequences and conformations and (2) *provable guarantees* that the optimal solution, the *flexible-backbone rigid-rotamer GMEC* (the lowest-energy conformation with rigid rotamers and when the backbone is allowed to flex) will not be pruned during the elimination stage in (1). Although traditional-DEE efficiently eliminates the majority of candidate conformations, it does not fulfill requirement (2), since this algorithm does not take into account possible changes in the energy interactions due to conformational changes in the protein backbone. In this article, we present BD, a novel DEE-based algorithm for backbone flexibility with rigid rotamers that fulfills *both* of the requirements above. Interesting future work may involve the derivation of a provable DEE-based algorithm for backbone flexibility with rotameric side-chain dihedral minimization—the marriage of MinDEE and BD (Table 1). The term *rigid-rotamer GMEC* is a retronym; virtually all previous DEE algorithms use rigid rotamers and have provable guarantees only with respect to the fixed-backbone rigid-rotamer GMEC (Table 1).

1.3 Contributions of the article

We make the following contributions in this article:

- (1) *BD*: An efficient DEE-based criterion for provably pruning rotamers that cannot be part of the flexible-backbone rigid-rotamer GMEC;
- (2) Similarly to the enhancements for traditional-DEE and MinDEE, we present algorithmic enhancements to the initial BD criterion for improved pruning efficiency;
- (3) We combine BD and the newly-derived enhancements in a provably accurate algorithm for the identification of the flexible-backbone rigid-rotamer GMEC;
- (4) We apply our new algorithms in redesigns of the core of the $\beta 1$ domain of protein *G* ($G\beta 1$) and of the adenylation

domain of the non-ribosomal peptide synthetase (NRPS) enzyme Gramicidin Synthetase A (GrsA-PheA). $G\beta 1$ is a small protein (56 residues) that is commonly used to test computational protein design algorithms (Gordon *et al.*, 2003; Pierce *et al.*, 2000). GrsA, in concert with GrsB, makes the natural antibiotic gramicidin S, so computational redesigns of GrsA can be an important step towards novel drug discovery (Stevens *et al.*, 2006). We compare the redesign results from BD to those from traditional-DEE and MinDEE for the same proteins.

2 APPROACH

A straightforward DEE-based approach for incorporating backbone flexibility with rigid rotamers could involve the generation of a discrete set S of backbone conformations and running traditional-DEE separately for each backbone. The fixed-backbone rigid-rotamer GMEC g_b for each backbone conformation $b \in S$ can then be identified, and the lowest-energy conformation for all backbone conformations in S will simply be $\underset{b \in S}{\operatorname{argmin}} E(g_b)$. Unfortunately, although such an

approach will be provably accurate with respect to S , it can make no guarantees about the identification of the flexible-backbone rigid-rotamer GMEC, unless the set S exhaustively covers the space of possible backbone conformations. The generation of the backbones in S would thus require fine sampling of the backbone dihedral angles of each residue in the protein. Hence, the size of the set S will be exponential in the number n of residues, thus making the naive approach computationally infeasible if provable guarantees with respect to the flexible-backbone rigid-rotamer GMEC are needed.

Instead of sampling the backbone dihedral angles and explicitly generating a discrete set S of backbones, we use both *ranges* of backbone angles and real-space restraint *volumes* on backbone movement. This approach is similar both to (Fung *et al.*, 2007) and to the treatment of side-chains in Lilien *et al.* (2005). Given a starting backbone conformation, we place a restraining box around each residue in the protein, which limits the displacement of that residue from its original pose: effectively, through kinematics, the restraining box also limits the range in the movement of backbone dihedrals. We now derive a novel DEE-based criterion

Table 1. Flexibility in DEE Algorithms

	Rotamer Library	Side-chain Dihedral Minimization	Backbone Flexibility	Optimal Solution	References
<i>traditional-DEE</i>	•			fixed-backbone rigid-rotamer GMEC	(Desmet <i>et al.</i> , 1992)
<i>MinDEE</i>	•	•		fixed-backbone minGMEC	(Georgiev <i>et al.</i> , 2006b,a)
<i>BD</i>	•		•	flexible-backbone rigid-rotamer GMEC	this work

A • sign shows that the corresponding flexibility type (a rotamer library, side-chain dihedral minimization, or backbone flexibility) is incorporated into the respective DEE algorithm (traditional-DEE, MinDEE, or BD).

that uses upper and lower *bounds* on rotameric interaction energies, within the specified ranges of backbone angles, to prune rotamers that are *provably* not part of the flexible-backbone rigid-rotamer GMEC.

2.1 Flexible-backbone DEE (BD)

Let the total energy of a conformation be defined as in Equation (1). Now, let $E_T(i_r|B_c)$ be the total energy of a conformation that has the rotamer identity r at residue position i , for a given backbone B_c ; let $E_T(g_0)$ be the total energy of the flexible-backbone rigid-rotamer GMEC g_0 . Also, let $E_r(B_c)$ be the template energy for backbone B_c ; let $E(i_r|B_c)$ be the self (intra-residue and residue-to-template) energy for rotamer i_r when backbone B_c is assumed, and let $E(i_r, j_s|B_c)$ be the pairwise interaction energy between rotamers i_r and j_s , again for backbone B_c . We will use a subscript g for rotamers belonging to the flexible-backbone rigid-rotamer GMEC (i_g, j_g , etc.), while B_g will be the backbone that gives the flexible-backbone rigid-rotamer GMEC.

