NuMA/Centrophilin: Sequence Analysis of the Coiled-Coil Rod Domain

David A. D. Parry
Department of Physics, Massey University, Palmerston North, New Zealand

ABSTRACT Nuclear mitotic apparatus protein (NuMA), also known as centrophilin, has been shown in previous work to contain a centrally located sequence of length 1485 residues that has both a heptad substructure and a high propensity for α-helix formation. Further analysis of this sequence here has revealed that NuMA will form a two-stranded coiled-coil structure with multiple (18) points at which the conformation is interrupted either by proline-containing segments or by discontinuities in the phasing of the heptad substructure. It has also been shown that the two chains will be parallel (rather than antiparallel), that they will lie in axial register, and that this arrangement will be stabilized by a large number of interchain ionic interactions. Interestingly the coiled-coil rod domain is also shown to lack any significant long-range periodicity in the linear distribution of either its acidic or its basic residues. Hence there is no direct evidence from the sequence data that NuMA molecules will aggregate to form closely packed filaments within nuclear space.

INTRODUCTION

The presence (or otherwise) of structural elements in the eukaryotic cell nucleus has remained a controversial topic over the years (Cook, 1988). Filaments or branched networks have been observed by electron microscopy but usually only after the nucleus has been subjected to quite severe preparative processing (Berezney and Coffey, 1974, 1977). Nuclear lamin filaments, however, are now well established and in the case of *Xenopus* oocytes have been shown to form a tetragonal meshwork on the inner side of the nuclear membrane (Aebi et al., 1986). More recently, Yang et al. (1992) have characterized and sequenced another nuclear protein which they have termed NuMA (nuclear mitotic apparatus) protein. This is a 238-kDa phosphoprotein which dissociates from condensing chromosomes at the earliest stage of prophase (and hence before the disintegration of the nuclear lamina). At the end of mitosis and before lamin aggregation on the chromosomes NuMA reassociates with telophase chromosomes. Yang et al. (1992) suggest, therefore, that NuMA may play an important role during the early stages of reassembly of the cell nucleus after mitosis. The possibility that NuMA molecules may aggregate into filaments and thus account for the presence of the observed structural elements in the nucleus was also discussed.

In their analysis of the sequence, Yang et al. (1992) noted about 20% homology between a long stretch of sequence in NuMA (1485 residues) and that constituting the coiled-coil rod domain sequences in α-fibrous proteins such as myosin and intermediate filament proteins. The NuMA sequence, like all α-fibrous proteins and those globular proteins which contain α-helical bundles, contains a heptad repeat of the form $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$, where *a* and *d* are positions commonly occupied by apolar residues. The presence of a heptad substructure in an amino acid sequence is now recognized as a strong indication that the conformation adopted will be that of an α-helical coiled-coil (Cohen and Parry, 1986, 1990, 1994). Yang et al. (1992) assumed that NuMA would dimerize to form a two-stranded rather than three-stranded coiled-coil rod, although no evidence was presented to support this particular choice. It was reported, however, that six short intervening segments containing non-helix-favoring proline residues would tend to break up the rod domain into seven coiled-coil segments and also that the heptad repeat was not continuous within these segments, although no details were given. If fully coiled-coil, then the rod domain would be about 1485 × 0.1485 nm in length i.e., 221 nm. (The average axial rise per residue is 0.1485 nm in coiled-coil proteins.) NuMA would thus have the longest coiled-coil rod domain of any molecule yet known. With this background in mind the coiled-coil domain of the NuMA sequence was reexamined in an effort to clarify several aspects of the structure, and, in particular, the likely number of strands composing the rod domain, the axial stagger between chains and their relative polarity (parallel or antiparallel), and the likelihood that NuMA molecules may aggregate into filamentous structures.

First, the level of homology (~20%) observed by Yang et al. (1992) between the heptad-containing regions of NuMA and several α-fibrous proteins is no higher than would be expected by chance, since all coiled-coil α-fibrous proteins when compared with one another show this same general level of sequence identity. The homology observed is thus merely a reflection of the underlying heptad substructure and has no further significance.

