In Silico Search for Functionally Similar Proteins Involved in Meiosis and Recombination in Evolutionarily Distant Organisms

Yuri F. Bogdanov*, Sergei Y. Dadashev and Tatiana M. Grishaeva

N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

Edited by E. Wingender; received 30 September 2002; revised and accepted 28 January 2003; published 1 March 2003

ABSTRACT: Evolutionarily distant organisms have not only orthologs, but also nonhomologous proteins that build functionally similar subcellular structures. For instance, this is true with protein components of the synaptonemal complex (SC), a universal ultrastructure that ensures the successful pairing and recombination of homologous chromosomes during meiosis. We aimed at developing a method to search databases for genes that code for such nonhomologous but functionally analogous proteins. Advantage was taken of the ultrastructural parameters of SC and the conformation of SC proteins responsible for these. Proteins involved in SC central space are known to be similar in secondary structure. Using published data, we found a highly significant correlation between the width of the SC central space and the length of rod-shaped central domain of mammalian and yeast intermediate proteins forming transversal filaments in the SC central space. Basing on this, we suggested a method for searching genome databases of distant organisms for genes whose virtual proteins meet the above correlation requirement. Our recent finding of the Drosophila melanogaster CG17604 gene coding for synaptonemal complex transversal filament protein received experimental support from another lab. With the same strategy, we showed that the Arabidopsis thaliana and Caenorhabditis elegans genomes contain unique genes coding for such proteins.

KEYWORDS: databases, computer analysis, functional proteomics, meiosis, cell ultrastructure, synaptonemal complex, coiled coil, Arabidopsis thaliana, Caenorhabditis elegans

INTRODUCTION

Genome databases have accumulated and processed data on the complete genome sequences of model eukaryotic organisms: yeast Saccharomyces cerevisiae, nematode Caenorhabditis elegans, fruit fly Drosophila melanogaster, and plant Arabidopsis thaliana. These organisms have been found to possess several hundreds of orthologous genes and proteins which are similar in their primary structure and play a common role. These evolutionarily distant organisms have such organelles as kinetochores, cell centers, synaptonemal complexes (SCs), etc., which partly or completely differ in their ultrastructure, notwithstanding the common functions. In many cases, these subcellular structures are built of different structural...
proteins. Here such proteins are referred to as functional analogs. To search for analogs, we developed a strategy which collates the data of computer analysis of the conformation and other physico-chemical properties of constitutive proteins with the data on in situ electron microscopic analysis of the linear parameters of cell organelles presumably built of these proteins.

Meiotic pairing of homologous chromosomes depends on SC, a specific protein structure formed in prophase I at the interface between aligned and paired homologs. Owing to SC, the homologs are held about 100 nm apart while crossing over is completed [Zickler and Kleckner, 1999]. The SC assembly is an important prerequisite to crossing over, formation of chiasmata (non-sister chromatid configuration which is formed after crossing over), and chromosome segregation in meiosis I.

In all eukaryotes, SC (Figure 1) has a universal general structure [Zickler and Kleckner, 1999]. Meiosis-specific protein axial cores of two homologous chromosome are arranged parallel to each other to produce two SC lateral elements (LE). The lateral elements are connected with protein transversal filaments (TFs). