Pergamon

Bioorganic & Medicinal Chemistry, Vol. 4, No. 5, pp 631–638, 1996

0968-0896/96 $15.00 + 0.00

S0968-0896(96)00056-9

Computational Screening of Combinatorial Libraries

Qiang Zheng* and Donald J. Kyle
Scios Nova, Inc., 820 West Maude Avenue, Sunnyvale, CA 94086, U.S.A.

Abstract—We first review existing computational methods with an intrinsic combinatorial feature, then describe a new computational method for screening combinatorial libraries using a recently developed multicopy sampling technique. The new method differs from the existing ones in that it can be used to screen simultaneously an entire library of molecules, instead of the individual molecules in a library. As an example, we have applied the method to study site-directed amino acid substitutions in a protein. After two rounds of library screening, we identified the energetically most stable substitutions along with their optimal conformations from all natural amino acids. In principle, the method is generally applicable to study ligand–host systems.

Introduction

Combinatorial syntheses and high throughput screening allow many compounds to be synthesized and screened in a time comparable to that normally required for the synthesis and screening of a single compound.[1,2] In contrast, computer-aided structural-based molecular design is, for the most part, limited to the modeling of a single compound at a time. This technological gap represents a significant challenge to molecular modeling at a time when its accuracy and reliability is already being scrutinized. Here, we present a brief analysis of the existing modeling methods that are relevant to combinatorial methods, then describe our new method for computational screening combinatorial libraries. This new method allows the simultaneously modeling of an entire library of compounds, in the context of a flexible binding site structure.

The combinatorial problem in computational screening is highly complex in that a given compound is always modeled indirectly via its conformations. For example, when compound X is said to have a lower binding energy to a host molecule than compound Y, the energies represent their respective optimal binding conformations. For this reason, molecular modeling always consists of two parts: conformational sampling and energetic evaluation. Since each compound can have a large number of conformations, the combinatorial number of all the conformations in a library of many compounds can be enormous. Moreover, since the host and ligands can undergo conformational adjustments upon ligand binding, the total combinatorial number of conformations of both the host and the ligands can be astronomical.

Conformational sampling

Elucidation of all the conformations of interest typically exceeds the time and computer hardware constraints, thus the success of molecular modeling relies on efficient sampling of low-energy conformations. The key to improving sampling efficiency is to find better ways to calculate conformational energy, which is a sum of many terms, including electrostatic and van der Waals interactions, hydrogen bonds, covalent bonds, bond angles, and rotatable dihedral angles. The electrostatic term accounts for most of the long range interactions between each pair of the atoms in a molecule, while the other terms represent short-range interactions between each atom and its neighbors. Since there are many more atom pairs than atoms (N^2 vs N, where N is the number of atoms in a molecule), the efficiency of a conformational sampling method is essentially determined by its efficiency of electrostatic calculation.

Consider a ligand–host complex, consisting of n and N atoms, respectively. The total number of electrostatic interactions of a given complex conformation can be grouped into three terms. (1) Self-interaction of the host [$N(N-1)/2$ pairs], (2) self-interaction of the ligand [$n(n-1)/2$ pairs], and (3) ligand–host interaction [nN pairs]. Suppose the time $\tau(1)$ required for the energy calculation for a single complex conformation is approximately proportional to the total number of electrostatic interactions:

$$\tau(1) \sim N(N-1)/2 + nN + n(n-1)/2. \quad (1)$$

The time $\tau(m)$ required for the energy calculation of m conformations is:

$$\tau(m) \sim m\tau(1) = m(N(N-1) + nN + n(n-1)/2)$$
$$\sim mN^2/2 \qquad \text{for } N \gg n > 1 \quad (2)$$

This computational time is proportional to the number of conformations, so the electrostatic calculation is *not* of a combinatorial nature. The inefficiency of this

calculation is primarily a result of the computational time spent on calculating the host's self-interaction (an 'overhead') instead of the ligand–host interaction. Moreover, the 'overhead' is carried over in the energy calculation of all conformations. Although a moderate conformational flexibility around the binding site in the host can be critical to ligand-binding, the overhead is too large to justify. The reason for having this overhead, which is inherent to many current modeling methods, is to account for the conformational flexibility of the host. One way to reduce the overhead is by limiting or eliminating altogether the conformational flexibility of the host.

Among the published methods for reducing this overhead, two are intrinsically efficient and combinatorial, namely the grid[3,4] and the multicopy sampling methods.[5,6] Both methods are based solely on the underlying physical principles of two-body interaction and, therefore, are generally applicable to the general ligand–host systems.