Second, it has been confirmed by DOTPLOT and Fourier transform techniques that the heptad-rich substructure does indeed extend between about residues 216 and 1700 (1485 residues), thus leaving a 215 residue (acidic) N-terminal domain and a 415 residues (basic) C-terminal domain. However, the heptad substructure is more broken up than indicated by Yang et al. (1992), and 19 segments have been identified in which the standard "rules" of heptad delineation

Received for publication 31 March 1994 and in final form 14 June 1994.
Address reprint requests to Dr. David A. D. Parry, Department of Physics, Massey University, Palmerston North, New Zealand.

0006-3495/94/09/1203/04 $2.00

The Amino Acid Sequence of NuMA/Centrophilin

```
1    M T L H A T R G A A  L L S W V N S L H V  A D P V E A V L Q L  Q D C S I F I K I I  D R I H G T E E G Q
51   Q I L K Q P V S E R  L D F V C S F L Q K  N R K H P S S P E C  L V S A Q K V L E G  S E L E L A K M T M
101  L L L Y H S T M S S  K S P R D W E Q F E  Y K I Q A E L A V I  L K F V L D H E D G  L N L N E D L E N F
151  L Q K A P V P S T C  S S T F P E E L S P  P S H Q A K R E I R  F L E L Q K V A S S  S S G N N F L S G S
201  P A S P M G D I L Q  T P Q F Q (M R R L K  K Q L A D E R S N R  D E L E L E L A E N  R K L L T E K D A Q
251  I A M M Q Q R I D R  L A L L) N E K Q A A  S R L E R K (E L E E  L R D K N E S L T M  R L H E T L K Q C Q
301  D L K T E K S Q M D  R K I N Q L S E E N  G D L S F K L R E F  A S H L Q Q L Q D A  L N E L T E E H S K
351  A T)(Q E W L E K Q A  Q L E K E L S A A L  Q D K K C L E E K N  E I L Q G K L S Q L  E E E L S Q L)Q D N
401  R R Q E K G E V L G  D(V L Q L E T L K Q  E A A T L A A N N T  Q L Q A R V E M L E  T E R G Q Q E A K L
451  L A E R G H F E E E  K Q Q L S S L I T D  L Q S S I S N L S Q  A K E E L E Q A S Q  A)H G(A R L T A Q Y
501  A S L T S E L T T L  N A T I Q Q Q D Q E  L A G L K Q Q A K E  K Q A Q L A Q T L Q  Q Q E Q A S Q G L R
551  H Q V E Q L S S) S L  K Q K E Q Q L K E V  A E K Q E A T R (Q D  H A Q Q L A T A A E  E R E A S L R E R D
601  A A L K Q L E A L E  K E K A A K L E I L  Q Q Q L Q V A N E A  R D S A Q T S V T Q  A Q R E K A E L S R
651  K V E E L Q A C V E  T A R Q E Q H E A Q  A Q V A E L E L Q L  R S E Q Q K A T E K  E R V A Q E K D Q L
701  Q E Q L Q A L K E S  L K V T K G S L E E  E K R R A A D A L E  E Q Q R C I S E L K  A E T R S L V E Q)H
751  K R E R K E L E E E  R A G R K G L E A R  L Q Q L G E A H Q A  G T(E V L R R E L A  E A M A A Q H T A E
801  S E C E Q L V K E V  A A W R E R Y E D) S  Q Q E E A Q Y G (A M  F Q E Q L M T L K E  E C E K A R Q E L Q
851  E A K E K V A G I E)  S H S (E L Q I S R Q  Q N E L A E L H A N  L A R A L Q Q V Q E  K E V R A Q K L A D
901  D L S T L Q E K M A  A T S K E V A R L E  T L V R K A G E Q Q  E T A S R E L V K) E  R A R A G D R Q R E
951  W L E E Q Q G R Q F  C S T Q (A A L Q A M  E R E A E Q M G N E  L E R L R A A L M E)  S Q G Q Q Q E E R G
