Automated modeling of coiled coils: application to the GCN4 dimerization region

Michael Nilges and Axel T. Briinger

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

A novel approach for the modeling of coiled coils through molecular dynamics is described and applied to the dimerization region of the yeast transcriptional activator GCN4. Initially, a model is created consisting of C° atoms only, representing an idealized coiled coil with infinite pitch. Human bias in the placing of the other atoms is reduced by an automatic building procedure using simulated annealing with simple geometric restraints. The resulting all-atom model is then allowed to relax during a short molecular dynamics run using an empirical energy function and weak restraints which reflect the coiled coil assumption. These models are then further refined using unrestrained molecular dynamics in water. In this report we test the model-building procedure on the known dimerization region of catabolite gene activator protein (CAP), part of which forms a coiled coil, and we predict the structure of the coiled coil dimerization region (the ‘leucine zipper’ domain) of GCN4. Several models are built, starting from different arrangements of the C° atoms in the initial structures. The final structures show similar crossing angles of the coiled coil, although this was not used as a restraint in the calculation. The leucines adopt a ladder-like conformation around the 2-fold axis of the coiled coil. A number of electrostatic interactions can be identified which may contribute to the stability of the helical structure of the monomers and of the dimer.

Key words: coiled coil/GCN4/leucine zipper/model building/molecular dynamics

Introduction

Coiled coils are ubiquitous structural motifs in proteins (Crick, 1952, 1953). Two or more helices in parallel (in this context means ‘same sense’) or antiparallel orientation can form a coiled coil by winding the helices around each other. Optimal packing is achieved by fitting the ‘knobs’ of one helix into the ‘holes’ of the other. Coiled coils are found in a number of fibrous proteins, e.g. keratin (Fraser and MacRae, 1973) and tropomyosin (Hodges et al., 1972), and globular proteins, e.g. catabolite gene activator protein (CAP) (Weber and Steitz, 1987), and serine amino acid synthetase (Cusack et al., 1990). Coiled coils are the proposed structure for the membrane-spanning helices of proteins, e.g. glycoplin and the EGF receptor (Welsh et al., 1985), and are the structure of the dimerization regions of a class of DNA-binding proteins, the ‘leucine zippers’ (Landschulz et al., 1988; Rasmussen et al., 1991). The occurrence of coiled coils in dimerization interfaces could be of functional importance, as dimerization is a prerequisite for DNA binding in the case of the leucine zippers, and signal transduction in the case of the EGF receptor. Furthermore, the dimerization can be sequence specific, as has been shown for the Fos-Jun leucine zipper (O’Shea et al., 1989a,b). Thus, coiled coils provide model systems for protein–protein recognition studies.

The simplicity of the coiled coil motif has already spawned modeling studies of coiled coils, using energy minimization of the idealized coiled coil (Parry and Suzuki, 1969; McLachlan, 1978). The mathematical description of an ideal coiled coil by Crick (1952), which was the basis of these studies, provides a low-resolution picture. However, a detailed high-resolution structure is required to address questions about the structural basis for stability, and sequence specificity of the coiled coil interaction. While X-ray crystallography and solution NMR spectroscopy are providing an increasing data base of coiled coil structures, the time may be ripe to attempt ab initio prediction. For this purpose, computer graphics modeling has a number of drawbacks. First, human bias in the building process is inevitable. Second, detailed considerations of chemical interactions are usually not possible and protein–solvent interactions are neglected. Third, computer graphics modeling is time-consuming and thus is difficult to sample a large number of possible conformations.

In this report, we propose an automated model building procedure where the models are generated by using molecular dynamics techniques with minimal human intervention. Assumptions and experimental knowledge about the modeled structure enter the calculation in the form of simple restraints. This makes it possible to clearly document the assumptions one makes, and keep them to a minimum. The automation of the building process enables one to repeat it often with varied initial conditions. In this way, one can obtain a representative sampling of the conformational space consistent with the constraints.

