Factors Influencing Accuracy of Computer-Built Models: A Study Based on Leucine Zipper GCN4 Structure

Liyang Shen, Robert E. Bruccoleri, Stanley Krystek, and Jiri Novotny
Department of Macromolecular Modeling, Bristol-Myers Squibb Research Institute, Princeton, New Jersey 08543-4000 USA

ABSTRACT A three-dimensional model of the leucine zipper GCN4 built from its amino acid sequence had been reported previously by us. When the two alternative x-ray structures of the GCN4 dimer became available, the root mean square (r.m.s.) shifts between our model and the structures were determined as ~2.7 Å on all atoms. These values are similar to the r.m.s. shift of 2.8 Å between the two GCN4 structures in the different crystal forms (C_2 and $P2_12_12_1$). CONGEN conformational searches were run to better understand the conditions that may determine the preference of different conformers in different environments and to test the sensitivity of our current modeling techniques. With a judicious choice of CONGEN search parameters, the backbone r.m.s. deviation improved to 0.8 Å and 2.5 Å on all atoms. The side-chain conformations of Val and Leu at the helical interface were well reproduced (1.2 Å r.m.s.), and the large side-chain misplacements occurred with only a small number of charged amino acids and a tyrosine. Inclusion of the crystal environment (C_2 symmetry), as a passive background, into the side-chain conformational search further improved the accuracy of the model to an r.m.s. deviation of 2.1 Å. Conformational searches carried out in the two different crystal environments and employing the AMBER protein/DNA forcefield, as implemented in CONGEN, gave the r.m.s. values of 2.2 Å (for the C_2 symmetry) and 2.5 Å (for the $P2_12_12_1$ symmetry). In the C_2 symmetry crystal, as much as 40% of the surface of each dimer was involved in crystal contacts with other dimers, and the charged residues on the surface often interacted with immobilized water molecules. Thus, occasional large r.m.s. deviations between the model and the x-ray side chains were due to specific conditions that did not occur in solution.

INTRODUCTION

The "leucine zipper" motif is a characteristic amino acid sequence found in dimeric DNA-binding proteins such as transcriptional regulatory proteins (e.g., GCN4) and transforming factors (e.g., fos and jun) (Landschulz et al., 1988; Kouzarides and Ziff, 1988). The motif consists of a heptad repeated five times in a stretch of 35 residues near the C-terminus, and it displays all of the characteristic patterns of α-helical coiled-coils (O'Shea et al., 1989).

Crick (1953) described the coiled-coil three-dimensional structure as a noncovalent dimer of left-handed superhelices with a superpitch 186 Å and 3.5 residues per helical turn. The coiled-coil heptads generally include hydrophobic amino acids in helix-helix contact positions, **a** and **d** according to the **a-g-a′-g′** helical wheel scheme of McLachlan and Stewart (1975; see also Schulz and Schirmer, 1979). Leucine zipper sequences are distinguished from other coiled-coil motifs by having a leucine in every **d**, **d′** position, whereas the other contact positions, **a**, **a′**, are less stringently conserved. NMR spectroscopy (Oas et al., 1990) and x-ray crystallography (O'Shea et al., 1991; Ellenberger et al., 1992; Marmorstein et al., 1992) confirmed that the GCN4 polypeptides form parallel α-helical supercoiled dimers with the hydrophobic **a** and **d** positions at the interface, such that side chains of one helix fit into the surface grooves of the other helix.

An analysis of the available x-ray crystallographic coiled-coil structures (Seo and Cohen, 1993) revealed a diversity of superhelical pitch values, from 150 to 200 Å, and even more frequent local pitch variations. These are associated with various amino acid sequence types in the apolar interface. On the other hand, a comparison of two crystallographic forms of the same leucine zipper sequence, GCN4 (O'Shea et al., 1991, PDB code 2ZTA; Ellenberger et al., 1992, PDB code 1YSA), shows that equivalent side chains may adopt different conformations in the different crystal forms (C_2 and $P2_12_12_1$, respectively) as exemplified by the 2.8-Å root-mean-square (r.m.s.) shift between the two structures (0.9 Å on the backbone). This degree of structural polymorphism is interesting and invites a further in-depth study into the conditions and interactions determining the preference of different conformers in the different crystal environment and in solution. For example, Junius et al. (1995) observed, by NMR, a rapid conformational exchange of the GCN4 side chains Asn 16 at the center of the dimer interface. Although, in the two different crystals, these side chains adopted strikingly different conformations (Fig. 1), their solution NMR signals were indistinguishable, implying an essentially symmetric conformation. Clearly, crystal conditions impose more asymmetric and selective conditions on the structure of small proteins, thereby "locking in" certain conformers. In the current paper, we attempt to investigate more closely such conditions and, at the same time, test the power of the current modeling algorithms.

Received for publication 12 September 1995 and in final form 27 November 1995.

Address reprint requests to Dr. Jiri Novotny, Department of Macromolecular Modeling, Bristol-Myers Squibb Research Institute, P.O. Box 4000, Princeton, NJ 08543-4000. Tel.: 609-252-6209; Fax: 609-252-6030; E-mail: novotny@bms.com.