1001 Q Q(E R E V A R L T  Q E R G R A Q A D L  A L) E K A A R A E L  E M R (L Q N A L N E  Q R V E F A T L Q E

1051 A L A H A L T E K E  G K D Q E L A K L R  G L E A A Q I K E L  E E L R Q T V K Q L  K E Q L A K)K E K E
1101 H A S G S G A Q S E  A A G R T E R T G R  K(L E A L R A E V S  K L E Q Q C Q K Q Q  E Q A D S L E R S L
1151 E A E R A S R A E R  D S A L E T L Q G Q  L E E K A Q E L G H  S Q S A L A S A Q R  E L A A F R T K V Q
1201 D H S K A E D E W K  A Q V A R G R Q E A  E R K N S L I S S L  E E E V S I L N R Q  V L E K E G E S K E
1251 L K R L V M A E S E  K S Q K L E E R L R  L L Q A E) T A S N S  (A R A A E R S S A L  R E E V Q S L R E E
1301 A E K Q R V A S E N  L R Q E L T S Q A E  R A E E L G Q E L K  A W Q E K F F Q K E  Q A L S T L Q L E H
1351 T S T Q A L V S E L  L) R (A K H L C Q Q L  Q A E Q A A A E K)R  H R E E L E Q S K Q  A A G G (L R A E L L
1401 R A Q R E L G E L I  R L R Q K V A E Q E  R T A Q Q L R A E K  A S Y A E Q L S M L  K K A H G L L A E E
1451 N R G L G E R A N L  G R Q F L E V E L D  Q A R)(E K Y V Q E L  A A V R A D A E T R  L A E V Q R E A Q S
1501 T A R E L E V M T A  K Y E G A K V K V L  E E R Q R F Q E E R  Q K L T A Q V E Q L  E V F Q R) E Q T K Q
1551 V E E L S K K L A D  S D Q A S K V Q Q Q  K L K A V Q A Q G G  E S Q Q (E A Q R L Q  A Q L N E L Q A Q L
1601 S Q K E Q A A E H Y  K L Q M E K A K T H  Y D A) K K Q Q N Q (E  L Q E Q L R S L E Q  L Q K E N K E L R A
1651 E A E R L G H E L Q  Q A G L K T K E A E  Q T C R H L T A Q V  R S L E A Q V A H A  D Q Q L R D L G K F)
1701 Q V A T D A L K S R  E P Q A K P Q L D L  S I D S L D L S C E  E G T P L S I T S K  L P R T Q P D G T S
1751 V P G E P A S P I S  Q R L P P K V E S L  E S L Y P T P I P A  R S Q A P L E S S L  D S L G D V F Q D S
1801 G R K T R S A R R R  T T Q I I N I T M T  K K L D V E E P D S  A N S S F Y S T R S  A P A S Q A S L R A
1851 T S S T Q S L A R L  G S P D Y G N S A L  L S L P G Y R P T T  R S S A R R S Q A G  V S S G A P P G R N
1901 S F Y M G T C Q D E  P E Q L D D W N R I  A E L Q Q R N R V C  P P H L K T C Y P L  E S R P S L S L G T
1951 I T D E E M K T G D  P Q E T L R R A S M  Q P I Q I A E G T G  I T T R Q Q R K R V  S L E P H Q G P G T
2001 P E S K K A T S C F  P R P M T P R D R H  E G R K Q S T T E A  Q K K A A P A S T K  Q A D R R Q S M A F
2051 S I L N T P K K L G  N S L L R R E A S K  K A L S K A S P N T  R S G T R R S P R I  A T T T A S A A T A
2101 A A I G A T P R A K  G K A K H
```

TABLE 1 Breakdown of residues in heptad segments

Residues	*a* (178)*‡	*b* (183)	*c* (182)	*d* (175)	*e* (179)	*f* (177)	*g* (179)
LIVMFYA	135	45	32	120	34	50	34
L	70	3	4	79	11	9	10
IV	32	2	8	2	8	5	2
A	22	36	16	28	14	34	19
DEKR	17	72	80	27	81	67	90
KR	14	25	29	15	43	30	32
DE	3	47	51	12	38	37	58

* The number in brackets indicates the number of residues *in toto* in this position.