Now, consider two conformations: one is the flexible-backbone rigid-rotamer GMEC, while the other differs from the first in two ways. First, the rotamer identity at residue i is changed from g to t (all other rotamers are the same). Second, the backbone conformation is changed from B_g to B_c . We then have:

$$E_T(i_t|B_c) \geq E_T(g_0). \quad (3)$$

First, let us define the following notation: let $\mathcal{V}(i_r) \subset \mathbb{R}^3$ be a restraining box around rotamer i_r , such that all atoms of i_r are confined to $\mathcal{V}(i_r)$. Let $C(i_r)$ be the set of conformations x of rotamer i_r for which $x(i_r) \subset \mathcal{V}(i_r)$. Similarly, we define the restraining volume $\mathcal{V}(t)$ and the set of backbone conformations $C(t)$ for the protein template t . $C(y)$ is simply the configuration space region, such that $x(y) \subset \mathcal{V}(y)$, for $y \in \{i_r, t\}$. Then,

$$E_{l_{\ominus}} = \min_{B_a \in C(t)} E_r(B_a); E_{u_{\oplus}} = \max_{B_a \in C(t)} E_r(B_a); \\ E_{l_{\oplus}} = E_{l_{\ominus}} - E_{l_{\ominus}},$$

Thus, $E_{l_{\oplus}}$ represents a lower bound on the template energy within the given backbone movement restraints. Similarly, $E_{l_{\ominus}}$ is an upper bound on the template energy and $E_{l_{\oplus}}$ is the interval of possible template energies.

We further define

$$E_{\ominus}(i_r) = \min_{z \in C(i_r), B_a \in C(t)} E(z|B_a); \\ E_{\ominus}(i_r, j_s) = \min_{z_1 \in C(i_r), z_2 \in C(j_s)} E(z_1, z_2).$$

Here, $E_{\ominus}(i_r)$ represents a lower bound on the sum of: (1) the energy interactions between the atoms of rotamer i_r , and (2) the energy interactions between the atoms of rotamer i_r and the template atoms. Similarly, $E_{\ominus}(i_r, j_s)$ is a lower bound on the pairwise energy between rotamers i_r and j_s within the restraining boxes around those two rotamers. The max terms $E_{\oplus}(i_r)$ and $E_{\oplus}(i_r, j_s)$ are defined analogously. We then define the following interval terms:

$$E_{\odot}(i_r) = E_{\oplus}(i_r) - E_{\ominus}(i_r); \\ E_{\odot}(i_r, j_s) = E_{\oplus}(i_r, j_s) - E_{\ominus}(i_r, j_s).$$

Now, using Equation (1) and taking the terms involving residue i out of the summations, we obtain:

$$E_T(i_t|B_c) = E_r(B_c) + E(i_t|B_c) + \sum_j E(i_t, j_g|B_c) \\ + \sum_j E(j_g|B_c) + \sum_j \sum_k E(j_g, k_g|B_c); \quad (4)$$

$$E_T(g_0) = E_r(B_g) + E(i_g|B_g) + \sum_j E(i_g, j_g|B_g) \\ + \sum_j E(j_g|B_g) + \sum_j \sum_k E(j_g, k_g|B_g), \quad (5)$$

where $j, k \neq i, k > j$.

Substituting Equations (4) and (5) into Equation (3), we have:

$$E_r(B_c) + E(i_t|B_c) + \sum_j E(i_t, j_g|B_c) \\ + \sum_j E(j_g|B_c) + \sum_j \sum_k E(j_g, k_g|B_c) \geq \\ E_r(B_g) + E(i_g|B_g) + \sum_j E(i_g, j_g|B_g) \\ + \sum_j E(j_g|B_g) + \sum_j \sum_k E(j_g, k_g|B_g), \quad (6)$$

where $j, k \neq i, k > j$.

Using the definitions of the E_{\ominus} and E_{\oplus} terms, we transform Equation (6) into:

$$E_{l_{\oplus}} + E_{\oplus}(i_t) + \sum_j \max_s E_{\oplus}(i_t, j_s) \\ + \sum_j E_{\oplus}(j_g) + \sum_j \sum_k E_{\oplus}(j_g, k_g) \\ \geq E_{l_{\oplus}} + E_{\ominus}(i_g) + \sum_j \min_s E_{\ominus}(i_g, j_s) \\ + \sum_j E_{\ominus}(j_g) + \sum_j \sum_k E_{\ominus}(j_g, k_g), \quad (7)$$

where $j, k \neq i, k > j$. Here, the range for the terms max and min is over the set of rotamers R_j for a given residue j .

Using the E_{\odot} definitions from above into Equation (7), we obtain:

$$E_{\oplus}(i_t) + \sum_j \max_s E_{\oplus}(i_t, j_s) \\ + E_{l_{\oplus}} + \sum_j E_{\odot}(j_g) + \sum_j \sum_k E_{\odot}(j_g, k_g) \\ \geq E_{\ominus}(i_g) + \sum_j \min_s E_{\ominus}(i_g, j_s), \quad (8)$$

where $j, k \neq i, k > j$.

We then define the *BD criterion* for a given rotamer i_r to be:

$$E_{\ominus}(i_r) + \sum_j \min_s E_{\ominus}(i_r, j_s) \\ - E_{l_{\oplus}} - \sum_j \max_s E_{\odot}(j_s) - \sum_j \sum_k \max_{s,u} E_{\odot}(j_s, k_u) \\ > E_{\oplus}(i_t) + \sum_j \max_s E_{\oplus}(i_t, j_s), \quad (9)$$

where $j, k \neq i, k > j$ and max is over the sets R_j and R_k for given residues j and k .

PROPOSITION 1. *When Equation (9) holds, rotamer i_r cannot be a part of the flexible-backbone rigid-rotamer GMEC and can thus be pruned from consideration for residue i .*

PROOF. We substitute the left-hand side of Equation (9) for the first two terms in the left-hand side of Equation (8) to obtain:

$$\begin{aligned} E_{\ominus}(i_r) + \sum_j \min_s E_{\ominus}(i_r, j_s) \\ - E_{r_{\ominus}} - \sum_j \max_s E_{\ominus}(j_s) - \sum_j \sum_k \max_{s,u} E_{\ominus}(j_s, k_u) \\ + E_{r_{\ominus}} + \sum_j E_{\ominus}(j_g) + \sum_j \sum_k E_{\ominus}(j_g, k_g) \\ > E_{\ominus}(i_g) + \sum_j \min_s E_{\ominus}(i_g, j_s), \end{aligned} \quad (10)$$

where $j, k \neq i, k > j$.

Since

$$E_{\ominus}(j_g) \leq \max_s E_{\ominus}(j_s); E_{\ominus}(j_g, k_g) \leq \max_{s,u} E_{\ominus}(j_s, k_u),$$

then Equation (10) becomes:

$$\begin{aligned} E_{\ominus}(i_r) + \sum_j \min_s E_{\ominus}(i_r, j_s) \\ - E_{r_{\ominus}} - \sum_j \max_s E_{\ominus}(j_s) - \sum_j \sum_k \max_{s,u} E_{\ominus}(j_s, k_u) \\ + E_{r_{\ominus}} + \sum_j \max_s E_{\ominus}(j_s) + \sum_j \sum_k \max_{s,u} E_{\ominus}(j_s, k_u) \\ > E_{\ominus}(i_g) + \sum_j \min_s E_{\ominus}(i_g, j_s), \end{aligned}$$

where $j, k \neq i, k > j$.

Simplifying, we obtain:

$$E_{\ominus}(i_r) + \sum_j \min_s E_{\ominus}(i_r, j_s) > E_{\ominus}(i_g) + \sum_j \min_s E_{\ominus}(i_g, j_s),$$

where $j \neq i$.

Thus, when Equation (9) holds, then $i_r \neq i_g$, so rotamer i_r can be provably pruned from further consideration when Equation (9) holds, since it cannot belong to the flexible-backbone rigid-rotamer GMEC. \square

For a given compact space of backbone conformations, Equation (9) compares a lower bound on the energy achievable when residue i has the specific rotamer identity r against an upper bound on the energy achievable with a competing rotamer t for the same residue. Similarly to the fixed-backbone DEE algorithms, Equation (9) is evaluated for each residue,

for each i_r and against different competitors, in order to identify as many dead-ending rotamers as possible. In contrast to both the fixed-backbone traditional-DEE and MinDEE criteria, BD takes into account possible changes in the energy interactions due to backbone conformational changes. These changes are represented by the E_{\ominus} terms in Equation (9). The E_{\ominus} terms in Equation (9) are a function only of the residue numbers and can thus be precomputed, so computing Equation (9) has the same complexity as Equation (2).

2.2 Extensions to BD

Analogously to the extensions for the fixed-backbone traditional-DEE and MinDEE, we derived (Fig. 1) four extensions to the initial BD criterion (Equation 9) for improved pruning efficiency; like BD, all of these extensions are provably-accurate with respect to the flexible-backbone rigid-rotamer GMEC.

3 ALGORITHM

The protein redesign algorithm for the identification of the flexible-backbone rigid-rotamer GMEC consists of two stages: pruning and enumeration. In the pruning stage, the BD analogs of the simple Goldstein (Fig. 1b), conformational splitting (Fig. 1e), and Goldstein pairs (Fig. 1d) criteria, as well as the MinBounds criterion and the DACS algorithm (Georgiev *et al.*, 2006a) are used to provably prune the majority of rotamer choices, thus reducing significantly the set of candidate conformations. The remaining conformations are then extracted in order of increasing *lower bounds* (Section 4) on their energies using the A^* algorithm (Georgiev *et al.*, 2006b; Leach and Lemon, 1998). The use of A^* eliminates the necessity to enumerate all remaining conformations. Since A^* generates conformations in order of increasing lower bounds on their energy, we are guaranteed to have obtained the flexible-backbone rigid-rotamer GMEC once the lower bound on the next conformation generated by A^* is greater than the minimum conformational energy computed in the search so far [see Proposition 2 in Georgiev *et al.* (2006b)].

For a given structural model, energy function, and rotamer library, the BD criterion and extensions (Fig. 1) guarantee that

(a) $E_{\ominus}(i_r) - E_{\oplus}(i_t) + \sum_{j \neq i} \min_s E_{\ominus}(i_r, j_s) - \sum_{j \neq i} \max_s E_{\oplus}(i_t, j_s) - E_{\Delta} > 0$
(b) $E_{\ominus}(i_r) - E_{\oplus}(i_t) + \sum_{j \neq i} \min_s (E_{\ominus}(i_r, j_s) - E_{\oplus}(i_t, j_s)) - E_{\Delta} > 0$
(c) $E_{\ominus}(i_r) - \sum_{x=1,T} C_x E_{\oplus}(i_{t_x}) + \sum_{j \neq i} \min_s \left(E_{\ominus}(i_r, j_s) - \sum_{x=1,T} C_x E_{\oplus}(i_{t_x}, j_s) \right) - E_{\Delta} > 0$
(d) $E_{\ominus}([i_r j_s]) - E_{\oplus}([i_u j_v]) + \sum_{h \neq i, j} \min_t (E_{\ominus}([i_r j_s], h_t) - E_{\oplus}([i_u j_v], h_t)) - E_{\Omega} > 0$
(e) $E_{\ominus}(i_r) - E_{\oplus}(i_t) + \sum_{j \neq h \neq i} \left(\min_s (E_{\ominus}(i_r, j_s) - E_{\oplus}(i_t, j_s)) \right) + (E_{\ominus}(i_r, h_v) - E_{\oplus}(i_t, h_v)) - E_{\Delta} > 0$

Fig. 1. Extensions to BD. (a) The initial BD criterion (Equation 9). (b) The fixed-backbone traditional-DEE Goldstein analog (Goldstein 1994) for BD. (c), (d), and (e) are the generalizations to BD of the (c) general and (d) dead-ending pairs Goldstein criteria (Goldstein, 1994), as well as (e) conformational splitting (Pierce *et al.*, 2000). Here, $E_{\Delta} = (E_{r_{\ominus}} + \sum_j \max_s E_{\ominus}(j_s) + \sum_j \sum_k \max_{s,u} E_{\ominus}(j_s, k_u))$, for $j, k \neq i; k > j$. Also, $E_{\Omega} = (E_{r_{\ominus}} + \sum_h \max_t E_{\ominus}(h_t) + \sum_h \sum_k \max_{t,w} E_{\ominus}(h_t, k_w))$, where $h, k \neq i, j; k > h$. Finally, $E_{\ominus}([i_r j_s]) = E_{\ominus}(i_r) + E_{\ominus}(j_s) + E_{\ominus}(i_r, j_s) (i \neq j)$, $E_{\ominus}([i_r j_s], h_t) = E_{\ominus}(i_r, h_t) + E_{\ominus}(j_s, h_t) (i, j \neq h)$, where $E_{\ominus} \in \{E_{\ominus}, E_{\oplus}\}$.

pruned rotamers are provably not part of the flexible-backbone rigid-rotamer GMEC. These criteria are provably correct assuming the lower (E_{\ominus}) and upper (E_{\oplus}) energy bounds are computed correctly. A lower bound $E_{\ominus}(i_r, j_s)$ for two rotamers i_r and j_s is *sound* if $E_{\ominus}(i_r, j_s) \leq \min_{c \in Y} E(i_r, j_s | c)$, where Y is the set of conformations for which: (1) $c \in Y$: residue positions i and j assume the particular rotamer identities i_r and j_s , respectively, and (2) all conformations in Y are restrained by the dihedral-angle bounds and $\mathcal{V}(\cdot)$ real-space bounds. That is, $E_{\ominus}(i_r, j_s)$ is sound if its value is lower than the minimum pairwise energy between rotamers i_r and j_s observed in conformations in Y . The soundness of the other E_{\ominus} terms is defined analogously. Several global optimization techniques are available to compute these bounds (Carr and Wales, 2005; Wales and Scheraga, 1999). For more efficient lower-bound computation, we use an approximation to steepest-descent minimization (Sec. 4). Empirically, we observe not only that the computed lower bounds are sound, but also that they are overly conservative, i.e. $E_{\ominus}(i_r, j_s) \ll \min_{c \in Y} E(i_r, j_s | c)$. This observation can be explained by the fact that the pairwise lower bounds are computed in the absence of some other side-chains (see Sec. 4). The soundness of the upper bounds E_{\oplus} is defined analogously to the lower-bounds case. In particular, since soft steric overlap between atoms is allowed (Sec. 4), it is unlikely that the energy between a pair of rotamers will increase from their initial energy upon conformation minimization. We thus compute $E_{\oplus}(i_r, j_s)$ as the initial energy between i_r and j_s ; the other E_{\oplus} terms are computed analogously (Section 4).

4 METHODS

A subset of the GrsA-PheA residues, consisting of 9 active site residues, a steric shell, the substrate ligand, and the AMP cofactor, (PDB id: lamu; Conti *et al.*, 1997) was used in our experiments. In general, the choice of flexible residues is a user-specified parameter. The choices that we made are designed to capture the important motions of the enzyme active site. The active site residues are (D235, A236, W239, T278, I299, A301, A322, I330, C331); these positions are allowed to either mutate to the set GAVLIFYWM of hydrophobic amino acids or to keep their wildtype amino acid identity. The side-chains for the active site residues are modeled using the Richardsons' rotamer library (Lovell *et al.*, 2000). The backbone (ϕ, ψ) dihedrals for these residues are also allowed to change during conformation energy minimization. Backbone conformational changes are restrained by allowing a maximum displacement of 1.5 Å for each C_{α} atom from its initial position in the PDB file and a maximum change of $\pm 3^{\circ}$ from the initial values for the flexible (ϕ, ψ) angles. The steric shell consists of 32 residues: the 30 residues with at least one atom within 8 Å of the ligand (Lilien *et al.*, 2005) and residues 277 and 298 (which are included so that the (ϕ, ψ) dihedrals for active site residues 278 and 299 can be properly defined). All residues in the steric shell have fixed side-chains and can only move from their initial position due to changes in the (ϕ, ψ) angles of the active site residues. The ligand is modeled using rotamers and is allowed to rotate and translate.

Similarly to Shah *et al.* (2004) and Georgiev *et al.* (2006a), the 12 core residues (3, 5, 7, 9, 20, 26, 30, 34, 39, 41, 52, 54) in G β 1 (PDB id: 1pga) (Gallagher *et al.*, 1994) are modeled as flexible using the same rotamer library and backbone flexibility procedure as for GrsA. The set AVLIFYW of hydrophobic amino acids, as well as the wildtype amino acid identity, is allowed at these core positions. The remainder of the G β 1 residues are modeled as part of the steric shell.

A lower bound on a conformational energy is computed as a sum of lower bounds on pairwise energy interactions (Georgiev *et al.*, 2006a). For a pair of rotamers i_r and j_s , a lower bound $E_{\ominus}(i_r, j_s)$ on the pairwise energy (Equation 9) is computed by performing a finite-differencing approximation to steepest-descent minimization of the (ϕ, ψ) backbone dihedrals subject to the $\mathcal{V}(\cdot)$ bounds in \mathbb{R}^3 (Section 2), and returns the minimum pairwise energy between i_r and j_s . So that i_r and j_s will have less steric constraint during minimization for the pairwise energy lower bound computation, all other flexible residues, except for Pro and Gly, are set to Ala. The upper bound $E_{\oplus}(i_r, j_s)$ is computed as the pairwise energy between i_r and j_s for the initial backbone conformation. This bound is much more conservative and is much faster to compute than the upper-bound algorithm in Georgiev *et al.*, (2006c). The energies involving the template are computed in a similar manner. The same restraints on backbone movement (as described above) are used. The computation of the side-chain dihedral lower and upper bounds is described in Georgiev *et al.*, (2006c).

A change in the ϕ angle for residue i rotates the whole structure between the N-terminus and residue i accordingly. Similarly, a change in the ψ angle for residue i rotates the structure between residue i and the C-terminus. Conformations are energy-minimized using a finite-differencing approximation to steepest-descent minimization. The energy function consists of the AMBER electrostatic and vdW terms (Cornell *et al.*, 1995; Weiner *et al.*, 1984) and the EEF1 pairwise solvation energy term (Lazaridis and Karplus, 1999). A dielectric of 20 and a solvation-energy scaling factor of 0.05 were used to make the computed energies predominantly dependent on vdW terms. For the redesign with MinDEE, rotamer dihedrals are assumed to be near the bottom of an energy well; dihedral energies are computed using the AMBER dihedral terms. Rotamers with a lower bound on the self-energy (intra-residue and residue-to-template energies) greater than 20 kcal/mol are pruned due to incompatibility with the template (De Maeyer *et al.*, 1997). Conformations for which at least one pair of atoms has a steric overlap of more than 1.5 Å before minimization are pruned from consideration. All experiments were performed on a 18-processor cluster.

5 RESULTS AND DISCUSSION

In this section we report the results from the application of our BD algorithm (Section 3) to redesign the core of G β 1 and to switch the substrate specificity of GrsA towards a novel substrate, Leucine. We further compare the BD results to traditional-DEE and MinDEE.

5.1 Backbone flexibility with traditional-DEE

When backbone flexibility is incorporated into the design model, the traditional-DEE criteria can be used instead of the BD criteria derived in Section 2, although in such a case the results are not provably-accurate. We now confirm via computational experiments that when traditional-DEE is used with backbone flexibility, the identification of the lowest-energy conformation, the flexible-backbone rigid-rotamer GMEC, can no longer be guaranteed. For the G β 1 redesign, we applied the BD algorithm (Section 3) to generate the flexible-backbone rigid-rotamer GMEC. When traditional-DEE was used instead, the rotamer belonging to the flexible-backbone rigid-rotamer GMEC at residue position 7, Leu rotamer 5 from the Richardsons' rotamer library (with χ angles -65° and 175°), was pruned, in favor of an Ile rotamer for the same residue. Thus, the flexible-backbone rigid-rotamer

Table 2. Minimum energy conformations with different types of minimization

	Optimum Computed	Mutations from wildtype	No minimization ^a	Side-chain dihedral minimization ^b	Backbone flexibility ^{c,d}
i.	fixed-backbone rigid-rotamer GMEC	7I/30L/39L/52W	-204.98*	-205.63	-214.58 (0.11)
	fixed-backbone minGMEC	7I/39L/52W/54L	-198.63	-208.88*	-209.35 (0.12)
	flexible-backbone rigid-rotamer GMEC	30L/39L/52W [†]	-202.33	-205.97	-215.47* (0.15)
ii.	fixed-backbone rigid-rotamer GMEC	A236M/A301G	-148.03*	-154.59	-155.98 (0.14)
	fixed-backbone minGMEC	A236M/A322M	-139.74	-160.28*	-160.26 (0.16)
	flexible-backbone rigid-rotamer GMEC	A236M/A322M	-139.74	-160.28	-160.26* (0.16)

After computing the rotamer sequences for the fixed-backbone rigid-rotamer GMEC, fixed-backbone minGMEC, and flexible-backbone rigid-rotamer GMEC for the redesigns of (i) G β 1 and (ii) GrsA-PheA, the energy for each of these rotamer sequences was computed for three cases: ^awith no minimization (the energy of the initial structure), ^bafter side-chain dihedral minimization, or ^cafter minimization with backbone flexibility. The computed energies are in kcal/mol. The energies of the fixed-backbone rigid-rotamer GMEC, fixed-backbone minGMEC, and flexible-backbone rigid-rotamer GMEC, each of which is guaranteed to be the respective minimum energy by the traditional-DEE, MinDEE, and BD algorithms, are starred (*). ^dThe backbone RMSD (in Å) between the ^ainitial structure and the ^cbackbone-minimized structure is shown in parentheses.

GMEC 30L/39L/52W (Table 2[†]) is erroneously *eliminated* by traditional-DEE during the pruning stage, confirming that traditional-DEE is not provably accurate with backbone flexibility.

5.2 Results with BD

In contrast to traditional-DEE, BD has provable guarantees with respect to the flexible-backbone rigid-rotamer GMEC. The application of the BD-based pruning algorithm (Section 3) efficiently reduced the number of conformations that had to be subsequently considered in the A^* enumeration stage by a factor of more than 10^7 (from the initial 8.39×10^{17}) for the G β 1 redesign and 10^5 (from the initial 4.78×10^{15}) for GrsA. The generation of the G β 1 flexible-backbone rigid-rotamer GMEC 30L/39L/52W took approximately one week. The generation of the two-point mutation flexible-backbone rigid-rotamer GMEC A236M/A322M for GrsA required more than 1.5 days (in a k -point mutation sequence, k residues are allowed to mutate simultaneously; A^* is easily modified to generate only up to k -point mutations). In both cases, the DEE pruning stage took around a minute and the remainder of the time was spent in the A^* enumeration stage. In comparison, the running time of traditional-DEE for the identification of the fixed-backbone rigid-rotamer GMEC was less than 1 min for both G β 1 and GrsA, while the generation of the fixed-backbone minGMEC by MinDEE took approximately one day for G β 1 and 5 h. for GrsA. Thus, although BD is capable of pruning a significant fraction of the possible conformations, its provable guarantees come at the expense of reduced pruning efficiency and considerably increased running time. In particular, there are three factors that influence the speed of the BD algorithm. (1) The inclusion of the E_{∞} terms in Equation (9) reduces the pruning efficiency of BD compared to traditional-DEE, since fewer rotamers can be provably pruned. This significantly increases the size of the input for the A^* conformation enumeration stage. (2) Whereas with traditional-DEE the first conformation generated by A^* is guaranteed to be the fixed-backbone rigid-rotamer GMEC, BD requires that a *set* of conformations be generated before the search can be provably

halted (Section 3). Since the provable halting condition described in Section 3 depends on the computed lower bounds on the conformational energies, so does the size of the set of generated conformations. Currently, however, the computed lower energy bounds are too conservative (i.e. the gap between the computed lower bound on a conformational energy and the actual energy of the conformation can be considerable; data not shown). Consequently, the size of the set of generated conformations increases significantly, resulting in an increased running time of the algorithm. This problem, although to a slightly lesser extent, is also present in MinDEE (Georgiev *et al.*, 2006b). Hence, an approach that uses more constraint to compute the lower energy bounds would prove beneficial both for BD and MinDEE designs. (3) The current implementation of the backbone minimizer is much slower than side-chain dihedral minimization. Although this is expected, a faster backbone minimization algorithm that does not sacrifice accuracy can significantly speed up the computation of the designs.

The significance of BD lies in its ability to generate lower-energy conformations than traditional-DEE. As Table 2 shows for the GrsA redesign, even after performing backbone minimization, the energy of the fixed-backbone rigid-rotamer GMEC is significantly higher (by more than 4 kcal/mol) than the energy of the flexible-backbone rigid-rotamer GMEC. Thus, for models that incorporate backbone flexibility, the BD criterion should be used instead of traditional-DEE when accuracy is preferred over speed.

Interestingly, for the GrsA redesign, the flexible-backbone rigid-rotamer GMEC identified by BD and the fixed-backbone minGMEC identified by MinDEE are obtained from the same sequence and initial rotamer identities (Table 2). From that initial conformation, flexible-backbone minimization and side-chain dihedral minimization resulted in virtually equal energies. In the G β 1 redesign, differences of almost 10 kcal/mol were observed between the energies resulting from side-chain and backbone minimization (Table 2). Experiments on a set of surface residues of G β 1 [a description of this system can be found in (Georgiev *et al.*, 2006a)] showed that in some cases side-chain minimization can achieve lower

energy levels than backbone minimization, while in others backbone minimization performs better (data not shown). Hence, in order to further improve the accuracy of the model, both minimization types could be incorporated simultaneously.

A somewhat surprising result (Table 2) is that even very small changes in the backbone can result in a significant improvement in the conformational energy. Since the ligand in the GrsA system is also allowed to rotate, translate, and flex, however, it was not certain to what extent the energy improvement was due to protein backbone movements. In order to determine whether small backbone perturbations can lead to considerably lower energies, we generated and energy-minimized (using flexible-backbone minimization) 1,000 conformations of our GrsA structural model without a ligand. The relationship between the magnitude of the backbone movement and the corresponding change in energy for the best 30 conformations (with lowest initial energy) is shown in Figure 2. This figure confirms that conformational energies indeed can be very sensitive to backbone movements. Since the non-minimized structures may contain some unoptimized interactions (e.g. some steric overlap between atoms), large energy decreases after minimization (such as the ones seen in the top right-hand corner of Figure 2) are not unexpected.

Another important observation is that all conformations plotted in Figure 2 have backbone RMSDs from their respective initial structures less than 0.3 Å, even though C_{α} atoms are allowed to move up to 1.5 Å (Section 4). The goal of the BD algorithm is not the design of novel backbone conformations, so significant backbone deviations should not be expected. Instead, the essence of BD is to allow small modifications to the backbone in order to adopt conformations that would otherwise score low (or even be erroneously pruned, like the G β 1 flexible-backbone rigid-rotamer GMEC 30L/39L/52W) by a fixed-backbone model.