0006-3495/96/03/1096/09 $2.00

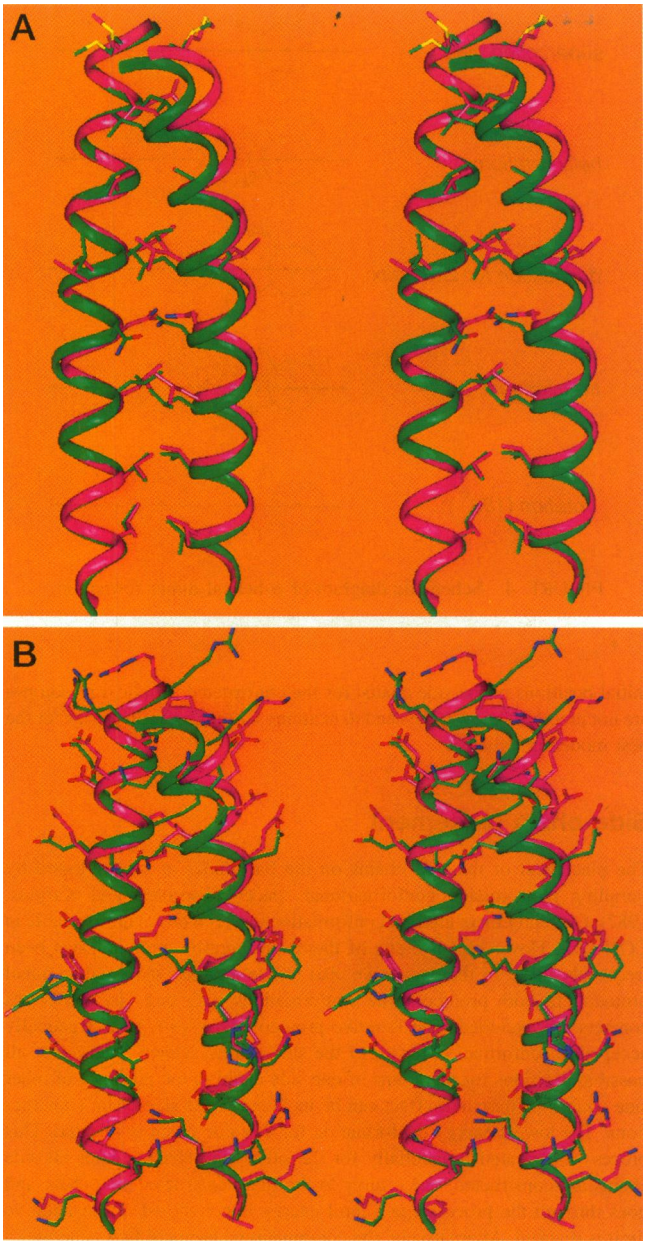

FIGURE 1 Stereo diagram of the least-squares superimposed GCN4 leucine zipper x-ray crystallographic structures (2.8 Å r.m.s.). α-Helical backbones are shown as ribbons, and the central hydrophobic residues (and the Asn 16) in the a and d positions are displayed as stick diagrams. In red, the O'Shea et al. (1991) structure (the C_2 unit cell); in green, the Ellenberger et al. (1992) structure (the $P2_12_12_1$ unit cell). The two alternative conformations adopted by the Asn 16 side-chain pair are clearly visible in the middle of the parallel dimer. Note that the C_2 structure is that of the leucine zipper dimer alone, whereas the $P2_12_12_1$ structure represents a part of a longer α-helical polypeptide in complex with its specific DNA. (*A*) Interface side chains; (*B*) surface side chains.

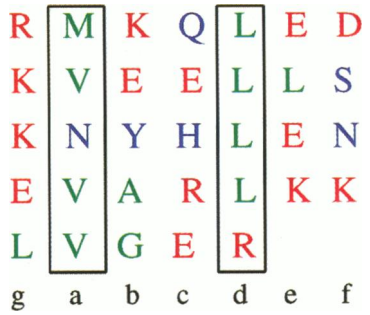

FIGURE 2 Amino acid sequence of the GCN4 leucine zipper. In red are the charge residues, in green hydrophobic residues, and in blue polar residues.

The three-dimensional structure of the GCN4 dimer has been modeled from its sequence (Fig. 2) by several groups. Before the x-ray structures of the dimer became available, Krystek et al. (1991) built models of the GCN4 and other coiled-coil leucine dimers using the molecular mechanics/conformational search program CONGEN (Bruccoleri, 1993). An alternative approach, simulated annealing, was employed by Nilges and Brunger (1991, 1993) and DeLano and Brunger (1994). More recently, Vieth et al. (1994) used a hierarchical approach, based on Monte Carlo dynamics driven by database-derived pseudopotentials, to generate a dynamical ensemble of representative dimer structures. Comparison of the models with the x-ray-derived structures (O'Shea et al., 1991; Ellenberger et al., 1992) showed that the supercoiled α-helical backbones were quite well reproduced in all the models: the respective backbone atom r.m.s. shifts were 1.1 Å for the model of Krystek et al., 1.3 Å for the model of Nilges and Brunger, and an average of 1.0 Å for an ensemble of the models of Vieth et al. DeLano and Brunger (1994) reported C_α atom r.m.s. shifts of their model and the O'Shea et al. (1991) x-ray structure as 0.7 Å, and the same C_α r.m.s. value was achieved by Krystek et al. (1991) model. For the model of Krystek et al., the all-atom r.m.s. deviation of the model from both of the x-ray structures (2.7 Å) was found to be about equal to that between the two x-ray structures themselves 2.8 Å. All-atom r.m.s. shifts of the models of Nilges and Brunger and Vieth et al. from the x-ray structures were not reported.

In this paper we systematically varied parameters of the CONGEN modeling procedure (superpitch; the skew angle of Harris et al., 1994) and those involved in the side-chain conformational searches (torsional angular grid; the form of the potential energy function) as we built alternative GCN4 models. In this way, we were able to determine the relative impact of dimer geometry and calculated energies on the accuracy of the final model. We also show that crystal lattice interactions played an important role in determining the local conformation of the major part of the GCN4 dimer surface. Once the crystal environment of the C_2 GCN4 x-ray structure (O'Shea et al., 1991) was explicitly included in the modeling protocol, the accuracy of the final model improved significantly. To model the crystal environment of the $P2_12_12_1$ structure (Ellenberger et al., 1992), which consists of DNA/GCN4 complexes, the AMBER DNA/protein force fields was implemented in CONGEN and used to model the GCN4 dimer in the static background of the

$P2_12_12_1$ crystal. The accuracy of this model was comparable to that obtained with the C_2 crystal environment.

MATERIALS AND METHODS

Construction of helical backbones

In constructing a supercoiled backbone, we start with the Pauling-Corey α-helical backbone structure by assigning the backbone torsion angles $\phi = -57°$ and $\psi = -47°$, which yields a helix of 3.6 residues/turn. We then gradually twist the α-helical backbone to a pitch of ~ 3.5 residues/turn. At each step of twisting, the torsion angle change is 0.5° followed by 100 cycles of Adopted Basis Newton-Raphson minimization (ABNR). We next supercoil the α-helical backbone to reach a supercoil pitch value, P, by using the matrix transformation

$$\begin{pmatrix} x_h \\ y_h \\ z_h \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_h \\ y_h \\ z_h \end{pmatrix} + \begin{pmatrix} R\cos\theta \\ R\sin\theta \\ 0 \end{pmatrix}$$

where R is the supercoil radius and $\theta = 2\pi z/P$. In this paper, we use R = 5.2 Å. The supercoiling is accomplished by taking a number of partial transformation steps; often 50 cycles of ABNR minimization are performed (Fig. 3). The Crick's supercoil pitch parameter p = 186 Å/turn was adopted for the GCN4 dimer in the previous study (Krystek et al., 1991). Recently, the pitch value was found to be diverse for different α-helical coiled coils. Therefore, in this work, we sampled the pitch value for the GCN4 leucine zipper dimer to find the best value.

