

**Prediction of the quaternary structure of coiled coils:
GCN4 Leucine Zipper and its Mutants**

By

Michal Vieth^a

Andrzej Kolinski^{a,b}

Charles L. Brooks, III^a

Jeffrey Skolnick^{a,*}

September 15, 1995

^a Departments of Molecular Biology and Chemistry
The Scripps Research Institute
10666 North Torrey Pines Road
La Jolla, California 92037
Phone: (619) 554-4821
Fax: (619) 554-6717

^bDepartment of Chemistry
University of Warsaw
Pasteura 1, 02-093 Warsaw
POLAND

Keywords: /GCN4 Leucine Zipper / Multimeric Equilibrium / Quaternary Structure Prediction / Quaternary Structure Stability / Protein Folding Simulations /

* To whom correspondence should be addressed

Abstract

A methodology for predicting coiled coil quaternary structure and for the dissection of the interactions responsible for the global fold is described. Application is made to the equilibrium between different oligomeric species of the wild type GCN4 leucine zipper and seven of its mutants that were studied by Harbury et al. Over the entire experimental concentration range, agreement with experiment is found in five cases, while in two other cases, agreement is found over a portion of the concentration range. These simulations suggest that the degree of chain association is determined by the balance between specific side chain packing preferences and the entropy reduction associated with side chain burial in higher order multimers.

1: Introduction

Due to their biological importance and inherent structural simplicity, coiled coils are the object of increasing attention. Among their many biological functions, they comprise a key motif of DNA^{1,2} and RNA^{3,4} binding proteins. Coiled coils exhibit a characteristic seven residue repeat $(abcdefg)_n^{5,6}$ which produces a native structure formed by two or more helices wrapped around each other with a left handed, helical supertwist⁷. Positions *a* and *d*, which form the helical interface, are mostly occupied by hydrophobic residues. Positions *b*, *c*, *e*, *f*, *g* are hydrophilic^{4,5}. Residues occupying the *g* and *e* positions tend to be charged and are believed to play a role in defining the mutual orientation of the helices. Furthermore, since coiled coils constitute the simplest quaternary structure, they represent a very useful model system for exploring the factors responsible for the stability and specificity of oligomeric proteins. In this context, Harbury *et al.*⁸ simultaneously substituted the four *a* residues of the GCN4 Leucine Zipper (Val9, Asn16, Val23, Val30) and the four *d* residues (Leu5, Leu12, Leu19, Leu26) by Leu, Ile and Val. The modified peptides were named according to the identity of the residues in the *a* and *d* positions (e.g., LI stands for the mutant with Leu (Ile) in all four of the *a* (*d*) positions) and are more than 90% helical. The IL mutant and the wild type populate dimeric species; II, LL, LV are trimeric, and LI is tetrameric. The VL mutant populates both dimeric and trimeric species, and the VI mutant populates multiple species.

This paper extends our previous predictions of the folding pathway and structure of the wild type GCN4 Leucine Zipper⁹ to the calculation of the equilibrium constant between different oligomeric species. Because of practical limitations on the direct simulation to predict the state of association of a collection of chains, we have developed a methodology that assumes a spectrum of parallel and antiparallel oligomers and attempts to estimate the equilibrium constants within the set of assumed species (schematically shown in Figure 1). The methodology is based on a new application of the classical Mayer and Mayer statistical mechanical approach¹⁰, but where we use a computer simulation to obtain the necessary variables for the statistical mechanical treatment. Most importantly, in the context of the model, the method allows for identification of the dominant interactions responsible for coiled coil quaternary structure.