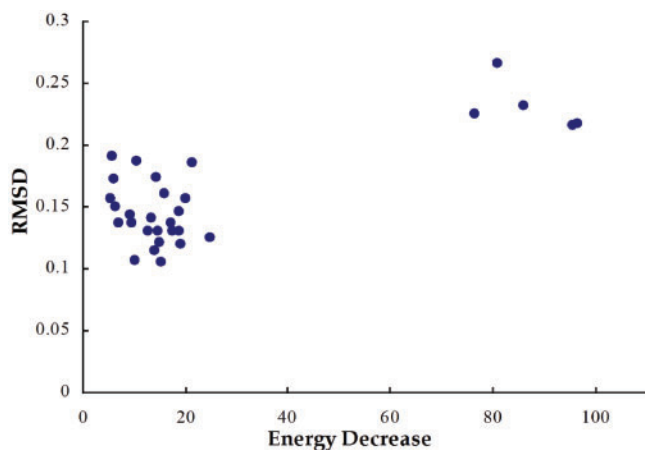


Fig. 2 . Backbone RMSD versus decrease in energy. The decrease in energy (in kcal/mol) resulting from backbone minimization of 30 conformations is plotted against the backbone RMSD values between the corresponding non-minimized and minimized structures.

6 CONCLUSION

In this article we presented BD, a novel provably accurate DEE-based algorithm for protein design with backbone flexibility. We also gave extensions to the initial BD criterion for improved pruning efficiency. When applied in redesigns of core residues of G β 1 and the active site of the NRPS enzyme GrsA-PheA, BD significantly reduced the set of candidate conformations for obtaining the flexible-backbone rigid-rotamer GMEC. The provable guarantees of BD, however, come at the expense of decreased pruning efficiency and increased running time compared to traditional-DEE. We showed experimentally that traditional-DEE is indeed not provably-correct with backbone flexibility, generating higher energy structures than BD. We can thus conclude that when improved accuracy is required, BD must be used instead of traditional-DEE. The different GrsA GMECs computed in Section 5 are currently being tested in our wetlab.

For a given structural model, energy function, and rotamer library, provably accurate algorithms such as traditional-DEE (for a fixed-backbone) and BD (for a flexible-backbone model), can guarantee the identification of the optimal solution for that model. In contrast, heuristic techniques such as Monte Carlo and genetic algorithms do not have such guarantees. It has been shown that when compared to experimental data, some heuristic predictions may have comparable quality to the predictions from a provably accurate algorithm (Desmet *et al.*, 2002). Heuristic approaches, however, cannot decouple the inaccuracy of the model from the inaccuracy of the algorithm. With a provably-accurate algorithm, the discrepancy between predictions and experimental data can be exclusively attributed to deficiencies in the model. Experimental feedback for improving the model can thus be more reliably incorporated with provably accurate algorithms.

The goal of the backbone flexibility model discussed in this article is not to identify novel backbone conformations that are significantly different from the initial protein backbone. Rather, our model allows small movements of the backbone in order to adapt for sequences and conformations that would otherwise be discarded by a fixed-backbone model. By increasing the maximum backbone movement allowed (Section 4), larger deviations from the initial backbone could be obtained. Increasing the bounds on the backbone movement, however, will also increase the magnitude of the lower energy bounds and the interval terms in Equation (9), which presents an added challenge for pruning efficiency. Improving the approach for computing the lower/upper energy bounds will thus be essential for enhancing the pruning and computational efficiency of the algorithm.

Similarly to MinDEE, BD can be used as a pruning filter in K^* , the ensemble-based protein design algorithm of (Lilien *et al.*, 2005), with incorporated backbone flexibility. This obtains a provably-good approximation algorithm for computing partition functions over ensembles simultaneously containing side-chain, backbone, and ligand flexibility. Moreover, since the backbone is allowed to flex, a negative design procedure can also be incorporated into the model, so that the design goal can not only improve a desired function but also impede certain functionality (e.g. to redesign GrsA so that

the specificity towards Leu is improved, but the specificity for other substrates is significantly reduced).

Since both side-chain dihedral minimization and backbone flexibility improve the accuracy of the model, interesting future work could simultaneously use both minimization mechanisms, while remaining provably accurate with respect to the corresponding optimal solution, the *flexible-backbone minGMEC*. The theoretical framework for BD will still be valid for such an algorithm, since the E_{\odot} terms in Equation (9) account for possible energy changes during minimization. However, the restraining boxes for the computation of the lower and upper energy bounds must be modified to incorporate possible rotamer movements when both minimization types are allowed. The major challenge for such an algorithm would thus be the increased computational requirements. However, the benefits of such a marriage could be substantial.

ACKNOWLEDGEMENTS

We thank Mr. J. MacMaster and Mr. A. Yan for help with different parts of this work, and all members of the Donald Lab for helpful discussions and comments. This work is supported by a grant to B.R.D. from the National Institutes of Health (R01 GM-65982).

Conflict of Interest: none declared.