Dimer generation

To form a dimer, we assume that the two GCN4 helices possess the same modeled backbone structure, as described above. The schematic illustration of the dimer formation process is shown in Fig. 4. We first rotate the supercoiled α-helical backbone along its helical axis, parallel to the z axis, by an ω angle, and then duplicate the backbone structure to generate atomic coordinates for the second α-helix. Next, the second α-helix is translated along the x axis followed by 180° rotation along its helical axis. The distance between the two helical axes is kept at twice the Crick's radius parameter. The relative orientation of the two helices is determined by the rotation angle ω. It is known that the interface of a leucine zipper dimer is formed by the residues in **a** and **d** positions. Nevertheless, because the

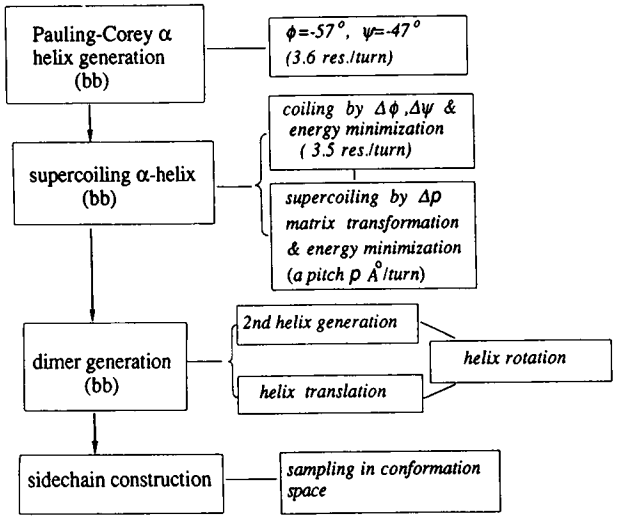

FIGURE 3 Model-building flow-chart for coiled-coil dimers.

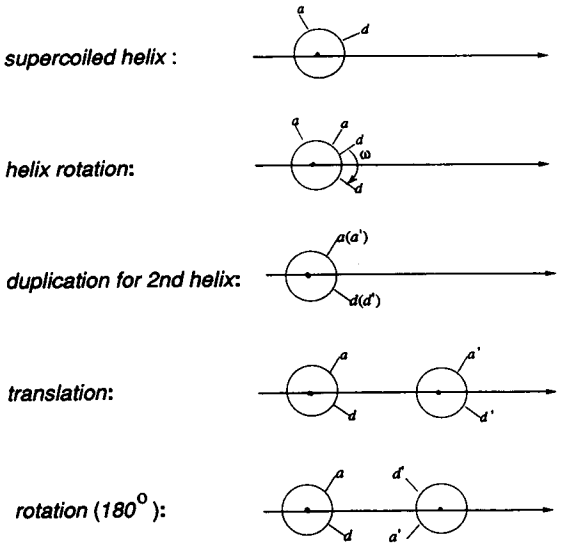

FIGURE 4 Schematic diagram of α-helical dimer formation.

initial positions of the side chains for the constructed α-helical backbones are not precisely known, we need to examine various rotation angles for the best modeling structure.

Side-chain placement

The placement of the side chains on the backbones is accomplished by sampling their torsion conformational space (Bruccoleri and Karplus, 1987). Currently there are five conformational search protocols available in CONGEN. However, only one of them, the iterative protocol, has been frequently used. In this work, we also utilize another, the combinational protocol. In both protocols, the conformations with bad van der Waals contacts are excluded. The iterative protocol starts with an energetically acceptable conformation for all of the side chains. It then regenerates all possible positions for the atoms of one side chain in the presence of other side chains, so that their effect can be included in the side-chain constructions. The lowest energy conformation for this side chain is selected. This process is repeated sequentially for the other side chains. After all side chains are constructed, one returns to the first constructed side chain and goes through the process again until energy is converged or the iterative limit is reached. Alternatively, the combinational protocol first generates a small number of the best side-chain conformations for each side chain independently. These side-chain conformations are then assembled in all possible combinations that do not have bad van der Waals contacts. The lowest energy conformation is finally selected. In addition to the search protocols, we also examine different angular search grids to understand their effect on modeling accuracy.

Model building in a crystalline environment

A part of our study involved construction of the GCN4 dimer side chains in the passive environment of the respective crystal lattices (either C_2, O'Shea et al., 1991; or $P2_12_12_1$, Ellenberger et al., 1992); that is, the CONGEN side-chain conformational searches were carried out in the presence of the surrounding GCN4 molecules and the crystallographically resolved water molecules, wherever those were sufficiently well resolved and bound to the protein. Intermolecular contact-accessible surfaces were calculated first, using the CONGEN atomic table analysis facility. The contact data were used to identify peptides in intermolecular van der Waals contacts and to construct a system where the "central" GCN4 molecule, with side chains to be modeled by CONGEN conformational searches, was

completely surrounded by its nearest-neighbor GCN4 and solvent molecules. For example, in the C_2 crystal lattice (O'Shea et al., 1991) the total of five symmetry-related GCN4 dimers and their bonded water molecules were included into the system. Exhaustive conformational searches were then carried out on all of the conformational degrees of freedom as described above (i.e., employing the angular grid of 30° and the steric repulsion; van der Waals avoidance threshold set to 50 kcal). The iterative side-chain placement protocol was used. Energetics of all the side-chain assemblies generated in the course of the search was evaluated using the CONGEN molecular mechanics potential that implicitly incorporated solvent effects (i.e., the dielectric constant set to $4r$), and all the inter- and intramolecular pairwise interactions (the distance cutoff set to 97 Å, i.e., exceeding the dimensions of the system).

RESULTS AND DISCUSSION

As before (Bruccoleri et al., 1986; Krystek et al., 1991), the structure of the GCN4 dimer was modeled starting from its sequence in a succession of the following steps (Fig. 3). 1) Ideal α-helical backbones in the Pauling and Corey (1953) geometry were supercoiled by a series of gentle twisting operations, each followed by energy minimization. 2) The backbones of the supercoiled dimer were generated by copying coordinates of the supercoiled α-helix and applying translational/rotational matrix transformations to achieve a predefined helix-helix skew angle (Figs. 4 and 5). 3) Side chains were constructed on the backbones using CONGEN conformational searches.

In each of these construction steps, important parameters of the protocol were varied, and the results obtained are described in the following sections. In general, the quality of the final or partial models was evaluated by calculating the r.m.s. shift from an x-ray crystallographic structure. We chose the GCN4 crystal structure of O'Shea et al. (1991) as a reference because of its high nominal resolution (1.8 Å) and its simplicity, that is, an isolated GCN4 leucine zipper supercoiled dimer only, compared to the crystal structure (normal resolution 2.9 Å) of Ellenberger et al. (1992), which contains the complete GCN4 basic region plus leucine zipper dimer in complex with the DNA. However, frequent comparisons with the Ellenberger structure are also reported.