The grid method removes the overhead completely by eliminating the conformational flexibility of the host. Each atom in the ligand interacts with the electrostatic field of the host discretized on a three-dimensional grid, thus greatly simplifying the electrostatic calculation of both host–host and the ligand–host interactions. The electrostatic field remains the same for all ligands. For a total of P grid points, the computational time $\tau_{grid}(m)$ required for the electrostatic calculation of m ligand-host conformations is:

$$\tau_{grid}(m) \sim NP + mn + mn(n-1)/2$$

$$\sim NP + mn^2/2$$

For $N \gg n > 1$ and $NP \gg n^2/2$, the grid electrostatic calculation of $m = 2NP/n^2$ ligand conformations requires twice the time needed by the traditional method for the calculation of a single conformation. Consider a protein host with 100 amino acids and a ligand of the size of a single amino acid, and a cubic grid with a 0.5 Å spacing and a 5 Å dimension (1000 grid points), the gain of sampling efficiency is $m/2 \sim 100{,}000$. In this regard, the grid method is intrinsically combinatorial.

The trade-off for this gain of conformational sampling speed is that the host conformation must be rigid, which may result in poor sampling quality. For example, consider two similar ligand conformations A and B, where A is at an energy minimum and the energy of B is high due to a small bad contact with the host. In a grid-based sampling, conformation B will be discarded. However, if the host molecule were flexible, a few steps of energy minimization of the ligand–host complex would quickly resolve the bad contact and would likely lead to a minimum-energy conformation very similar to A.

The multicopy sampling method [also known as the locally enhanced sampling (LES)[5] and the multiple copy simultaneous sampling (MCSS)[6] methods] can be viewed as an intermediate approach between the traditional and the grid sampling methods. It allows limited conformational flexibility of the host, while substantially reducing the electrostatic calculation of the host self-interaction. In this approach, a fictitious ligand–host complex consisting of one host and multiple ligands (copies) is simulated. The copies are energetically transparent to one another, while each interacts with the host normally. The host experiences the average force from all the copies. This is similar to, but not exactly the same as, the mean-field approximation routinely used in statistical mechanics.[7] The gain of sampling efficiency comes from the self-interaction of the host being calculated once for all of the copies. The computational time $\tau_{multi}(m)$ required for the electrostatic calculation of m ligand–host conformations is:

$$\tau_{multi}(m) \sim N(N-1)/2 + mnN + mn(n-1)/2$$

$$\sim N^2/2 + mnN \qquad \text{for } N \gg 1.$$

Thus, the multicopy electrostatic calculation of $m = N/2n$ ligand conformations requires twice the time as needed by the traditional method for the calculation of a single conformation, indicating that the multicopy sampling is also combinatorial. Consider a protein host with 100 amino acids and a ligand of the size of a single amino acid, the gain of sampling efficiency is $m/2 = 25$-fold. While this gain is apparently much lower than that of the grid method, the quality of multicopy sampling is significantly improved, since the host is conformationally flexible.[8] Moreover, the quality of sampling is also improved due to the effectively smoothed energy surface in multicopy sampling.[8,9] The gain from a smoothed energy surface in terms of sampling efficiency and structural evaluation is a unique feature of multicopy sampling and its full potential remains to be explored.

Structural evaluation

There are two different types of structural evaluation. One is to determine the optimal conformation from a pool of low-energy conformations for a single ligand. A more difficult one is to determine the most favorable ligand along with its optimal conformation(s) from a pool of many conformations of many different ligands. The latter is critical for a true implementation of computational screening of combinatorial libraries and is the focus of this work. In contrast to the significant progress toward improving conformational sampling, the search for accurate and reliable criteria for structural evaluation has proven to be challenging.[10,11] One natural explanation is that there is no guarantee of success for the use of an empirical energy function to describe the complex biomolecular structure and interaction. The binding affinity of a ligand–host complex is essentially determined by the energy and entropy differences between the bound conformation and many

unbound conformations of the ligand. While the energy and entropy of the bound conformation are relatively easy to calculate, an accurate estimate of the entropy and average energy of the unbound conformations is difficult in the absence of an exhaustive conformational sampling.

Two approaches have been used for estimating ligand binding affinities. One is the free-energy integration/perturbation method based on molecular dynamics.[12,13] This method is theoretically rigorous, and has been widely used to calculate the relative binding affinity between two ligands, along with their bound conformations. If successfully applied, this method can provide insight and detailed information of molecular interaction that is difficult or impossible to obtain via experimental means. Its main disadvantages are that the thermodynamic calculation is computationally extensive and its final outcome is sensitive to many structural and energetic details that may be difficult to meet in practice.[11] The efficiency of this approach can be significantly enhanced with multicopy sampling method, if some technical problems are resolved.[14,15]

Another approach is to replace the expensive sampling of the unbound conformations with an empirical scoring function, which usually consists of energy, contact surface area, geometric fit, solvation and entropy.[16–19] The calculation of such a scoring function requires much less time than what would be required to perform a thermodynamic integration. For many practical drug design problems, an accurate prediction of the binding affinity of ligands may be neither realistic, nor necessary, since a drug's pharmacological profile in a living organism may depend on many factors beyond the scope of molecular modeling. A quick, approximate and reliable affinity ranking of many ligands is often sufficient. For this reason, the empirical approach has been widely adapted in computer-aided drug design.