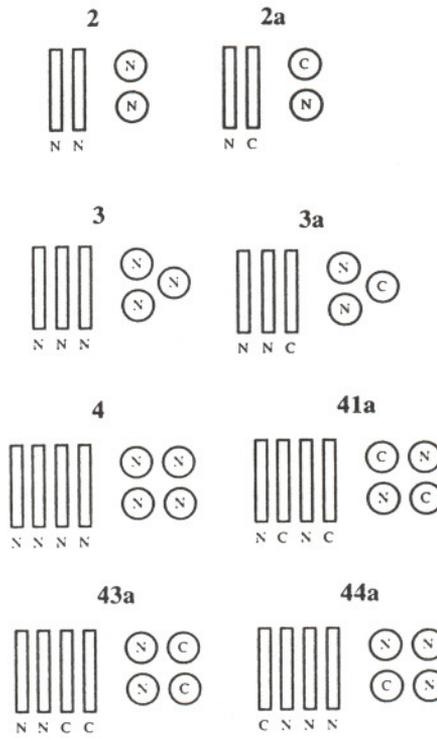


Figure 1

Schematic drawing of the interhelical orientations studied. 2 represents parallel dimers, 2a antiparallel dimers, 3 parallel trimers, 3a antiparallel trimers, 4 parallel tetramers. 41a, 43a and 44a represent possible antiparallel tetramers studied in this work.

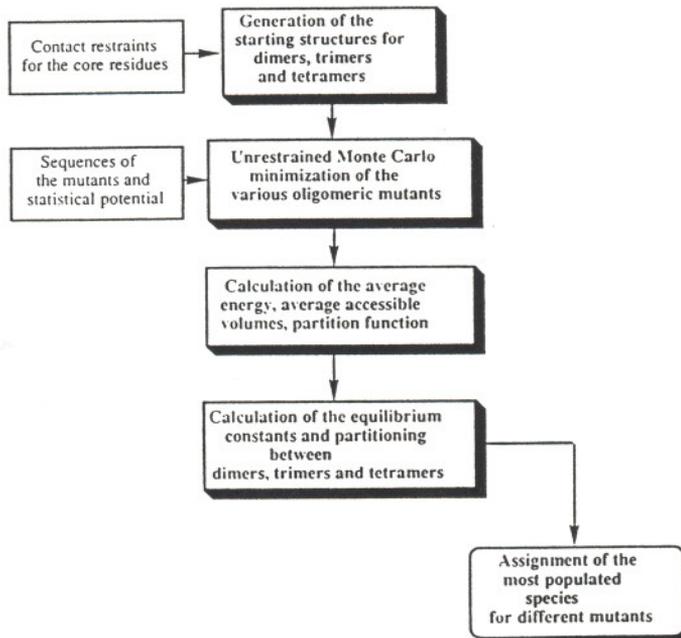


Figure 2

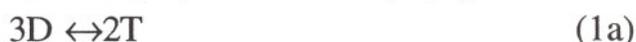
For each mutant, every oligomer is generated and subjected to unrestrained, isothermal Monte Carlo simulations under conditions (energy function, temperature) identical to those for which the wild type GCN4 Leucine Zipper was refined⁹. Then, the partition functions for each mutant in every oligomeric state are calculated, and the most populated species are assigned for the relevant chain concentration.

2: Method

An overview of the entire simulation methodology is presented in Figure 2. The lattice model used to estimate the equilibrium constants is based on an α -carbon representation of the protein backbone and a multiple rotamer, single ball representation of the side chains and is described elsewhere^{11,12}. The entire potential energy parameter set is available by anonymous ftp¹³ and includes potentials that reflect intrinsic secondary structural preferences, hydrogen bonds, the preference of amino acids to be buried or exposed, side chain pair interactions, and terms that reflect cooperative side chain packing in proteins.

3: Protocol for extracting the equilibrium constant from a simulation

In order to compare with experiment, we have to calculate the equilibrium constants associated with the dimer, D, trimer, T, and tetramer, R, species.



The equilibrium constants¹⁰ are:

$$K_{DT} = \frac{\{T\}^2}{\{D\}^3} \quad (2a)$$

$$K_{DR} = \frac{\{R\}}{\{D\}^2} \quad (2b)$$

with $\{D\}$, $\{T\}$ and $\{R\}$ the concentration of dimer, trimer and tetramer, respectively. Statistical mechanics relates the equilibrium constants from Eq. 2a,b to the configurational partition functions as^{10,14} :

$$K_{DT} = \frac{VZ_{\text{int},T}^2 \sigma_D^3}{Z_{\text{int},D}^3 \sigma_T^2} \quad (3a)$$

$$K_{DR} = \frac{\{R\}}{\{D\}^2} = \frac{VZ_{\text{int},R} \sigma_D^2}{Z_{\text{int},D}^2 \sigma_R} \quad (3b)$$

with V the total volume of the system and σ_g is the symmetry number ($\sigma=2!$, $3!$, $4!$ for homo dimers, trimers and tetramers, respectively). $Z_{\text{int},\gamma}$ are the integrals corresponding to integration over the internal coordinate degrees of freedom (also called the internal partition function) for oligomer γ , respectively.

4: Calculation of the internal partition function

Consider a system comprised of $3N_\gamma$ coordinates. Here, N_γ corresponds to the number of distinct structural elements, i.e., the number of C_α s and side chain centers of mass. The probability of having a conformation (with the first group fixed in space) inside a $3N-3$ dimensional volume element, centered about $\mathbf{r}=(\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4, \dots,\mathbf{r}_{N_\gamma})$ is:

$$P_u(\mathbf{r}) = \frac{\exp(-E(\mathbf{r})/kT)}{Z_{\text{int},\gamma}} \left(\prod_{i=2}^{N_\gamma} dv_i \right) \quad (4a)$$

$E(\mathbf{r})$ denotes the energy of the internal conformational state \mathbf{r} , k is Boltzmann's constant, and T is the temperature. Eq.4a can be used to precisely calculate the internal partition function $Z_{\text{int},\gamma}$ provided that the corresponding probabilities can be obtained (e.g., from a Monte Carlo simulation).

Now, let us concentrate on the calculation of $P_u(\mathbf{r})$. We first fix the origin at the coordinates of the first C_α ^{15,16}. The coordinates of the second C_α are expressed in a spherical coordinate system (R_2,θ_2,ϕ_2) whose origin is at the first C_α . Similarly, the third C_α is expressed in terms of coordinates (R_3,θ_3,ϕ_3) expressed with respect to an origin located at the second C_α . The configurational partition function is independent of (R_2,θ_2,ϕ_2) , which comprise three Euler angles. Hence, the probability of seeing a specific value of (R_2,θ_2,ϕ_2) , $P(\theta_2)P(\phi_2)P(\phi_3)$, is just $1/8\pi^2$ ¹⁵ and:

$$\begin{aligned} P_u(\mathbf{r}_2,\mathbf{r}_3\dots\mathbf{r}_{N_\gamma}) &= P(\theta_2)P(\phi_2)P(\phi_3)P(R_2,R_3,\theta_3\dots) = \\ &= \frac{1}{8\pi^2} P(R_2,R_3,\theta_3\dots) \end{aligned} \quad (4b)$$

The probability calculated from the Monte Carlo simulation, $P(\mathbf{r})$, has to be corrected due to the fact that only a portion, Ω , of the entire range of Euler angles is sampled during the course of the simulation. Thus, $P(\mathbf{r})$ is given by:

$$P(\mathbf{r}) = \frac{1}{\Omega} P(R_2,R_3,\theta_3\dots) \quad (4c)$$

Substituting Eq. 4c into Eq 4b gives the probability that the system is free to assume all possible orientations of (θ_2,ϕ_2,ϕ_3) :

$$P_u(\mathbf{r}_2,\mathbf{r}_3\dots\mathbf{r}_{N_\gamma}) = \frac{\Omega}{8\pi^2} P(\mathbf{r}) = Q_{MC} P(\mathbf{r}) \quad (4d)$$

For all of the simulations, we calculate Q_{MC} (the correction term for sampling a limited range of rotations) as the average number of observed two consecutive C_α - C_α vectors divided by the total number of possible two consecutive C_α - C_α

vectors. In all cases, this number is close to $1/80$, and, for simplicity, we can assume that our simulations sample a unique orientation of the molecule in space (that is, $QMC \approx 1/8\pi^2$).