Interhelical skew angle

In building the GCN4 dimer **a**, **d** interface, and without any prior knowledge of the structure, it becomes necessary to examine variations of helix-helix skew angle (θ, cf. Fig. 5; Harris et al., 1994) to find an angle that yields the "best" model. Table 1 compares r.m.s. shifts between the reference x-ray structure and a set of models with various helix-helix skew angles. In constructing these models, the superpitch value was fixed at 186 Å, whereas the helical backbone dimer was constructed using the predetermined skew angle values. Side chains were then constructed onto the backbones via conformational searches as described in Materials and Methods. Clearly, $\theta = 23°$ gives the best results (r.m.s. 1.3 Å on the side chains in the a, d positions, 2.8 Å on all atoms). Skew angles diverging from $\theta = 23°$ by as little as 15° gave unsatisfactory results. In what follows, the skew angle of 23° has been used throughout in the dimer backbone construction.

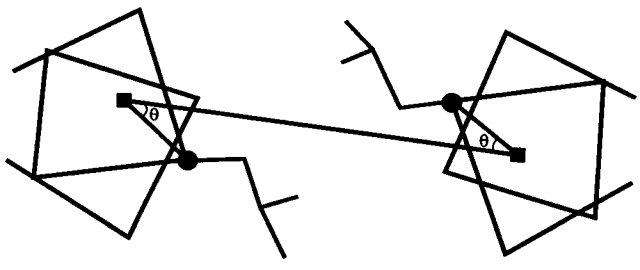

FIGURE 5 Definition of the skew angle, θ, in an α-helical dimer (Harris et al., 1994).

Supercoil pitch

Table 2 reports the r.m.s. shifts between the x-ray structure and the models constructed using different supercoil pitches (keeping the skew angle of 23°). The superpitch value of 147.2 Å gave the best results (backbone r.m.s. 0.8 Å, all-atom r.m.s. 2.5 Å), consistent with the actual superpitch of the GCN4 sequence. Interestingly, the accuracy of positioning the interface Leu and Val residues did not seem to be very sensitive to the change of supercoil pitch parameter (cf. Table 2).

Side-chain search protocols

There are several side-chain placement protocols in CONGEN (Bruccoleri and Karplus, 1987). Although all of them use uniform sampling of side-chain conformational space to generate stereochemically acceptable protein conformers, they differ in the way the "best" conformations are generated and selected. The two search protocols employed and compared here were ITER and COMBINATION. In both protocols, the criterion for selection of the individual, "best" side-chain conformation was the sum of the calculated covalent and noncovalent energies of that side chain (see Materials and Methods for details). The ITER protocol iteratively sampled all of the side-chain rotatable torsion in a predefined order (in this case, from N to C terminus) until

TABLE 1 Relative helix-helix orientation and its influence on the quality of the final model

Rotation angle (ω)*	Skew angle (θ)*	r.m.s. shift (Å)‡		
		Val and Leu	Backbone	Total
−120°	171.5°	6.6	4.4	7.0
30°	36°	4.2	2.5	4.1
80°	4.5°	2.1	1.2	22.9
95°	23°	1.3	1.1	2.8
120°	42.5°	1.9	1.3	3.1
140°	79°	2.9	1.8	3.6

*See Figs. 4 and 5 for an explanation of the angles ω and θ.
‡With respect to the x-ray GCN4 dimer structure (O'Shea *et al.*, 1991). The angular grid of CONGEN side-chain conformational search was 60° for the Lys and Arg residues and 30° for the rest of the residues.

TABLE 2 Supercoil pitch of the helix backbone and its influence on the quality of final model*

Supercoil pitch (Å/turn)	r.m.s. shift (Å)‡		
	Backbone	Val and Leu	Total
205.7	1.3	1.4	3.1
186.0§	1.1	1.3	2.8
164.8	0.9	1.3	2.6
147.2	0.8	1.3	2.5
127.0	0.8	1.3	2.6

*The other parameters of this modeling experiment: the skew angles of the two helices A and B, separated by 10.4 Å, are of $\theta = 23°$, and the GCN4 side chains constructed by the iterative protocol with search angular grid of 60° for the Lys and Arg residues and 30° for the rest of the residues.
‡With respect to the x-ray GCN4 dimer structure (O'Shea *et al.*, 1991).
§The "canonical" coiled-coiled pitch given by Crick (1953).

TABLE 4 Angular grid of CONGEN side-chain conformational search and its influence on the quality of final model*

Angular grid	r.m.s. shift (Å)‡		
	Arg and Lys	His and Tyr	Total
120°	4.3	3.5	2.8
(30°, 60°)§	3.8	1.4	2.5
30°	4.1	2.2	2.6

*The other parameters of this modeling experiment: the same backbone modeled of the skew angles of the two helices A and B, separated by 10.4 Å, and of $\theta = 23°$, the supercoil pitch of 147.2 Å/turn; the iterative protocol for the side-chain construction with $\epsilon = 4r$.
‡With respect to the x-ray GCN4 dimer structure (O'Shea *et al.*, 1991).
§The value of 30° applied to all side-chain torsional angles except for Lys and Arg side chains, which were searched, for convenience, on a 60° grid.

the energy of the system stopped changing. The result of the ITER sampling process was the single "best" conformer of the whole structure. In the COMBINATION protocol, on the other hand, two turns of each helix were sampled at a time (i.e., 14 side chains total), and all of the stereochemically acceptable conformations were retained. The final "best" structure was the lowest energy conformer of the set of all the possible combinations of the partial two-turn searches. As Table 3 shows, the two procedures yielded essentially identical results.

Angular search grid of side-chain torsional sampling

The angular grid on which each side-chain torsion is sampled is an important factor potentially influencing the modeling accuracy (Bruccoleri and Karplus, 1987). In general, the finer the grid the greater the accuracy of the model. In practice, limits of time and computer resources dictate the fineness of the grid. Table 4 reports the results of search grid variations. The general tendency of the results is as expected: a finer search grid tends to produce more accurate models than a larger search grid. However, the above tendency is not absolute. Thus, values with a grid of 60° for the Lys and Arg side chains and of 30° for the rest of the side chains are smaller than those with grid = 30° for all of the side chains. Comparing the second and third columns in Table 4, one notices that the change of the angular grid for Arg and Lys also affects results obtained for His and Tyr.