A disadvantage of the empirical approach is that its scoring function is different from the energy function used in conformational sampling. Thus, the low-energy conformations sampled do not necessarily have better empirical scores, and vice versa. To increase the chance of finding conformations with better scores, more conformations must be sampled and subjected to scoring. However, this will reduce the efficiency of an empirical scoring function that is introduced to avoid performing extensive conformational sampling in the first place.

The above analysis of conformational sampling and structural evaluation leads to a list of features one must consider while developing efficient computational methods for the high-throughput screening of combinatorial libraries:

(1) A conformational sampling method with true combinatorial nature.

(2) A uniform target function for both conformational sampling and structural evaluation.

(3) An intrinsically (without resorting to any empirical means) smoothed energy surface.

(4) A substantially higher throughput than the free-energy integration method.

While some of these features are included in existing modeling methods, until now no single method offers them all.

A New Method for Computation Screening of Combinatorial Libraries

We now describe our new computational method into which we have attempted to include all of these features. It is based on a combination of the multicopy sampling method and a new criteria for structural evaluation.[20] While, in principle, the method is applicable to the general ligand–host systems, we chose to model the site-directed amino acid substitutions in a protein with known three-dimensional coordinates. Specifically, we studied the mutation of the conserved residue, Phe14, in the first zinc finger domain of protein Zif268 for the following reasons.[21] (1) The wild type amino acid Phe can be assumed to be most favorable at position 14, since it is highly conserved among known zinc fingers.[22] (2) Experimental data is available for several substitutions, Tyr, His and Leu[23,24] so that we can evaluate our calculated results against the known data. (3) The zinc finger domain is among the smallest independently folding units with 32 residues, making it a manageable test system on limited computer resources. Since our multicopy calculation requires large computer memory, memory space saved by using a small protein can be used to increase the structural diversity of copies. (4) We know an energy function that is accurate in describing the interaction between Phe14 and its surrounding hydrophobic core of the protein.[8]

Our objective was to reproduce the following experimental results on the zinc-finger system via computational screening of the combinatorial libraries comprised of all natural amino acid substitutions at position 14. Specifically, the experimental results indicate that the wild-type Phe is energetically most favorable, and substitutions by the ring-containing side chains of Tyr or His decrease the protein's thermodynamic stability.[23] The thermodynamic stability can be further decreased by a Leu substitution.[24]

Computational method

The modifications of energy and force calculation as required by multicopy sampling have been described previously.[25] All calculations were performed with our modified version of CHARMm22 on Silicon Graphics

workstations.[26] The All Hydrogen parameter set was used with a non-bonded cutoff distance of 10 Å. The solvation effects were approximated by using a distant-dependent dielectric of 4R, where R is the distance between two interacting atoms. This dielectric approximation was chosen based on our previous experience on modeling the same zinc-finger system.[8,25] Energy minimization was performed with the Adopted Basis Newton Raphson routine with an energy tolerance of 0.5 kcal/mol in 50 steps of energy minimization. The protein was constrained to its known crystallographically determined coordinates with a harmonic force of 1.0 kcal/mol Å per atom. This constraint was previously determined as sufficiently strong to prevent protein distortion, yet sufficiently weak to allow the host to undergo conformational changes to enhance conformational sampling.[8] All RMSDs were calculated for non-hydrogen atoms without rotation and translation.

A combinatorial library of 19 natural amino acids (Pro and Gly were excluded due to their irregular side chains) was constructed as follows. For each amino acid, 10 (an arbitrary number) random conformations were generated by randomizing all rotatable sidechain dihedral angles. The entire library was then inserted to the protein at position 14, as illustrated in Figure 1(A). This fictitious library-protein complex can be considered to be mathematically equivalent to the superposition of 190 normal proteins which differ only at position 14. For convenience, each conformation in the library was named a 'copy', and the remainder of the protein was named the protein. To increase statistical significance, 10 such random libraries were generated for screening.