REFERENCES

- Carr, J.M. and Wales, D.J. (2005) Global optimization and folding pathways of selected α -helical proteins. *J. Chem. Phys.*, **123**, 1–12.
- Chazelle, B. et al. (2004) A semidefinite programming approach to side-chain positioning with new rounding strategies. *INFORMS J. Comput. Comput. Biol. Special Issue*, **16**, 380–392.
- Conti, E. et al. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of Gramicidin S. *EMBO J.*, **16**, 4174–4183.
- Cornell, W. et al. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.
- Dahiyat, B. and Mayo, S. (1997a) De novo protein design: fully automated sequence selection. *Science*, **278**, 82–87.
- Dahiyat, B.I. and Mayo, S.L. (1997b) Probing the role of packing specificity in protein design. *PNAS*, **94**, 10172–10177.
- De Maeyer, M. et al. (1997) All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination. *Fold. Design*, **2**, 53–66.
- Desjarlais, J.R. and Handel, T.M. (1999) Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.*, **289**, 305–318.
- Desmet, J. et al. (1992) The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, **356**, 539–542.
- Desmet, J. et al. (2002) Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins*, **48**, 31–43.
- Dunbrack, R.L. (2002) Rotamer libraries in the 21st century. *Cur. Op. Struct. Biol.*, **12**, 431–440.
- Fung, H.K. et al. (2007) Novel formulations for the sequence selection problem in de novo protein design with flexible templates. *Optim. Methods Software*, **22**, 51–71.
- Gallagher, T. et al. (1994) Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry*, **33**, 4721–4729.
- Georgiev, I. et al. (2006a) Improved pruning algorithms and divide-and-conquer strategies for dead-end elimination, with application to protein design. *Bioinformatics*, **22**, e174–183. Proceedings of International Conference on Intelligent Systems for Molecular Biology (ISMB), Fortaleza, Brazil (2006).
- Georgiev, I. et al. (2006b) A novel minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. In *Proceedings of The Tenth Annual International Conference on Research in Computational Molecular Biology (RECOMB)*. Springer Berlin, Venice, Italy, pp. 530–545, RECOMB 2006, Lecture Notes in Computer Science, LNBI 3909.
- Georgiev, I. et al. (2006c) A novel minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. Technical Report 570, Dartmouth Computer Science Dept., <http://www.cs.dartmouth.edu/reports/abstracts/TR2006-570>.
- Goldstein, R. (1994) Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J.*, **66**, 1335–1340.
- Gordon, D. et al. (2003) Exact rotamer optimization for protein design. *J. Comput. Chem.*, **24**, 232–243.
- Gordon, D.B. et al. (1999) Energy functions for protein design. *Cur. Op. Struct. Bio.*, **9**, 509–513.
- Harbury, P.B. et al. (1998) High-resolution protein design with backbone freedom. *Science*, **282**, 1462–1467.
- Harbury, P.B. et al. (1995) Repacking protein cores with backbone freedom: Structure prediction for coiled coils. *PNAS*, **92**, 8408–8412.
- Hellings, H. and Richards, F. (1991) Construction of new ligand binding sites in proteins of known structure: I. computer-aided modeling of sites with predefined geometry. *J. Mol. Biol.*, **222**, 763–785.
- Jaramillo, A. et al. (2001) Automatic procedures for protein design. *Comb. Chem. High Throughput Screen*, **4**, 643–659.
- Jin, W. et al. (2003) De novo design of foldable proteins with smooth folding funnel: Automated negative design and experimental verification. *Structure*, **11**, 581–591.
- Kuhlman, B. and Baker, D. (2000) Native protein sequences are close to optimal for their structures. *PNAS*, **97**, 10383–10388.
- Kuhlman, B. et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
- Lasters, I. and Desmet, J. (1993) The fuzzy-end elimination theorem: correctly implementing the side chain placement algorithm based on the dead-end elimination theorem. *Protein Eng.*, **6**, 717–722.
- Lazaridis, T. and Karplus, M. (1999) Effective energy function for proteins in solution. *PROTEINS: Structure, Function, and Genetics*, **35**, 133–152.
- Leach, A. and Lemon, A. (1998) Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. *Proteins*, **33**, 227–239.
- Lilien, R. et al. (2005) A novel ensemble-based scoring and search algorithm for protein redesign, and its application to modify the substrate specificity of the Gramicidin Synthetase A phenylalanine adenylation enzyme. *J. Comput. Biol.*, **12**, 740–761.
- Lim, W.A. et al. (1994) The crystal structure of a mutant protein with altered but improved hydrophobic core packing. *PNAS*, **91**, 423–427.
- Looger, L. et al. (2003) Computational design of receptor and sensor proteins with novel functions. *Nature*, **423**, 185–190.
- Looger, L. and Hellings, H. (2001) Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: Implications for protein design and structural genomics. *J. Mol. Biol.*, **307**, 429–445.
- Lovell, S. et al. (2000) The penultimate rotamer library. *Proteins*, **40**, 389–408.
- Najmanovich, R. et al. (2000) Side-chain flexibility in proteins upon ligand binding. *Proteins*, **39**, 261–268.
- Pierce, N. et al. (2000) Conformational splitting: a more powerful criterion for dead-end elimination. *J. Comput. Chem.*, **21**, 999–1009.
- Pierce, N. and Winfree, E. (2002) Protein design is NP-hard. *Protein Eng.*, **15**, 779–782.
- Ponder, J. and Richards, F. (1987) Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.*, **193**, 775–791.
- Shah, P. et al. (2004) Preprocessing of rotamers for protein design calculations. *J. Comput. Chem.*, **25**, 1797–1800.
- Stevens, B.W. et al. (2006) Redesigning the PheA domain of Gramicidin Synthetase leads to a new understanding of the enzyme's mechanism and selectivity. *Biochemistry*, **45**, 15495–15504.

- Street,A. and Mayo,S. (1999) Computational protein design. *Structure*, **7**, R105–R109.
- Su,A. and Mayo,S.L. (1997) Coupling backbone flexibility and amino acid sequence selection in protein design. *Prot. Sci.*, **6**, 1701–1707.
- Voigt,C. *et al.* (2000) Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.*, **299**, 789–803.
- Wales,D.J. and Scheraga,H.A. (1999) Global optimization of clusters, crystals, and biomolecules. *Science*, **285**, 1368–1372.
- Weiner,S. *et al.* (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765–784.
- Zanghellini,A. *et al.* (2006) New algorithms and an in silico benchmark for computational enzyme design. *Prot. Sci.*, **15**, 2785–2794.