We have tested our automated building procedure on the coiled coil dimerization interface of catabolite gene activator protein (CAP) (Weber and Steitz, 1987). We have then predicted a family of possible structures for the leucine zipper domain of the yeast transcriptional activator GCN4 (Jones and Fink, 1982; Hinnebusch and Fink, 1983), for which a high resolution X-ray structure will become available shortly (Rasmussen et al., 1991). We discuss the conformation of this family of structures. Furthermore, we present evidence for the electrostatic interactions which may contribute to the stability of each monomer, and of the dimer in solution.

Materials and methods

The automated model-building procedure consists of four stages (Table I). First, a C° model of the dimer is built in an idealized conformation consisting of two parallel helices. This initial C° model corresponds to a coiled coil with infinite pitch. Second, the remaining atoms are built in an automated way, starting with the C° atom positions. The third stage is a refinement of the whole model by restrained molecular dynamics and minimization. These first three stages are carried out without explicit solvent, using only a crude approximation of solvent screening. The final stage is an unrestrained molecular dynamics simulation with explicit inclusion of solvent. While it is assumed that the
modeled folding motif is a coiled coil, the parameters of the coiled coil, such as crossing angle and separation, are left self-adjusting during all stages of the building procedure. By repeating the procedure with different initial conditions, a family of structures can be produced.

**Initial C° model**

The initial C° model consists of two parallel helices, with a separation between the two helix axes of 10.4 Å, which is the separation found in fiber diffraction studies of coiled coils (Fraser and MacRae, 1973). The helix rise is 1.5 Å, the twist is 102.9° (3.5 residues per turn), and thus the pitch of the coiled coil is infinite. [The pitch \( P \) (in Å), the difference \( \Delta t \) (in rad) between actual twist of the helices and 3.5 residues per turn, the helix rise \( h \) (in Å), and the separation \( r_0 \) (in Å) of the two helices are related by the equation

\[
P = \left( \frac{2\pi}{\Delta t} \right) \left[ h^2 - (r_0 \Delta t)^2 \right]^{\frac{1}{2}}
\]

(Fraser and MacRae, 1973). For \( \Delta t = 0 \) rad, we obtain \( P = \infty \) Å.] The helices are slightly overwound by \(-2.9^\circ\), the normal value found for \( \alpha \)-helices for the twist being 100° (3.6 residues per turn). Thus, every seventh residue is exactly in the same angular position, and in the dimer, the ‘knobs’ in one straight helix fit exactly into the ‘holes’ of the other.

The orientations of the helices are chosen such that the hydrophobic residues (at positions ‘a’ and ‘d’ in Figure 1) are interacting with each other. We define the orientations of the helices by the angle \( \theta \), which is measured as the angle between the C° atom of the ‘d’ residue in one helix, the long axis of the helix, and the connecting line between the two helix axes (see Figure 1). We have varied the initial orientations within a range of \(-\pm 15^\circ\) around the angle \( \theta = 35^\circ \), which maximizes the hydrophobic contact area.

**Building the remaining atoms**

The remaining backbone and side-chain atoms are built by a simulated annealing protocol using packing and stereochemical restraint.

### Table I. Automated model building procedure

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Build C° atoms of coiled coil with infinite pitch: two parallel helices separation = 10.4 Å twist = 3.5 residues/turn</td>
</tr>
<tr>
<td>2</td>
<td>Build remaining atoms using (C° — all) building procedure</td>
</tr>
<tr>
<td>3</td>
<td>Molecular dynamics refinement without explicit water: coiled coil restraint (separation &lt; 10.4 Å) ( \alpha )-helical restraint (b-bond distance &lt; 3.2 Å)</td>
</tr>
<tr>
<td>4</td>
<td>Molecular dynamics refinement with explicit water: no restraints</td>
</tr>
</tbody>
</table>