The target function used for screening substitutions was $D_{\text{clustering}}$, which we have defined for each substituted amino acid as:[20]

$$D_{\text{clustering}} = (\text{RMSD}_0 - \text{RMSD})/\text{RMSD}_0,$$

where, RMSD_0 and RMSD are the average root-mean-square deviations of all possible pairs (e.g. a total of $10(10-1)/2=45$ pairs from 10 copies) of the copies before and after energy minimization, respectively. The normal range of $D_{\text{clustering}}$ is between 0 and 1. Substitutions with larger $D_{\text{clustering}}$ are considered more favorable. The optimal conformation for each substitution is identified by the largest cluster of its copies after energy minimization.

Results

Initial screening

The average and standard deviation of $D_{\text{clustering}}$ obtained from the screening of 10 different 190-copy libraries are shown in Figure 2A. Three observations were drawn. First, substitutions of Phe, His, His+(protonated His), Leu, Glu, Tyr, Gln, Asp and Asn scored higher than the other substitutions, whose upper-range of $D_{\text{clustering}}$ (Cys) is lower than the average $D_{\text{clustering}}$ of Asn. Second, four best substitutions Phe, His, His+, and Tyr suggest a preference for a ring-containing side chain at position 14. This provides a clear guidance for the design of pooling strategy for sub-library screening. Third, more than half of the random copies of the wild-type Phe clustered to near its native conformation after energy minimization (Fig. 1B). Given the enor-

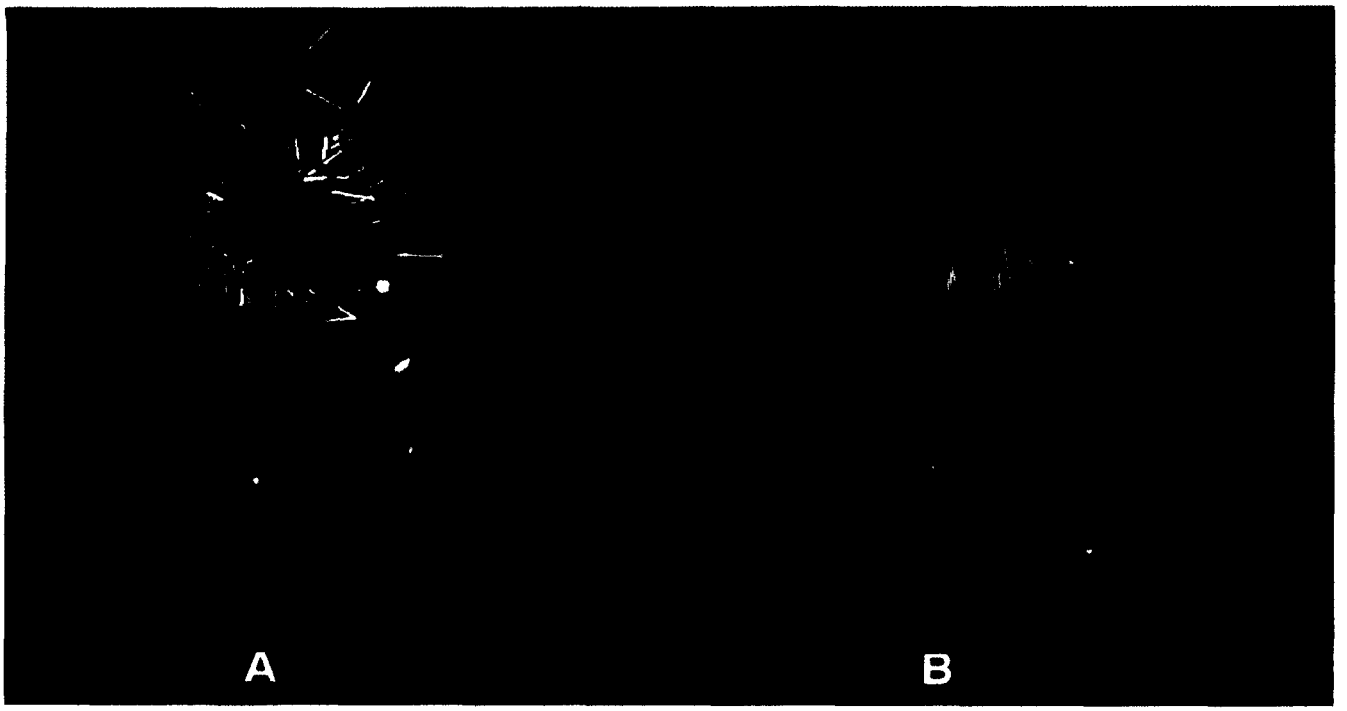

Figure 1. A combinatorial library of 19 amino acid substitutions constructed at position 14 of the zinc-finger protein whose α-carbon trace is shown in blue ribbon, along with the zinc ion (sphere). (A) The initial library consists of 190 random initial copies, 10 for each amino acid. (B) The copies (yellow) of the highest scoring substitution Phe side chain after energy minimization, along with its native conformation (red).

mous complexity of a 190-copy library of 19 amino acids, it is encouraging that a simple multicopy energy minimization can lead to such detailed results.