By calculating the fraction of time a system spends in a given state (\mathbf{r}), a dynamic Monte Carlo method provides $P(\mathbf{r})$. Note that \mathbf{r} can be any conformational state. However, in what follows, because it occurs most frequently, the most probable state is used. However, for systems having substantial conformational fluctuations, the probability $P(\mathbf{r})$ cannot be reliably calculated due to the poor sampling statistics. This requires that a number of simplifying approximations be made.

5: Local volume factorization

To enrich the sampling, the probability $P(\mathbf{r})$ of the entire structure being in the $3N_\gamma-3$ dimensional volume element (centered about the most probable conformational state) is approximated as the product of the $N_\gamma-1$ independent probabilities that each group is in a 3 dimensional box centered around its most probable state. By treating each group separately, the statistics are greatly enhanced, relative to the case when we require the simultaneous occurrence of a set of \mathbf{r} . That is,

$$P(\mathbf{r}) \cong \prod_{i=2}^{N_\gamma} P_{i,\max}(\mathbf{r}_i) \quad (5a)$$

Eq. 5a is referred to as the local volume factorization approximation. Since the choice of the first bead as the origin of our internal coordinate system is arbitrary, to remove this arbitrariness, the total probability $P(\mathbf{r})$ is better approximated as the product of N_γ independent probabilities divided by their geometric mean:

$$P(\mathbf{r}) \cong \frac{\prod_{i=1}^{N_\gamma} P_{i,\max}(\mathbf{r}_i)}{\left(\prod_{i=1}^{N_\gamma} P_{i,\max}(\mathbf{r}_i)\right)^{1/N_\gamma}} = \left(\prod_{i=1}^{N_\gamma} P_{i,\max}(\mathbf{r}_i)\right)^{1-1/N_\gamma} \quad (5b)$$

(If $P(\mathbf{r})$ were accurately calculated, the results would be independent of the choice of origin.) The most probable position of the i -th group is computed from the trajectory as the location of maximal frequency of occupation of a given cubic volume element of length $dr=2.6\text{\AA}$. Using the local volume factorization approximation (Eq. 5a), the internal partition function for oligomer γ is obtained from:

$$Z_{\text{int},\gamma} \cong 8\pi^2 \exp(-E(\gamma, \mathbf{r}) / kT) \left(\prod_{N_\gamma} \frac{dv_i}{P_{i,\text{max}}(\mathbf{r}_i)} \right)^{1-1/N_\gamma} \quad (5c)$$

$E(\gamma, \mathbf{r})$ corresponds to the energy of the most probable conformation of oligomer γ . While this approximation is not exact, for test energy functions having a similar character as those used here, the local volume factorization approximation gives satisfactory estimates for the partition functions; errors in the equilibrium constants are on the order of 10-20%.¹⁷

6: Results

The method described above was applied to the various oligomeric states depicted in Figure 1 for the wild type and a number of mutants of the GCN4 Leucine Zipper⁸. All helical orientations (including three antiparallel orientations of helices in tetramers) were considered for the LL mutant of the GCN4 Leucine Zipper. The computed free energies for the antiparallel species are considerable higher (by $\sim 5kT/\text{monomer}$) than those of the parallel species. This energy difference is sufficiently great that antiparallel structures of this mutant can be dismissed. The energy difference mainly arises from unfavorable charge interactions. The preferential stability of parallel over antiparallel species seems to hold for all other mutants. However, we ignore the possibility of higher order aggregates (i.e., pentamers, dimers of trimers etc.) which, in principle, might occur^{13,18}.

Employing Eq. 2-6, the partitioning between dimers, trimers and tetramers has been calculated for each mutant. Due to the limited accuracy of our energy function, as well as the approximations used in the probability calculations, we restrict ourselves to the prediction of the dominant species for each mutant at a given concentration. Thus, the partitioning at low ($2\mu\text{M}$) and high ($200\mu\text{M}$) concentration that corresponds to the concentration range studied experimentally is calculated in Table I. For all cases considered, we find that over the experimentally measured concentration regime⁸, the predicted dominant species is the same. However, because of the law of mass action in the low concentration regime (about $2\mu\text{M}$), the population of lower order oligomers increases.