### Table II. (C° — all) building procedure

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Assign and fix C° atoms</td>
</tr>
<tr>
<td>2</td>
<td>Assign all other coordinates to C° positions</td>
</tr>
<tr>
<td>3</td>
<td>Molecular dynamics refinement without explicit water: coiled coil restraint (separation &lt; 10.4 Å) ( \alpha )-helical restraint (b-bond distance &lt; 3.2 Å)</td>
</tr>
<tr>
<td>4</td>
<td>Molecular dynamics refinement with explicit water: no restraints</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1) Assign and fix C° atoms</td>
</tr>
<tr>
<td>2</td>
<td>2) Assign all other coordinates to C° positions</td>
</tr>
<tr>
<td>3</td>
<td>3) Annealing</td>
</tr>
<tr>
<td>4</td>
<td>4) Minimization</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>( K_{gb} )</th>
<th>( K_{vdw} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0 ( \times 10^6 )</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>10 ( \times 500 )</td>
<td>0.0001 ( \times 0.1 )</td>
</tr>
<tr>
<td>1000 ( \rightarrow 300 )</td>
<td>500</td>
<td>4 [radial reduced](^{1})</td>
</tr>
</tbody>
</table>

| (4) | Minimization | 500 | L-J potential\(^{1}\) |

\(^{1}\)Every 0.1 ps, the energy constant \( K_{gb} \) for the bond-length and bond-angle terms are multiplied by 1.25; those for the chirality and planarity terms by (1.1)^2. \( K_{gb} \) is measured in kcal mol Å^2 for bonds, and in kcal molÅ^2 for the bond angle, chirality and planarity terms. These energy constants are initially adjusted so that the sum of bond and angle energies is approximately equal to the kinetic energy at 100 K. The values for chirality and planarity energy constants are chosen as 0.01 times the values for bonds and angles.

\(^{2}\)Every 0.2 ps, the energy constant \( K_{gb} \) for the bond length and bond angle terms are multiplied by 1.1, those for the chirality and planarity terms by (1.1)^2. The maximum value of the energy constants for bonds and angles is 500 kcal molÅ^2 and 500 kcal molÅ^2 respectively, and 200 kcal molÅ^2 for planarity and chirality restraints.

\(^{3}\)The repulsive nonbonded potential has the form \( E_{vdw} = K_{vdw}([\max(0, (r^{2}d_{min} - R^{2})])^{2} \) where \( R_{vdw} \) is the sum of the two van der Waals radii, calculated from the distance at the minimum of the Lennard–Jones potential, \( s \) is a scale factor, and \( K_{vdw} \) is the energy constant. This is increased by doubling it every 0.2 ps from 0.0001 to 0.003 kcal molÅ^2, and by multiplying it by 1.25 from 0.003 to 0.1 kcal molÅ^2. The scale factor \( s \) is set to 1.

\(^{4}\)The scale factor \( s \) in the repulsive potential is set to 0.8.

\(^{5}\)The temperature is controlled by the heat bath coupling method (Berendsen et al., 1984) with a friction coefficient of 100 ps\(^{-1}\). Every 0.1 ps, the temperature of the heat bath is reduced by 50 K.

\(^{6}\)Standard Lennard–Jones van der Waals potential with PARAM19 parameters (Brooks et al., 1983). No electrostatic interaction is used throughout the side-chain building calculation.

---

M.Nilges and A.T. Brünger
Stage II of the model building procedure has obvious uses in homology modeling, and building of side chains during X-ray refinement once the C<sup>α</sup> positions have been chosen. We have tested it for a number of globular proteins with known crystal structures (data not shown). The success and efficiency of our method is comparable to other side-chain building methods (Novotny et al., 1988; Holm and Sander, 1991; Lee and Sabbin, 1991). The procedure seems to be fairly robust as regards to the precision of the C<sup>α</sup> atom positions (data not shown).