Sub-library screening

Ten sub-libraries were generated for the high scoring substitutions of Phe, His, His+, Leu, Glu, Tyr, Gln, Asp, and Asn in the initial library screening. Each sub-library consisted of 90 random copies, 10 for each of the nine substitutions. The average and standard deviation of $D_{clustering}$ obtained from screening these sub-libraries are shown in Figure 2B. Since the sub-libraries were less complex than the original ones, the signal-to-noise ratio of the $D_{clustering}$ was significantly improved. The wild-type Phe is clearly the most favorable substitution, followed by Tyr, His, and His+. This conclusion is further supported by that the standard deviation of $D_{clustering}$ of Phe is significantly smaller than that of Tyr, His and His+. Leu, Asp, Gln, and Asn are definitely less favorable than Phe, Tyr, His and His+, whose upper-range of $D_{clustering}$ is less than the average $D_{clustering}$ of Phe, Tyr, His or His+. These results confirm the previous indication of the preference of a ring-containing side chain. Finally, the optimal conformation of Phe is clearly identifiable for more than 80% of its random initial copies cluster to the native conformation after energy minimization.

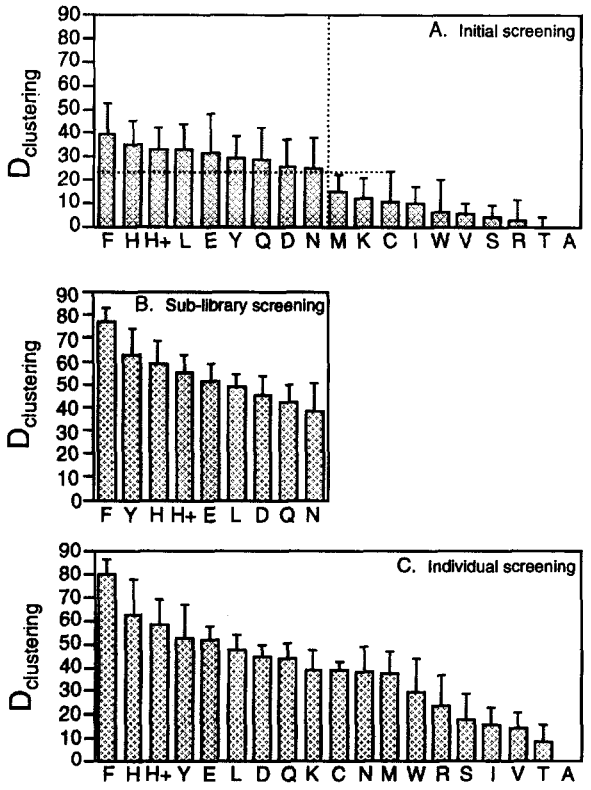

Figure 2. The $D_{clustering}$ profiles (the average and standard deviation of $D_{clustering}$) from the screening of the original 190-copy libraries (A), the 90-copy sub-libraries (B), and the 10-copy individual libraries (C). The average and standard deviation of each amino acid substitution in each library are calculated based on the screening of 10 different libraries with random initial conformations. In each screening, the amino acid substitutions are placed in the descending order of their average $D_{clustering}$ scores.

These results are in qualitative agreement with the published experimental mutagenesis data collected from several zinc fingers (Table 1). Whereas the Tyr and His mutants maintain similar zinc-finger folds, the His mutant is thermally less stable than the wild-type.[23] For example, Phe vs His: pH stability midpoints are 4.0 vs 4.5, and amide proton exchange rates in D_2O are 18 h vs 20 min. The Leu-mutant, while also maintaining the zinc-finger fold, causes marked decrease of thermal stability.[24] At pH 5.5, the Leu-mutant is $<90\%$ folded, while the wild-type is $>95\%$ folded at pH 6.3. Amide protons exchange rates in D_2O occur in hours for the wild-type and minutes for the Leu-mutant.

Individual screening

As an assurance of the initial and sub-library screenings, it was necessary to examine whether all the favorable substitutions were included in the sub-library, and whether the rank-order scoring of the sub-library screening was similar to that of the individual screening. For this purpose, 10 different 10-copy random libraries were generated for each of the 19 substitutions.