Table 1 compares the predictions with the experimentally determined degree of chain association. For 4 of the 8 cases, the predictions completely agree with the experimentally determined dominant species⁸. For the LV mutant, while trimers, in agreement with experiment, are always the dominant species, at low concentrations, given the uncertainty in the calculation, dimers may be populated. In the offending case of the IL mutant, trimers and dimers are assigned to be the dominant species which is in partial contrast to the experiment which indicates that only dimers are present. This may reflect the inaccuracy of the potential, as well as an accumulation of errors in the entropy calculation (for this mutant, the entropy increases with the degree of association). For the VL mutant, dimeric species negligibly contribute, and over the entire concentration regime, trimers are assigned to be the dominant species. In contrast, experiment indicates that both dimers and trimers are populated⁸. For the VI mutant, trimers are predicted to be

the only species over the entire concentration regime, whereas experiment shows that multiple species are populated⁸.

Table 1

Comparison of the predicted state of association with experiment

Mutation a d	Dominant species from		Predicted Population ⁱ	Predicted Population ⁱ
	Experiment	Simulation	2 μ M	200 μ M
wild type	2	2	99.5:0.5:0	95:5:0
I L	2	2, 3	65:32:0	19:61:19
I I	3	3	0:100:0	0:100:0
L I	4	4	2:23:75	0:9:91
V I	?	3	0:100:0	0:100:0
L V	3	3	49:51:0	15:85:0
V L	(2,3)	3	2:98:0	0.5:99.5:0
L L	3	3	33:67:0	8:92:0

i) Predicted percentage of the dimers, trimers and tetramers, respectively.

The investigation of the individual contributions to the free energy indicate that the dominant contribution to the effective entropy change (60-90%) comes from the side chains; the effective entropy change for the backbone is smaller, but non negligible. The greatest contribution to the entropy (largest accessible volume) comes from the C-terminal ends of the molecules. This prediction is consistent with the crystal structures of the wild type dimer and the LI tetramer. In both cases, the last two C-terminal residues are highly disordered^{2,13}.

In the case of the wild type GCN4 Leucine Zipper, relative to dimers, trimers are favored energetically (by about 2 kT/monomer) and disfavored entropically. In the wild type, trimers are more stable than dimers, but over the experimental concentration regime only dimers are predicted. Asn 16 in the wild type destabilizes the trimer (dimer) locally in the vicinity of residues 14-18 by 6.1 kT (4.1 kT) per monomer (plus a constant value that reflects the effect of the mutation on the unfolded state). Other parts of the wild type trimer play a stabilizing role; consequently, compensation effects are present. Our calculations

indicate that the effect of a single point mutation is not local, but propagates for at least one helical turn. This is in agreement with studies¹⁹ on tropomyosin fragments, where compensation effects are also present. We find that the N16V mutation stabilizes trimers more than dimers by roughly 2.8 kT per monomer.

Our calculations also suggest that short range, intrinsic secondary structure preferences favor lower order oligomers. Furthermore, the reduction in side chain effective entropy on burial in the core of trimers and tetramers also favors lower order oligomers. Long range interactions (burial preferences, cooperative side chain packing interactions, and side chain pairwise interactions, the last being the most specific) favor higher order oligomers. In higher order multimers, side chains in the core (*a* and *d* residues) are more buried and experience additional favorable hydrophobic interactions. The competition between short range and long range interactions and the effective side chain entropy change are the major factors that determine the dominant species for the mutants studied here.