‡ Note that 76% of the *a* and *d* positions are occupied by apolar residues. Leucine dominates *a* and *d* and the branched sidechain residues (IV) occur predominantly in *a*. Regarding the charged residues the *e* position is the most basic and the *g* position is the most acidic, a feature seen in other two-stranded α-fibrous proteins. Positions *b*, *c*, and *f* (the outermost positions) are all clearly net acidic.

are followed (Cohen and Parry, 1986, 1990). These segments, together with the heptad position at which they commence, are as follows: 216–264, *a;* 277–352, *g;* 353–397, *b;* 412–491, *e;* 494–558, *b;* 579–749, *e;* 783–819, *b;* 829–860, *b;* 864–939, *e;* 965–990, *f;* 1003–1022, *e;* 1034–1096, *d;* 1122–1275, *a;* 1281–1361, *b;* 1363–1379, *a;* 1395–1473, *a;* 1474–1545, *e;* 1585–1623, *g;* 1630–1700, *g* (Fig. 1, Table 1). Some of the breaks in the coiled-coil rod domain are proline-rich, others merely represent discontinuities in the phasing of the heptad substructure. All are likely to have some effect on the ability of the rod to bend, kink, or otherwise allow some deformation to occur.

Third, calculation of the number of interchain ionic interactions between the coiled-coil segments as a function of relative axial stagger (primarily between oppositely charged residues in the *e* and *g* positions) yields a high total (74 interactions) for a parallel in-register chain arrangement, i.e., 0.41 ionic interactions per heptad pair (cf. 0.23–0.77 for two-stranded coiled-coil molecules; Conway and Parry, 1990). For relative staggers of ± 1 or ± 2 heptads, the scores are 0 in both cases. Antiparallel arrangements all have low scores and are highly improbable, since heptad-containing segments of differing lengths would not permit maximal overlap to occur and the percentage of possible coiled-coil structure in the rod domain would necessarily be much reduced as a consequence. Furthermore, there are no examples yet available of an antiparallel stranded α-fibrous protein.

The fourth feature of the structure that can be deduced from the sequence is the likely number of strands. Several indicators provide support for the hypothesis that the molecule is two-stranded. 1) Valine and isoleucine residues with their branched side chains are found far more commonly in position *a* than in position *d* (32:2). This is consistent with the crystallographic studies of Alber, Kim, and colleagues on GCN4-leucine zipper mutants in which isoleucine and leucine residues in *a* and *d*, respectively, correlate specifically to two-stranded structures (Harbury et al., 1993). 2) The charged to apolar residue ratio of the rod domain is 1.04. Two-stranded coiled-coils typically have values of ~1.0 and three-stranded structures have values of ~0.8 (Parry and Cohen, 1990). 3) The average number of residues per heptad segment is 78 (cf. 80 and 45 residues, respectively, for two- and three-stranded structures). 4) The percentages of charged residues in *e* and in *g*, of apolar residues in *e* and in *g*, of acidic residues in *a* and in *d*, and of basic residues in *a* all lie closer to those found in an "average" two-stranded coiled-coil than in a three-stranded structure (Conway and Parry, 1990, 1991). Only the percentage of basic residues in position *d* lies closer to that expected of a three-stranded "average" structure. 5) There are, as yet, no examples of a three-stranded coiled-coil occurring intracellularly; all are two-stranded. It seems reasonable, therefore, to believe that NuMA will indeed adopt a two-stranded structure as proposed by Yang et al. (1992).