TABLE 3 Side-chain search protocols: comparison of the final model to the x-ray*

	r.m.s. shift (Å)			
	Case I ($\epsilon = 4r$, grid = 30°)		Case II ($\epsilon = 80$, grid = 120°)	
	Iteration	Combination	Iteration	Combination
Total	2.6	2.9	2.8	2.6
His and Tyr	1.2	3.5	2.7	1.5

*O'Shea *et al.*, 1991.
The modeled backbones are the same with supercoil pitch = 147.2 Å/turn and the skew angle $\theta = 23°$.

Dielectric constant and energy scaling factors

The form of potential energy function used to evaluate the generated conformations is perhaps the single most critical determinant of successful modeling. It has been known that in vacuo potentials poorly discriminate between correct and incorrect models, whereas modified potentials that incorporate solvent effects through surface-dependent terms and effective dielectric constants have better discriminatory power (Novotny et al., 1984; Bruccoleri et al., 1988). In this work we investigated several different forms of the effective dielectric constant to approximate solvent screening effects, e.g., $\epsilon = r$, $\epsilon = 4r$ and $\epsilon = 80$. The in vacuo Coulomb formula, $\epsilon = 1$, was also used for comparison. Table 5 shows that, in our case, the overall modeling results do not seem to be much influenced by the exact forms of the effective dielectric constants used. Nevertheless, for individual residue types the modeling accuracy varies, and the formally charged residues are, as a rule, considerably affected. For the Arg and Lys residues $\epsilon = 4r$ gave the best r.m.s. values, whereas for the His and Tyr residues $\epsilon = 1$ generated the best result.

TABLE 5 Dielectric constant and its influence on the quality of final model*

Dielectric constant (ϵ)	r.m.s. shift (Å)‡		
	Arg and Lys	His and Tyr	Total
4r	3.8	1.4	2.5
r	4.0	2.2	2.6
80	4.2	1.5	2.6
1	4.1	1.2	2.6

*The other parameters of this modeling experiment: the same backbone modeled of the skew angles of the two helices A and B, separated by 10.4 Å, and of $\theta = 23°$; the supercoil pitch of 147.2 Å/turn; the iterative protocol for the side chains construction with the angular search grid of 60° for the Lys and Arg residues and 30° for the rest of the residues.
‡With respect to the x-ray GCN4 dimer structure (O'Shea *et al.*, 1991).

TABLE 6 Modeled backbone versus x-ray backbone: comparison of the final model to the x-ray*

	r.m.s. shift (Å)‡		
	Total	Arg and Lys	His and Tyr
Modeled backbone	2.6	4.1	2.2
X-ray backbone	2.0	2.7	2.9

*Iterative protocol for the side chains construction with the angular search grid of 30° and $\epsilon = 4r$.
‡With respect to the x-ray GCN4 dimer structure (O'Shea *et al.*, 1991).

We also experimented with adjusting the individual potential energy terms by arbitrary scaling. This included i) setting all of the covalent terms to zero; ii) turning off electrostatic interactions; and iii) scaling the van der Waals interactions by a factor of 10. These experiments did not give consistent or satisfactory results and are not being reported.

Modeled backbone versus x-ray backbone in side-chain construction

As shown in Table 2, the GCN4 polypeptide backbone can be modeled to agree with the x-ray structure to within 0.8 Å. The accuracy of side-chain modeling, however, has not exceeded 2.5 Å r.m.s. difference. To further explore the limits of side-chain modeling, we used the x-ray crystallographic backbone for side-chain construction and compared the results to those obtained with the de novo modeled backbone (Table 6). Overall, there was a significant improvement in side-chain placement when the x-ray backbone was used. The improvement was particularly pronounced for the Arg and Lys residues (r.m.s. 1.4 Å), whereas for the His and Tyr residues the r.m.s. actually became 0.7 Å larger.

Crystal lattice contacts

So far we have been concerned with modeling an isolated GCN4 dimer. However, in the crystal structure of O'Shea et al. (1991), about 40% of each dimer surface is in tight contact with the other dimer molecules of the crystal lattice. Crystal contacts of the Ellenberger et al. (1992) structure occlude about 20% of the dimer surface, with about 6% of the intermolecular contact area being due to DNA-protein contacts. Altogether, 2049 Å^2 was buried in the surface of a single GCN4 dimer within the C_2 crystal lattice, about half of the contacts involving nonpolar atoms, mostly carbons (1084 Å^2 or 53%). In the $P2_12_12_1$ crystal lattice, on the other hand, 906 Å^2 of the single GCN4 dimer was buried, with polar contacts predominating (509 Å^2 or 56% polar atoms involved in the contacts). A series of modeling experiments was therefore run where the C_2 symmetry-related GCN4 dimers and crystallographically resolved water molecules were included as a passive background. A marked improvement in modeling accuracy ensued. In Table 7 we compare the result of the GCN4 side-chain construction in the crystal with the best result of the isolated dimer, using the x-ray backbone to support the side-chain construction in both cases. The all-atom r.m.s. of the run with crystal environment was 1.7 Å. It is particularly interesting to see the marked improvement in the placement of the helix-helix interface residues, Leu and Val, and the His and Tyr rings.

To obtain a broader perspective on crystal lattice effects, CONGEN side-chain construction runs were also carried out using the two different crystal environments (the C_2 and $P2_12_12_1$) and the AMBER force-field. The AMBER parameters were implemented in CONGEN to allow us a simultaneous construction of both protein and DNA in the $P2_12_12_1$ crystal (the DNA atomic parameters were not available in the CONGEN force field). In both of these runs, the total side-chain r.m.s. was comparable (2.2 Å for the C_2 structure, 2.5 Å for the $P2_12_12_1$ structure), and similar trends were observed in systematic deviations of side-chain groups such as hydrophobic, charged, and aromatic rings (Table 7, Figs. 6 and 7). Overall, the intermolecular interactions were dominated by the van der Waals contacts and steric (excluded volume) constraints, as evidenced by the calculated total electrostatic and Lennard-Jones energy terms: 1.1 kcal/mol (electrostatic) and 85.1 kcal/mol (van der Waals) in the C_2 crystal environment and 19.1 kcal/mol (electrostatic) and 29.4 kcal/mol (van der Waals) in the $P2_12_12_1$ environment. It is also interesting to compare the conformations of the side-chain pairs Leu 6, Leu 12, Leu 19, Leu 26, and Asn 16, respectively, in the **a** and **d** positions of the dimer interface (Table 8 and Figs. 6 and 7). By and large, the conformations of Leu/Leu pairs of the two α-helices in the dimer were symmetrical, with a clear preference for the *-g*, *-g* rotamer. The Asn 16 side-chain pairs, how-