Harbury *et al.*⁸ attribute the different levels of stability of various GCN4 mutants to the preferential relative angular packing of different side chains. In the known crystal structures, parallel packing occurs at the *a* positions in tetramers and *d* positions in dimers, whereas perpendicular packing occurs at the *a* positions in dimers and *d* positions in tetramers. Acute packing is exhibited by trimers. Based on the most populated rotamer, Ile and Val side chains prefer to pack in the perpendicular or acute fashion, and Leu in the parallel fashion. This is perhaps the reason why LI forms trimers, IL tetramers and II, trimers. In our model, however, the related term doesn't exhibit such a trend, but this may be partially due to the fuzziness of the simplified side chain representation. Our explanation of specificity is based on the competitive effects of the pairwise interactions, side chain packing (long range interactions favor higher order species) and side chains orientational packing preferences (short range interactions favor lower order oligomers); together with the loss of configurational entropy (which favors lower order oligomers). Consequently, on average, we see the population of the statistically most favorable rotamer in the majority of species, and there is no selection based on the lowest energy of rotamers.

7: Conclusion

In this paper, a new application of the Mayer and Mayer approach to calculate the equilibrium constant between dimeric, trimeric, and tetrameric coiled coils has been described. This approach, combined with a lattice protein model, successfully predicts the state of association of the majority of different mutants of the GCN4 leucine zipper. Based on the detailed dissection of the interaction energy, local interactions were found to stabilize lower order oligomers, whereas tertiary/quaternary interactions stabilize higher order oligomers. In most cases, the internal entropy of side chains was found to stabilize low order oligomers. The main difference in the population of different oligomeric species of various mutants arises from the interplay between different interaction environments for the *a* and *d* positions in dimers, trimers, and tetramers, differential packing preferences and the effective entropy change associated with side chain burial.

Acknowledgments

This work was supported in part by NIH Grants GM-38794, GM-37554, and FIRCA PA-91-77.

References

1. A. R. Ferre-D'Amare, G. C. Prendergast, E. B. Ziff, S. K. Burley, *Nature*, **363**, 38-45 (1993).
2. E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science*, **254**, 539-544 (1991).
3. D. W. Banner, M. Kokkinidis, D. Tsernoglou, *J. Mol. Biol.*, **196**, 657-675 (1987).
4. C. Cohen, D. A. D. Parry, *Trends Biochem. Sci.*, **11**, 245-248 (1986).
5. R. S. Hodges, A. S. Saund, P. C. S. Chong, S. A. St-Pierre, R. E. Reid, *J. Biol. Chem.*, **256**, 1214-1224 (1981).
6. A. D. McLachlan, M. Stewart, *J. Mol. Biol.*, **98**, 293-304 (1975).
7. F. H. C. Crick, *Acta Cryst.*, **6**, 689-697 (1953).
8. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science*, **262**, 1401-1407 (1993).
9. M. Vieth, A. Kolinski, C. L. Brooks III, J. Skolnick, *J. Mol. Biol.*, **237**, 361-367 (1994).
10. J. E. Mayer, M. G. Mayer, *Statistical Mechanics*. ed. New York: John Wiley & Sons, Inc. , chapter 9,(1963).
11. A. Kolinski, J. Skolnick, *Proteins*, **18**, 353-366 (1994a).
12. A. Kolinski, J. Skolnick, *Proteins*, **18**, 338-352 (1994b).
13. M. Vieth, J. Skolnick, A. Kolinski, A. Godzik, *available via anonymous ftp from ftp.scripps.scripps.edu in pub/(MCDP.info, MCDP.Z)* (1994).
14. R. Fowler, E. A. Guggenheim, *Statistical Thermodynamics*. Cambridge: University Press (1960).
15. D. R. Herschbach, *The Journal of Chemical Physics*, **31**, 1652-1661 (1959).
16. N. Davidson, *Statistical Mechanics*. New York: McGraw-Hill Book Company, Inc. (1962).
17. M. Vieth, A. Kolinski, J. Skolnick, *J. Chem. Phys.*, **102**, 6189-6193 (1995).
18. J. Chmielewski, *J. Am. Chem. Soc.*, **116**, 6451-6452 (1994).
19. M. E. Holtzer, A. Holtzer, *Biopolymers*, **30**, 985-993 (1990).