A fast Fourier transform analysis has been undertaken to determine whether any significant periodicities occur in the linear distribution of the acidic or the basic residues in the rod domain. Common periods in this pair of residue groupings have been found in all α-fibrous proteins that form filaments (e.g., myosin, paramyosin, intermediate filaments). Those α-fibrous proteins lacking such periods (e.g., laminin, M-protein, gp17) exist in vivo in a molecular nonfilamentous form. It is not yet known whether other α-fibrous proteins that contain significant periods in their charged residue distributions (e.g., plectin, desmoplakin, bullous pemphigoid antigen) form filaments or networks. In the case of NuMA, the fast Fourier transform technique has failed to find any evidence of a long-range significant period in any residue grouping, let alone a common period in the acidic and the basic residues. Not only has the rod domain as a whole been studied but each of the major continuous fragments of the rod domain has also been investigated individually. These negative results suggest that NuMA does not form a close packed

FIGURE 1 The amino sequence of NuMA (Yang et al., 1992) is given in the usual one-letter code (A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine). The N-terminal domain (residues 1–215) and the C-terminal domain (residues 1701–2115) are separated by a largely coiled-coil rod domain of length 221 nm. The *a* and *d* positions in the heptad-containing coiled-coil segments are indicated by ●, and each coherent heptad-containing segment is enclosed within brackets. Proline residues in the rod domain are underlined and bold, and rarely occur within an α-helix-rich structure except within the first turn of α-helix.

filamentous structure of the type seen in myosin thick filaments or in intermediate filaments. It would also seem certain that the 23-nm period visualized in some nuclear filaments does not arise from NuMA aggregates and is more likely attributable to intermediate filaments or intermediate filament-like filaments in which this period is a characteristic feature. The possibility that some form of head-to-tail aggregation of NuMA molecules resulting from charge interactions between the acidic N-terminal domain and the basic C-terminal domain cannot, of course, be eliminated. A loose and semi-regular meshwork of molecules thus remains a structural option. If NuMA exists purely in the molecular form then it would seem that the role of the coiled-coil domain is to provide a means by which the two chains can aggregate and to keep spatially apart the functional N- and C-terminal domains. More work will be required to ascertain the precise roles of the terminal domains in vivo.

REFERENCES

Aebi, U., J. B. Cohn, L. Buhle, and L. Gerace. 1986. The nuclear lamina is a meshwork of intermediate filament type filaments. *Nature*. 323: 560–564.

Berezney, R., and D. S. Coffey. 1974. Identification of a nuclear protein matrix. *Biochem. Biophys. Res. Commun.* 60:1410–1417.

Berezney, R., and D. S. Coffey. 1977. Nuclear matrix: isolation and characterization of a framework structure from rat liver nuclei. *J. Cell Biol.* 73:616–637.

Cohen, C., and D. A. D. Parry. 1986. Alpha-helical coiled coils—a widespread motif in proteins. *Trends Biochem. Sci.* 11:245–248.

Cohen, C., and D. A. D. Parry. 1990. α-Helical coiled-coils and bundles: how to design an α-helical protein. *Protein Struct. Funct. Genet.* 7:1–15.

Cohen, C., and D. A. D. Parry. 1994. α-Helical coiled-coils: more facts and better predictions. *Science*. 263:488–489.

Conway, J. F., and D. A. D. Parry. 1990. Structural features in the heptad substructure and longer range repeats of two-stranded α-fibrous proteins. *Int. J. Biol. Macromol.* 12:328–334.

Conway, J. F., and D. A. D. Parry, 1991. Three stranded α-fibrous proteins: the heptad repeat and its applications for structure. *Int. J. Biol. Macromol.* 13:14–16.

Cook, P. R. 1988. The nucleoskeleton: artefact, passive framework or active site? *J. Cell Sci.* 90:1–6.

Harbury, P. B., T. Zhang, P. S. Kim, and T. Alber. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*. 262:1401–1407.

Parry, D. A. D., and C. Cohen. 1990. Structure of the spectrin superfamily: a three α-helix motif. *In* The Living Cell in Four Dimensions. G. Paillotin, editor. AIP Conference Proceedings 226. American Institute of Physics, New York. 367–377.

Yang, C. H., E. J. Lambie, and M. Snyder. 1992. NuMA: an unusually long coiled-coil related protein in the mammalian nucleus. *J. Cell Biol.* 116: 1303–1317.