The average and standard deviation of $D_{clustering}$ obtained from screening these individual libraries are shown in Figure 2C. The rank-order scoring is very similar to that of the sub-library screening. Substitution Phe remains as having the highest scoring, followed by His, His+, and Tyr, although their relative order is altered from what was observed in the initial and sub-library screenings. The standard deviation of $D_{clustering}$ of Phe remains significantly smaller than that of Tyr, His, and His+. The rank-order of the next four substitutions of Glu, Leu, Asp and Gln is unchanged. The only substitution that slips out of the top-9 scoring list is Asn. Even in this case, its average $D_{clustering}$ is only marginally lower than that of Lys and Cys. Finally, the optimal side-chain conformation of Phe, corresponding to the largest cluster of copies for each amino acid, is readily identifiable, since 90% of its random copies cluster to the native conformation after energy minimization (Figure 3).

In summary, with a few minor exceptions, the $D_{clustering}$ profiles of the sub-library and individual screenings are remarkably similar. This demonstrates that the individual screening of 19 substitutions can be effectively replaced by two rounds of the library screening.

Energy screening

Since the library screening is driven by the multicopy energy minimization, one might imagine that the rank-order scoring of $D_{clustering}$ merely reflects the rank-order scoring of energy. To demonstrate that this is not the case, the rank-order scorings of energy of the original libraries, sub-libraries and the individual libraries were

Table 1. Correlation between the $D_{clustering}$ scores and the experimental results

Substitution	Initial library screening[b]	Sub-library screening	Individual screening	Stability (exp[a])
Phe	40±13 (1)[b]	77±6 (1)	80±7 (1)	Wild-type
His/His+[c]	34±10 (2)	57±9 (3)	61±13 (2)	↓
Tyr	29±10 (5)	63±12 (2)	53±15 (3)	↓
Leu	33±11 (3)	49±6 (5)	48±7 (5)	↓↓

[a] ↓ Denotes less stable; ↓↓ denotes even less stable. See text for details.
[b] Denotes the $D_{clustering}$ rank-order number of a substitution.
[c] Represents the average scoring and ranking of His and His+.

calculated and the results are shown in Figure 4. The energy of a given substitution used in scoring is the minimized energy of a complex consisting of the protein and 10 copies of the substitution. Substitutions of Phe, Tyr, His and His+ do not consistently rank favorably in the three energy screenings, and there is no indication of the preference for a ring-containing side chain. These results are not surprising since the favorability of any substitution is not totally determined by its energy of the optimal conformation. Rather, it is determined by the free-energy difference between the optimal conformation and many unfolded conformations.

Discussion and Conclusion

The only assumption made in this work is that the wild-type Phe is the most favorable substitution at a highly conserved position in the zinc finger fold. Although our computational screening method does not rely on this assumption, our interpretation of the experimental mutagenesis data in terms of the $D_{clustering}$ scoring does. While the assumption that the wild-type is most thermodynamically stable may not always be valid, it is consistent with the experimental cited in this work. Moreover, a similar assumption has been widely used in deriving empirical energies based on the contact frequencies of amino acids from known protein coordinates.[27,28] On the other hand, the correlation between the $D_{clustering}$ scoring and protein stability is only qualitative, since the mutagenesis data is taken from several different zinc-fingers although their share the same three-dimensional fold.

While the concept of $D_{clustering}$ originated from our empirical observations, its relationship to the association energy and entropy can be rationalized as follows. First, during energy minimization, large energy decreases →large conformational changes →high clustering. For each amino acid substitution, if its initial copies are completely random and most of the