**Preliminary refinement without explicit solvent**

The initial all-atom model is allowed to relax during a short molecular dynamics run with an empirical energy function using the CHARMM PARAM19 parameters (Brooks et al., 1983). A minimal set of restraints is added which reflects our assumptions: the two monomers are helical, and they form a dimer. The latter is achieved by using a weak half-parabolic restraint with an energy constant of 2.5 kcal mol/Å<sup>2</sup> on the local separation between the two helices (Figure 1). This local helix—helix separation is measured in the following way (see Figure 1). First, we compute the geometric center for each set of seven consecutive C<sup>α</sup> atoms. As they form almost exactly two full helix turns, these geometric centers approximately lie on the helix axis. The distances between equivalent geometric centers approximate the local helix—helix separation near the central atom of the seven. The local helix—helix restraint is active only if the center-to-center distance exceeds 10.4 Å. The actual center-to-center distances in the refined structures are well below this value (see below), so that the restraint is normally inactive and serves only to prevent the monomers from moving too far apart in the beginning of the molecular dynamics run. In particular, this restraint leaves the monomers free to rotate around their respective axes, and slide along their axes with respect to each other, and imposes no restraint on the angle between the two helices. To ensure that the monomers stay helical during the calculation, the backbone—backbone hydrogen bonds are strengthened by adding a weak half-parabolic restraint with an energy constant of 2.5 kcal mol/Å<sup>2</sup>, which keeps the hydrogen bond donor—acceptor distance below 3.2 Å.

Initial velocities are assigned at 500 K. The system is then cooled down to 300 K over a period of 4 ps, during which harmonic restraints on the C<sup>α</sup> atoms (Brucoleri and Karplus, 1986), which are applied initially, are slowly removed. This is followed by 5 ps molecular dynamics at 300 K and 1000 steps of conjugate gradient minimization. The non-bonded list is updated whenever an atom has moved by >0.25 Å and the cutoff for the electrostatic interaction is 9.5 Å.

In order to decrease the computational requirements, the calculations in stage III are performed without explicit water; but solvent effects are mimicked by using a dielectric constant of ϵ = R, where R is the distance between two interacting atoms. A switching function (Brooks et al., 1983) is employed between 5 and 9 Å. Judging from the work of Loncharich and Brooks (1989), this seemed to be the best compromise between computational efficiency and quality of the calculation as judged by r.m.s. fluctuations and r.m.s. differences from X-ray crystal structures. To further account for solvent screening, the charges on side chain atoms of lysine, arginine, histidine, aspartate and glutamate residues are scaled by a factor of 0.3.

**Final refinement in explicit solvent**

The preliminary refined model is subjected to molecular dynamics with explicit inclusion of solvent, and with the coiled coil and helical restraints removed. We use the OPLS non-bonded parameters (Jorgensen and Tirado-Rives, 1988) with parameters for bond lengths, bond angles, dihedral and improper dihedral angles from AMBER (Weiner et al., 1984). The cutoff for the non-bonded interaction is 11 Å. The preliminary refined model is immersed in a cylinder of TIP3P (Jorgensen et al., 1983) water molecules. Waters overlapping with protein atoms by >2 Å are removed. The dimensions of the cylinder are chosen such that the minimum distance from any protein atom to the cylinder boundary is 7 Å. The water molecules are prevented from evaporating by a stochastic boundary at the cylinder circumference (Brooks and Karplus, 1982; Brünger et al., 1984; Brooks et al., 1985); periodic boundary conditions are used in the direction of the cylinder axis. One hundred steps conjugate restraints are applied to the C<sup>α</sup> atoms to account for solvent screening, the charges on side chain atoms of lysine, arginine, histidine, aspartate and glutamate residues are scaled by a factor of 0.3.

**Final refinement in explicit solvent**

The preliminary refined model is subjected to molecular dynamics with explicit inclusion of solvent, and with the coiled coil and helical restraints removed. We use the OPLS non-bonded parameters (Jorgensen and Tirado-Rives, 1988) with parameters for bond lengths, bond angles, dihedral and improper dihedral angles from AMBER (Weiner et al., 1984). The cutoff for the non-bonded interaction is 11 Å. The preliminary refined model is immersed in a cylinder of TIP3P (Jorgensen et al., 1983) water molecules. Waters overlapping with protein atoms by >2 Å are removed. The dimensions of the cylinder are chosen such that the minimum distance from any protein atom to the cylinder boundary is 7 Å. The water molecules are prevented from evaporating by a stochastic boundary at the cylinder circumference (Brooks and Karplus, 1982; Brünger et al., 1984; Brooks et al., 1985); periodic boundary conditions are used in the direction of the cylinder axis. One hundred steps conjugate

![Diagram](image)

**Fig. 1.** The orientation of the single helices is described by the angle between the C<sup>α</sup> atom of the 'd' residue in one helix (the zipper-leucine), the long axis of the helix, and the connecting line between the two helix axes. θ is set to 30° for both helices in this figure. The distance between the two helix axes in the dimer is approximated by the distance between the geometric centers of seven equivalent and consecutive C<sup>α</sup> atoms (this is only exact in the case of helices with 3.5 residues per turn). The distance restraint (10.4 Å) is applied between these geometric centers.