TABLE 7 Crystal packing effect: comparisons of various runs

	r.m.s. shift (Å)				
	Total	Val and Leu	Asp and Glu	His and Tyr	Arg and Lys
Isolated dimer (CONGEN f.f.)	2.0	1.0	1.7	2.9	2.7
Isolated dimer (AMBER f.f.)	2.5	0.7	1.5	2.9	3.7
Dimer in C_2 (CONGEN f.f.)	1.7	0.5	1.1	1.4	2.6
Dimer in C_2 (AMBER f.f.)	2.3	0.8	1.7	2.1	3.5
Dimer in $P2_12_12_1$ (AMBER f.f.)	2.5	0.9	1.4	2.1	3.8

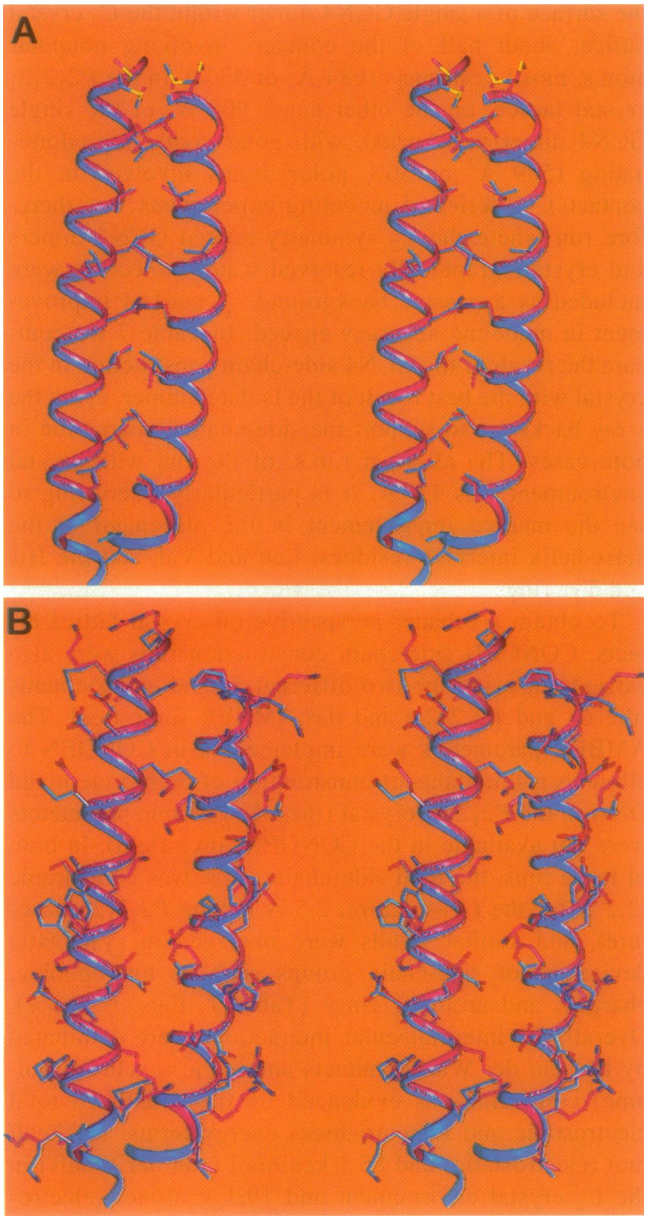

FIGURE 6 Stereo diagram of the least-squares superimposed GCN4 leucine zipper O'Shea et al. (1991) x-ray crystallographic structure (*in red*) and the model generated in the C_2 crystal lattice environment (AMBER force-field) (*in blue*). α-Helical backbones are shown as ribbons, and the central hydrophobic residues (and the Asn 16) in the a and d positions are displayed as sticks. The overall r.m.s. of this superposition was 2.2 Å, 0.7 Å on the interface Leu and Val side chains. (*A*) Interface side chains; (*B*) surface side chains.

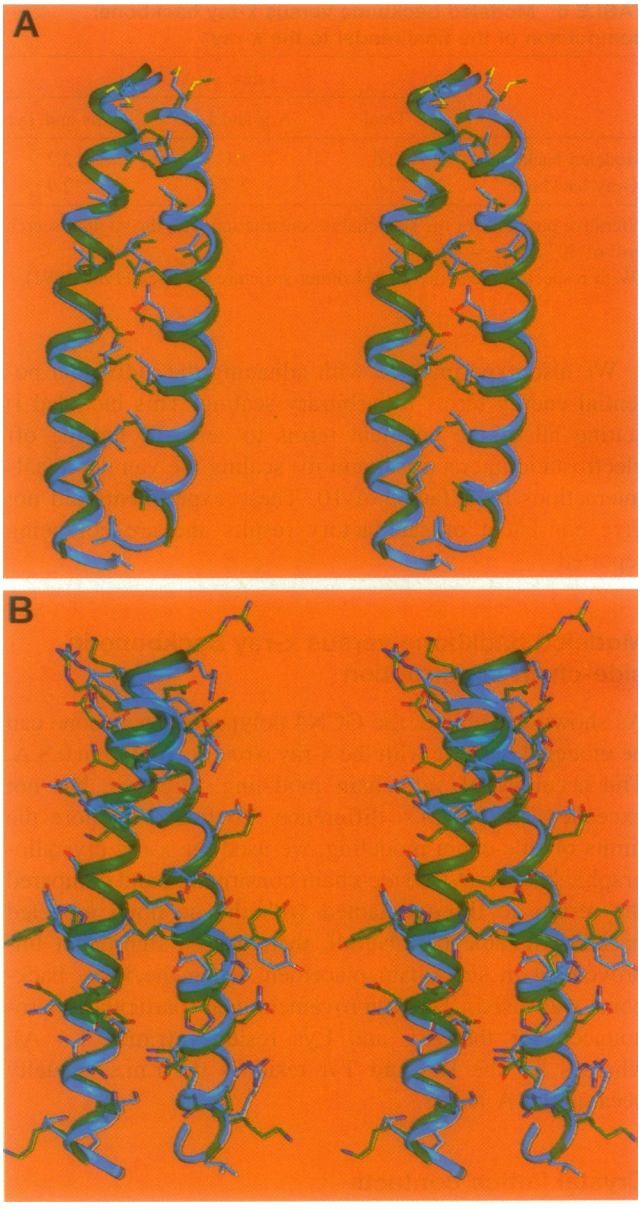

FIGURE 7 Stereo diagram of the least-squares superimposed GCN4 leucine zipper Ellenberger et al. (1992) x-ray crystallographic structure (*in green*) and the model generated in the $P2_12_12_1$ crystal lattice environment (AMBER force-field) (*in blue*). α-Helical backbones are shown as ribbons, and the central hydrophobic residues (and the Asn 16) in the a and d positions are displayed as sticks. The overall r.m.s. of this superposition was 2.5 Å, 0.9 Å on the interface Leu and Val side chains. (*A*) Interface side chains; (*B*) surface side chains.

ever, showed both an asymmetry and a lack of strong conformational preference, although the *-g*, *-g* rotamer did occur more often than the others (cf. Table 8).