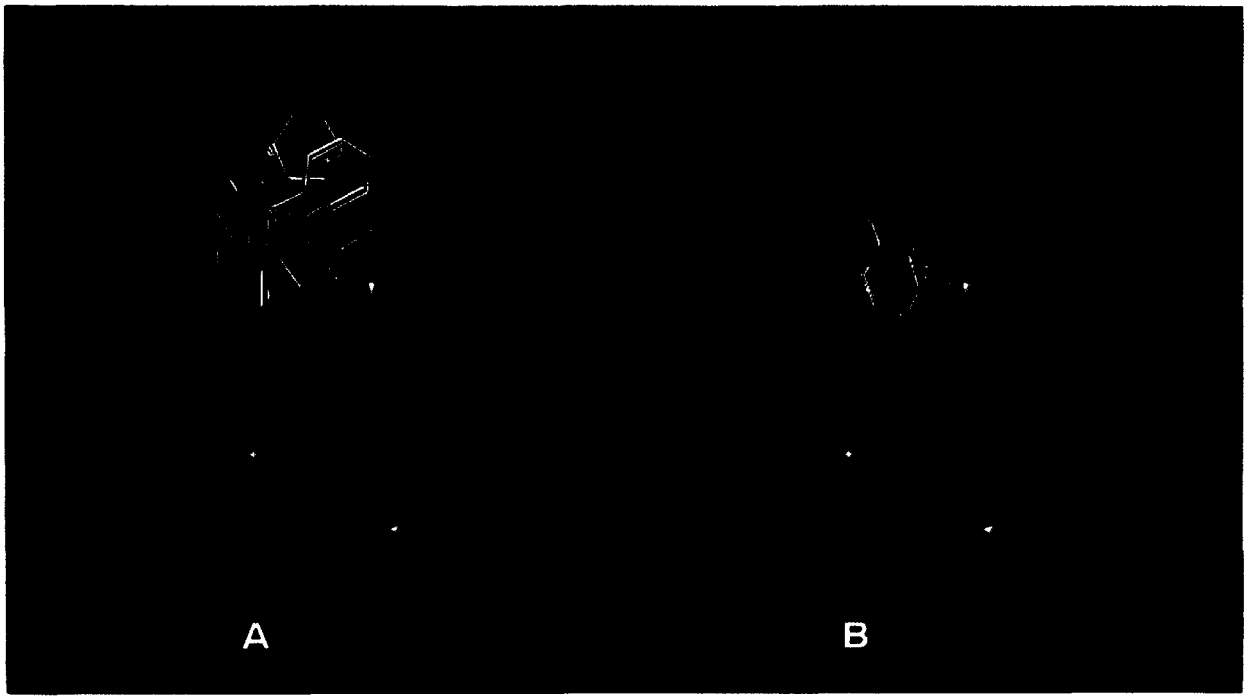

Figure 3. The clustering of the random initial copies of Phe at position 14 in the zinc-finger protein after the multicopy energy minimization: the α-carbon trace and the zinc (blue); the native conformation of the Phe side chain (red); and the copies (yellow) of the Phe side chain, before (A) and after (B) energy minimization.

minimized copies converge to a folded conformation, the energy decrease during minimization is an approximate measure of the average energy difference between the folded and unfolded conformations. Therefore, $D_{clustering}$ reflects the energy difference between a folded and many unfolded conformations. Second, for each amino acid substitution, $RMSD_0$ is a measure of the size of the conformational space spanned by the initial random copies. RMSD is a measure of the size of the conformational space spanned by the minimized copies. During energy minimization, copies clustered to their nearest energy minima, thereby reducing the conformational space. Minima with larger attraction basins contribute more to the reduction of the average conformational space as measured by $RMSD_0 - RMSD$. A more intrinsic measure is given by $D_{clustering} = (RMSD_0 - RMSD)/RMSD_0$, where the sidechain conformational change is measured relative to the size of the conformational space of random individual copies. The introduction of the normalization factor of $RMSD_0$ is critical, since it provides an absolute measure of the association affinity of each ligand, thus eliminating the need to compare one ligand to another, a technique used in the free-energy integration approach.

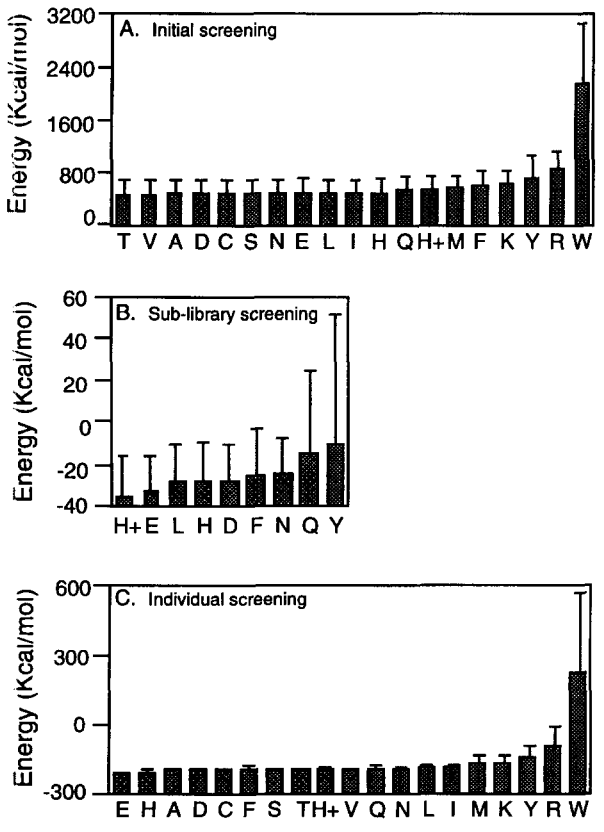

Figure 4. The energy profiles (the average and standard deviation of energy) from the screening of the original 190-copy libraries (A), the 90-copy sub-libraries (B), and the 10-copy individual libraries (C). The average and standard deviation of each amino acid substitution in each library are calculated based on the screening of 10 different libraries with random initial conformations. In each screening, the amino acid substitutions are placed in the increasing order of their average energies.