**Fig. 2.** Sequences of the modeled coiled coil regions of CAP (Weber and Stenz, 1987) and GCN4 (Jones and Fink, 1982)
gradient minimization are carried out to remove bad non-bonded
counters, followed by 5 ps of molecular dynamics at 300 K with
a time-step of 2 fs. Bond lengths and the geometry of the TIP3P
water are constrained with the SHAKE algorithm (Ryckaert et
al., 1977). The last coordinate set represents a predicted structure
of our automated modeling procedure.

All calculations (building of the initial Cα model and the side
chains, dynamics, minimization and analysis of the results) were
carried out with the program X-PLOR (Brünger, 1990), running
on a Cray YMP and a Convex C220.

Results and discussion

Dimerization domain of CAP

We tested stages I—III of the model building procedure (building
the initial Cα model, building the remaining atoms, and
molecular dynamics without explicit solvent; see Table I) on the
known structure of the coiled coil that forms part of the dimeriza-
tion interface of the CAP protein (Weber and Steitz, 1987). The
modeled portion of the sequence of CAP is shown in Figure 2.
The final refinement in explicit solvent was omitted as the coiled
coil is largely surrounded by the protein matrix rather than water.
In order to improve sampling of the conformational space of the
coiled coil, we used a number of different starting orientations
θ around the orientation θ = 35° which maximizes the
hydrophobic contact area (see Figure 1). The starting orienta-
tions were symmetric around the 2-fold of the coiled coil. We
chose six initial conformations with θ varying from 25 to 50°
in 5° steps. Stages II and III of the automated model building
procedure (Table I) were applied individually to each of them.

Figure 3 shows a superposition of the predicted structures
and the crystal structure (Weber and Steitz, 1987). The crossing angle
of the coiled coil self-adjusted itself to an average of 21° in the
models, compared with 22° in the X-ray structure. θ (as defined
in Figure 1) at residue T17 is 39° in the models, while it is 41°
in the X-ray structure. For five turns of helix, from residue M10
to the C-terminus, the interhelical distances in the two structures
are virtually identical: 9.6 Å in the X-ray structure, compared
with an average of 9.5 Å in the models. For residues D1 to M10,
the inter-helical distance is 10.2 Å in the X-ray structure, while
it is predicted in the model to be only 8.9 Å. The smaller inter-
helical distance between D1 and M10 is presumably due to the
absence of the surrounding protein matrix.

These differences correspond to relatively small atomic r.m.s.
differences. The r.m.s. difference for backbone atoms,

Fig. 3. Stereoplot of the X-ray structure (heavy line) and predicted structures (thin lines) of CAP. The spheres highlight Cα atoms in the 'd' position in the
X-ray structure. The right pair of images is for normal stereo viewing.

Fig. 4. Eisenberg and McLachlan (1986) free energy (solid line), and actual
interaction energy between protein and solvent as used in the simulation
dashed line), for one of the molecular dynamics runs in explicit solvent
(stage IV in Table 1)
disregarding the first and last three residues, between the predicted structures and the crystal structure varies between 0.6 and 0.9 Å, which compares to a r.m.s. difference between the models of 0.7 Å. In light of the generally larger atomic r.m.s. differences reported for molecular dynamics calculations in vacuo [e.g. Figure 1 of Levitt and Sharon (1988) for a comparison of several calculations for BPTI] this is a promising result. Note also that, in contrast to the calculations compared by Levitt and Sharon (1988), our calculations do not start from the exact X-ray structure coordinates, but from rather distorted structures (i.e. two parallel helices), and converge towards the X-ray structure.