Figs. 8 and 9 provide a more detailed analysis of the results obtained in the side-chain construction runs in the C_2 crystal environment with the CONGEN force field. Fig. 8 shows the residue r.m.s. values for the isolated dimer (*dotted line*) and for the dimer in the crystal (*solid line*). Among the 62 residues of the GCN4 leucine zipper dimer, 37 display r.m.s shifts from the O'Shea et al. (1991) structure below 1.0 Å. Seventeen additional residues have their r.m.s. between 1.0 Å and 2.0 Å, and only eight residues show a r.m.s. of 2.0 Å or more. In the isolated dimer, on the other hand, only 20 residues had r.m.s. values below 1.0 Å, 24 residues between 1.0 Å and 2.0 Å, and 17 residues above 2.0 Å. In Fig. 9 we show the GCN4 dimer surface color-coded by r.m.s. values on the one hand, and crystal contacts on the other hand, thus

TABLE 8 The side-chain conformations at the helix-helix contact positions of GCN4 models constructed within the C_2 and $P2_12_12_1$ crystal environment

	Dimer in C_2		Dimer in $P2_12_12_1$	
	Segment A	Segment B	Segment A	Segment B
Leu 5				
χ_1	$-g$	t	$-g$	t
χ_2	$-g$	$+g$	$-g$	$-g$
Leu 12				
χ_1	$-g$	$-g$	$-g$	$-g$
χ_2	$-g$	$-g$	$-g$	$-g$
ASN 16				
χ_1	$-g$	t	$+g$	$-g$
χ_2	$-g$	$-g$	$+g$	$-g$
Leu 19				
χ_1	$-g$	$-g$	$-g$	$-g$
χ_2	$-g$	$-g$	$-g$	$-g$
Leu 26				
χ_1	$-g$	$-g$	$-g$	$-g$
χ_2	$-g$	$-g$	$-g$	$-g$

The letters g and t denote the *gauche* and *trans* torsional ranges ($\pm g$ = $\pm 60° \pm 30°$, t = $180° \pm 30°$.

making the correlation between the two quantities graphically obvious. It can be seen that most of the side chains that were difficult to model correctly in the isolated dimer were those involved in extensive crystal lattice contacts.

CONCLUSIONS

We carried out a series of studies to explore the key factors that influence the modeling accuracy for the GCN4 leucine zipper dimer. The basic conclusions are as follows. 1) In backbone construction, the experiments exploring various supercoil pitch values showed the best r.m.s. agreement with the GCN4 x-ray structure (r.m.s. 0.8 Å) for the pitch 147.2 Å/turn. This value differs from the canonical supercoil pitch of Crick, 186 Å/turn, but is close to the one measured by Seo and Cohen (1993) on selected x-ray crystallographic structures. 2) In dimer generation, we found that the variation of the helix skew angle θ has a strong influence on modeling accuracy. 3) In side-chain construction, comparable results were obtained with the CONGEN conformational searches where all side-chain torsional angles were sampled globally (iteration protocol), or where shorter segments (two helical turns) were built first and combined later. 4) In energetic evaluation of the generated conformations, we showed that the use, in the Coulomb equation, of the

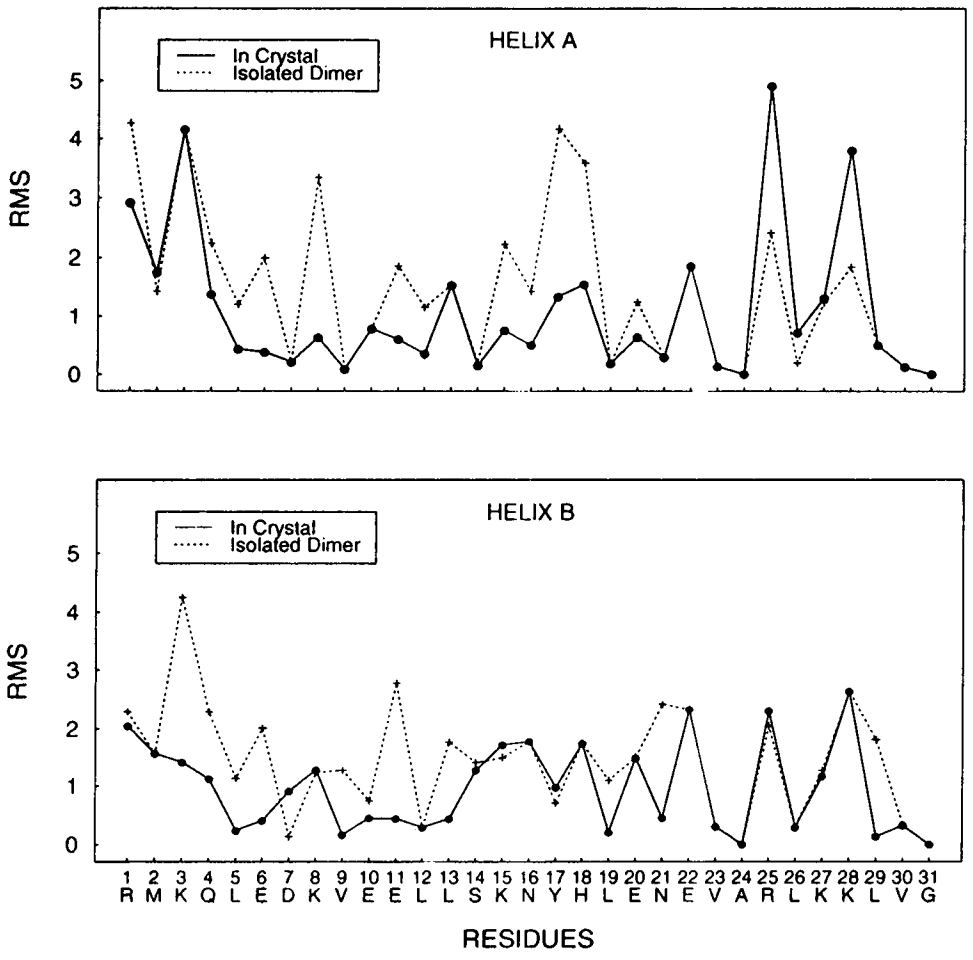

FIGURE 8 Root-mean-square (RMS) shifts of side chains in the crystallographic (C_2) and modeled (CONGEN force-field) GCN4 leucine zippers. The dotted line is the result modeled for the isolated GCN4 dimer, and the solid line is the result for the dimer modeled in the C_2 crystal environment.