The combination of the multicopy sampling and $D_{clustering}$ methods, provides a natural platform for computational screening of combinatorial libraries. Indeed, the simultaneous modeling of 19 substitutions can be naturally viewed as a computer simulation of the screening of a combinatorial peptide library. In this view, the method also provides an efficient means to explore structure–activity information for the sub-library construction and screening. For example, the high $D_{clustering}$ of Phe, Tyr, His, and His+ would suggest to create sub-libraries to explore all possible ring-containing side chains.

Since the multicopy sampling method was proposed 5 years ago, its sampling efficiency has been the focus of attention. Little attention has been given to utilizing the large number of conformations generated by the new method to improve the accuracy and reliability of structural evaluation.[29] One of the most difficult elements in structural evaluation is the calculation of conformational entropy. This calculation is crucial since the ultimate goal of conformational sampling is to search for global energy minima with a large attraction basin (i.e. large conformational entropy). The clustering of copies, as an inherent feature and a natural outcome of multicopy sampling, contains a wealth of entropic information that has just begun to be explored.

Finally, our entire procedure of library construction and screening is independent of atomic details, thus readily applicable to combinatorial libraries of non-natural amino acids and small molecules for studying site-directed substitutions or ligand–receptor interactions. We would caution that given the uncertainties in discriminating between native vs misfolded conformations of a given molecule,[10] it is highly challenging to directly evaluate heterogeneous molecules without resorting to empirical rules. While this work presents preliminary evidence and rationale of a possible relation between $D_{clustering}$ and the association energy and entropy, the exact nature and quantitative characterization of this relation remain to be elucidated.

References

1. Gallop, M. A.; Barret, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994**, *37*, 1233.

2. Gordon, E. M.; Barret, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. *J. Med. Chem.* **1994**, *37*, 1385.

3. Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849.

4. Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. *J. Med. Chem.* **1989**, *32*, 1083.

5. Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161.

6. Miranker, A.; Karplus, M. *Proteins* **1991**, *11*, 29.

7. Zheng, Q.; Rosenfeld, R.; Kyle, D. J. *J. Chem. Phys.* **1993**, *99*, 8892.

8. Zheng, Q.; Kyle, D. J. *Proteins* **1994**, *19*, 324.

9. Roitberg, A; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277.

10. Novotny, J.; Rashin, A. A.; Bruccoleri, R. E. *Proteins* **1988**, *4*, 19.

11. Shi, Y. Y.; Mark, A. E.; Wang, C. X.; Huang, F.; Berendsen, H. J. C.; van Gunsteren, W. F. *Prot. Engng.* **1993**, *6*, 289.

12. Strraatsma, T. P.; McCammon, J. A. *Methods in Enzymol.* **1991**, *202*, 497.

13. Kollman, P. *Chem. Rev.* **1993**, *93*, 2395.

14. Verkhivker, G.; Elber, R.; Nowak, W. *J. Chem. Phys.* **1992**, *97*, 7838.

15. Straub, J. E.; Karplus, M. *J. Chem. Phys.* **1991**, *94*, 6737.

16. Novotny, J.; Bruccoleri, R. E.; Saul, F. A. *Biochemistry* **1989**, *28*, 4735.

17. Kuntz, I. D. *Science* **1992**, *257*, 1078.

18. Bohm, H. J. *J. Comput-Aided Mol. Design.* **1994**, *8*, 243.

19. Vajda S.; Weng, Z.; Rosenfeld, R.; Delisi, C. *Biochemistry* **1994**, *33*, 13977.

20. Zheng, Q.; Kyle, D. J. *Proteins* **1996**, *24*, 209.

21. Pavletich, N. P.; Pabo, C. O *Science* **1991**, *252*, 809.

22. Gibson, T. J.; Postma, J. P. M.; Brown, R. S.; Argos, P. *Protein Engng.* **1988**, *2*, 209.

23. Qian, X.; Weiss, M. A. *Biochemistry* **1992**, *31*, 7463.

24. Mortishire-Smith, R. J.; Lee, M. S.; Bolinger, L.; Wright, P. E. *FEBS Lett.* **1992**, *296*, 11.

25. Zheng, Q.; Rosenfeld, R.; DeLisi, C.; Kyle, D. J. *Protein Sci.* **1994**, *3*, 493.

26. Brooks, B. R; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

27. Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534.

28. Covell, D. G.; Jernigan, R. L. *Biochemistry* **1990**, *29*, 3287.

29. Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisi, C. *J. Mol. Biol.* **1993**, *234*, 515.

(*Received in U.S.A.* 13 *September* 1995; *accepted* 13 *November* 1995)