Side-chain positions at the dimerization interface (Figure 3), which are not restrained at any point during the calculation, are similar in the two structures. The atomic r.m.s. differences for the non-hydrogen atoms of the buried side chains (L3, S7, M10, L14, T17, V21) between the models and the X-ray structure vary between 1.1 and 1.8 Å. The average r.m.s. difference between models for these atoms is 1.3 Å. Side chains outside the dimerization interface are more disordered (the average r.m.s. difference between models for all non-hydrogen side chain atoms is 2.7 Å, which is not surprising considering the missing surrounding protein matrix and solvent). Note that some side chains are in unsymmetrical conformations in the X-ray structure.

### Prediction of GCN4 structure

A coiled coil motif is the conformation suggested by Landschulz et al. (1988) for the dimerization region of a new class of transcriptional DNA binding proteins, termed the ‘leucine zipper’. It was originally proposed on the basis of sequence similarity studies. The sequences exhibit a conserved heptad repeat of leucines, and many of the sequences show the 4-3 repeat of hydrophobic residues which is characteristic of coiled coils (Hodges et al., 1972; McLachlan and Stewart, 1975). The coiled coil character of leucine zippers has been corroborated by a variety of experimental evidence, e.g. circular dichroism (O'Shea et al., 1989a,b), two-dimensional 	extsuperscript{1}H nuclear magnetic resonance (Oas et al., 1990; Saudek et al., 1990), and preliminary X-ray diffraction data (Rasmussen et al., 1991). Vinson et al. (1989) modeled the dimerization and DNA-binding domain of rat liver nuclear protein (C/EBP) complexed with DNA with computer graphics and proposed a model for the interaction of C/EBP with DNA (the ‘scissors grip’ model).

For our study of the conformation of the leucine zipper domain of GCN4, we have modeled the 37 C-terminal residues, starting at residue 245 (see Figure 2). The first of the ‘zipper’ leucines, which are in the ‘d’ position, is residue number 9 in our convention (see Figures 1 and 2).

The same starting orientations were chosen as in the calculations with the CAP dimer. Stage IV of our procedure, i.e. molecular dynamics in explicit solvent, proved necessary for the leucine zipper models, as the GCN4 helices are highly charged and fully solvated. An additional starting structure for the fourth refinement stage was obtained by averaging the six structures after stage III. Each of these seven structures was placed in a cylinder of water with 70 Å length and 18 Å radius (see Materials and methods). The simulated system comprised 800 protein atoms and ~1900 water molecules. Each calculation required ~2 h CPU time on a Cray YMP.

Artifacts present in the structures after stage III, such as headgroups of lysines folded back to the protein, have disappeared in the solvated models after stage IV (not shown). Figure 4 shows two parameters that were monitored during the molecular dynamics runs, the interaction energy between solvent and protein, and an approximate solvation free energy calculated according to Eisenberg and McLachlan (1986). Note that this latter energy was computed for the analysis only and was not part of the energy function during the molecular dynamics calculations. The Eisenberg and McLachlan free-energy term has been used successfully to discriminate between correctly and incorrectly folded structures (Novotny et al., 1988), and to detect vacuum artifacts in structures calculated from NMR data by restrained molecular dynamics in vacuo (Chiche et al., 1989). As Figure 4 shows, the structures seem to have relaxed enough for the purpose of our structure prediction after 5 ps of molecular dynamics. The total energies at the end of the seven calculations are comparable (the standard deviation of the energy values of the seven structures is <1% of the average total energy).