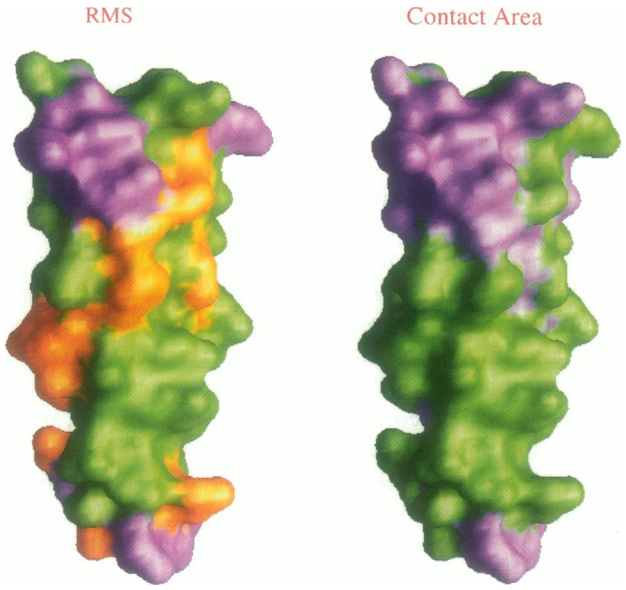

FIGURE 9 Surface representation of the CONGEN-modeled GCN4 leucine zipper dimer in the crystal (C_2) environment. On the left, the surface color-coded by r.m.s. values; on the right, the surface color-coded by intermolecular contact areas. In green the residual r.m.s. is less than 1.0 Å (*left*) and the residual crystal contact area larger than 40 Å^2 (*right*); in purple the residual r.m.s. is larger than 2.0 Å (*left*) and the crystal contact area less than 40.0 Å^2 (*right*); in brown the residual r.m.s. is between 1.0 Å and 2.0 Å (*left*). The picture was generated with the program GRASP (Nicholls et al., 1991).

effective constant $\epsilon = 80$ or $\epsilon = 4r$ gave comparable results. Finally, 5) we demonstrated that the crystal packing effects are of overwhelming importance in determining individual side-chain positions. When the crystal environment was included in the GCN4 dimer construction as a passive background, the r.m.s. agreement between the model and the x-ray structure was significantly improved (total r.m.s. = 1.7 Å).

REFERENCES

Bruccoleri, R. E. 1993. Application of systematic conformational search to protein modeling. *Mol. Simulation.* 10:151–174.

Bruccoleri, R. E., E. Haber, and J. Novotny. 1988. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature.* 335:564–568.

Bruccoleri, R. E., and M. Karplus. 1987. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers.* 26:137–168.

Bruccoleri, R. E., J. Novotny, P. Keck, and C. Cohen. 1986. Two-stranded α-helical coiled-coils of fibrous proteins. *Biophys. J.* 49:79–81.

Crick, F. H. C. 1953. The Fourier transform of a coiled-coil. *Acta Crystallogr.* 6:689–697.

DeLano, W. L., and A. T. Brunger. 1994. Helix packing in proteins: prediction and energetic analysis of dimeric, trimeric and tetrameric GCN4 coiled coil structures. *Proteins.* 20:105–123.

Ellenberger, T. E., C. J. Brandl, K. Struhl, and S. C. Harrison. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein-DNA complex. *Cell.* 71:1223–1237.

Harris, N. L., S. R. Presnell, and F. E. Cohen. 1994. Four helix bundle diversity in globular proteins. *J. Mol. Biol.* 236:1356–1368.

Junius F. K., J. P. Mackay, W. A. Bubb, S. A. Jensen, A. S. Weiss, and G. F. King. 1995. Nuclear magnetic resonance characterization of the Jun leucine zipper domain: unusual properties of coiled-coil interfacial polar residues. *Biochemistry.* 34:6164–6174.

Kouzarides, T., and E. Ziff. 1988. The role of the leucine zipper in the Fos-Jun interaction. *Nature.* 336:646–651.

Krystek, S. R., R. E. Bruccoleri, and J. Novotny. 1991. Stability of leucine zipper dimers estimated by an empirical free energy method. *Int. J. Pept. Protein Res.* 38:229–236.

Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science.* 240:1759–1764.

Marmorstein, R., M. Carey, M. Ptashne, and S. C. Harrison. 1992. DNA recognition by GCN4: structure of a protein-DNA complex. *Nature.* 356:408–414.

McLachlan, A. D., and M. Stewart. 1975. Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. Mol. Biol.* 98: 293–304.

Nicholls, A., K. A. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins.* 11:281–296.

Nilges, M., and A. T. Brunger. 1991. Automated modeling of coiled coils. Application to the GCN4 dimerization region. *Protein Eng.* 4:649–659.

Nilges, M., and A. T. Brunger. 1993. Successful prediction of the coiled coil geometry of the GCN4 leucine zipper domain by simulated annealing: comparison to the x-ray structure. *Proteins.* 15:133–146.

Novotny, J., R. E. Bruccoleri, and M. Karplus. 1984. An analysis of incorrectly folded protein models. *J. Mol. Biol.* 177:787–818.

Oas, T. G., L. P. McIntosh, E. K. O'Shea, F. W. Dalquist, and P. S. Kim. 1990. Secondary structure of a leucine zipper determined by nuclear magnetic resonance spectroscopy. *Biochemistry.* 29:2891–2894.

O'Shea, E. K., J. D. Klemm, P. S. Kim, and T. Alber. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science.* 254:539–544.

O'Shea, E. K., R. Rutkowski, and P. S. Kim. 1989. Evidence that the leucine zipper is a coiled coil. *Science.* 243:538–542.

Pauling, L., R. B. Corey, and H. R. Branson. 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA.* 37:205–211.

Schulz, G. E., and R. H. Schirmer. 1979. Coiled-coil α-helix. *In* Principles of Protein Structure. C. R. Cantor, editor. Springer-Verlag, New York. 79–81.

Seo, J., and C. Cohen. 1993. Pitch diversity in α-helical coiled coils. *Proteins.* 15:223–234.

Vieth, M., A. Kolinski, C. L. Brooks III, and J. Skolnick. 1994. Prediction of the folding pathways and structure of the GCN4 leucine zipper. *J. Mol. Biol.* 237:361–367.

Zhang, L., and J. Hermans. 1993. Molecular dynamics study of structure and stability of a model coiled coil. *Proteins.* 16:384–392.