Superpositions of the initial structures after stage II are shown in Figure 5, and superpositions of the final structures after stage IV are shown in Figure 6 for the backbone atoms and leucine side chains, and Figure 7 for all non-hydrogen atoms. All structures converged roughly to the same global conformation. Figure 8 shows, as a function of residue number, the average r.m.s. difference between the initial structures after stage II and the final structures, and the average pairwise r.m.s. difference among the final structures. Overall, the average pairwise r.m.s. difference is 1.7 (1.2) Å for backbone atoms and 2.9 (2.3) Å for all non-hydrogen atoms, where the numbers in parentheses are computed without the first and last turn of the helices. This is approximately half as large as the average r.m.s. shift from the initial to the final structures, which is 3.1 (2.3) Å for backbone.
Fig. 6. Two views of the superposition of the backbone atoms of all seven modeled GCN4 structures, with the leucine side chains at position ‘d’. The spheres show the Cα positions of the leucines, averaged over the seven structures. The DNA binding side is at the bottom of the figure. The right pair of images is for normal stereo viewing.
Fig. 7. Superposition of all non-hydrogen atoms of the modeled GCN4 structures. The spheres show the C° positions of the leucine side chains at position 'd', averaged over the seven structures. The DNA binding site is at the bottom of the figure. The right pair of images is for normal stereo viewing.

Fig. 8. Average r.m.s. differences between initial structures after side-chain building (stage II) and final structures (stage IV) (a), and average pairwise r.m.s. differences between the seven final structures (b), as a function of residue number. Dashed line, backbone atoms; solid line, all non-hydrogen atoms.

Fig. 9. Local crossing angle \( \phi(i) \) between the two helices as a function of residue number \( i \). This angle at residue number \( i \) was estimated as the average of two pseudo-torsion angles between points on the helix axes two turns apart:

\[
\phi(i) = \frac{\phi^+(i) + \phi^-(i)}{2}
\]

with

\[
\phi^+(i) = P^g_i - P^g_{i+7} - P^g_{i+7} - P^b_i,
\]

\[
\phi^-(i) = P^g_i - P^g_{i-7} - P^g_{i-7} - P^b_i
\]

where \( P^g_i \) is the geometric center of seven consecutive C° atoms in monomer \( A \) with residue number \( i \) at the center.
Fig. 10. Superposition of two helical turns of the seven models in three views along the long axis of the coiled coil. The backbone atoms, and all non-hydrogen atoms of some side chains are shown. (a) A pair of 'zipper' leucines (in the 'd' position) (b) The highly conserved asparagine residue N20 (in the 'a' position). (c) Interaction between the side chains of K31 (in the 'e' position) and E26 (in the 'g' position). The right pair of images is for normal stereo viewing.

Fig. 11. Inter-monomer side-chain contact map for GCN4. The distances were measured in each individual model from the geometric centers of the non-hydrogen side-chain atoms starting at the γ position. The distances were then averaged over the seven models.
Automated modeling of coiled coils

Inter-helix charge-charge contact map

Fig. 12. Average electrostatic energy between pairs of residues in the seven final structures, between the two monomers (a), and within one of the monomers (b).

Intra-helix charge-charge contact map

Atoms and 4.1 (3.2) Å for all non-hydrogen atoms. Residues in the dimer interface show smaller r.m.s. differences among the seven structures (Figure 7). Clearly, the first and last turns of the helices of the models are less well ordered than the middle section, as there are fewer conformational constraints on residues near the ends.

Variation of conformation and energy along the sequence

The geometry of the coiled coil can be characterized by the crossing angle between the two helices, the angle θ (as defined in Figure 1), and the inter-helical distance. θ is ~37°, and the inter-helical distance is 9.8 Å throughout most of the molecule. Interestingly, the distance is slightly smaller around L9 (9.5 Å), and ~1 Å larger for the N-terminal turn of helix (10.6 Å), which is at the DNA-binding end of the coil. The crossing angle at different positions along the sequence is shown in Figure 9. The angle exhibits a small systematic variation, namely, it is larger towards the C-terminal end of the molecule. This variation is significant even considering the standard deviation of several degrees, as each individual model shows this variation. In other words, each of the seven models unwinds slightly towards the N-terminal (DNA-binding) end of the molecule. The leucines
form a ladder-like structure (Figure 10a), similar to the anti-parallel coiled coil in the structure of the serine amino acid synthetase (Cusack et al., 1990).

Figure 11 is an inter-monomer side-chain contact map. The distances were measured in each individual model from the geometric centers of the non-hydrogen side-chain atoms starting at the g position in one monomer with the one at the g position in the other. The map clearly illustrates the proximity of the two monomers. The map likely shows the proximity of the non-hydrogen side-chain atoms starting at the helix—helix interface (at the a and d positions), and in some cases of side chains at the e position in one monomer with the one at the g position in the other.

Figure 12 shows the inter- and intra-helical electrostatic interaction energy on a residue-by-residue basis averaged over the seven models (a ‘charge—charge contact map’). Negative (attractive) electrostatic energies between two residues show up as darker than background, and positive (repulsive) electrostatic energies as lighter than background. Rather than representing free energy contributions, this plot reflects the spatial arrangement of the charges during the simulation. Note that repulsive interactions are largely absent.

The four strongest pairs of charge—charge contacts may contribute to the stability of the dimer, namely, E10—R5, K12—E10, E24—K19 and K31—E26 (Figure 12a and Table III). R5 is in the basic region of the DNA-binding domain, and it can conceivably interact with DNA backbone in the protein—DNA complex. In several of the models, the distance between N20 and O2 of E26 is smaller than 4 Å, which qualifies as a saltbridge. This is also apparent in Figure 10(c), which shows a view of the backbone atoms from residue 25 to residue 32, with the side chains of K31 and E26. A saltbridge can also be formed between E24 and K19. This saltbridge occurs less often in our models.

Figure 12(b) shows the intra-helical charge—charge contact map for one of the helices. As suggested by Landschulz et al. (1988) for the sequence of C/EBP, there are a number of putatively helix-stabilizing attractive interactions, mostly over one turn of the helix (see also Table III). As one would expect, a characteristic pattern of i — i + 4 attractive interactions is observed between E14 and R29 (Figure 12b). A particularly strong interaction over two turns (E14—K7) may facilitate the branching of the dimer at this point.

Of special interest is the conformation of the side chain of residue N20 (Figures 7 and 10b). This residue is highly conserved in the known leucine zipper sequences (Vinson et al., 1989) and may play a special role, as recent experiments have shown that it is not required for dimerization (Hu et al., 1990). In our models, the polar N20 side chain is buried and thus inaccessible for water. We note that in some of the models hydrogen bonds between the N of N20 and the carbonyl oxygen of residue 19 of the opposite monomer are formed. This appears as a slightly negative electrostatic energy in Figure 12(a) between residues 19 and 20.

Conclusions

We have presented a largely automated model-building procedure for coiled coils and used it to predict the structure of the dimerization domain (‘leucine zipper’) of GCN4. The aim was to build models of sufficient quality to be able to extract meaningful information on structure and, to some degree, on energetics from them. To ensure this, we included as a final stage in the model-building procedure a molecular dynamics run in explicit solvent. The methodology is applicable to structure prediction of other coiled coils in both fibrous and globular proteins.

The family of predicted structures of GCN4 converge from different starting conditions to similar coiled coil conformations with the leucines in a ladder-like arrangement. In particular, the local crossing angle is similar in all models, without using the crossing angle as a restraint. The angle between the two helices decreases towards the N-terminus, i.e. the dimer unwinds slightly towards the DNA binding end. We expect that the actual solution structure of GCN4 falls within this family of predicted models. A number of attractive charge—charge interactions emerged which may contribute to the stabilization of the helices and the dimer.

Application of the technique to mutations of GCN4 will hopefully contribute to the understanding of the wealth of experimental data obtained through site-directed mutagenesis (Hu et al., 1990). Furthermore, we hope to be able to address the structural basis for the sequence specificity of dimerization in the Fox-Jun system (O'Shea et al., 1989b; M.Nilges and A.T.Brünger, in preparation).

Acknowledgements

We thank Paul Sigler for drawing our attention to the leucine zipper protein Dushyant Pathak, Thomas Simonson and Herbert Treutlein for helpful discussions. This work was in part supported by a grant from the Pittsburgh Supercomputing Center (grant No DMB870007P).

References


Automated modeling of coiled coils


Received on February 13, 1991, accepted on May 22, 1991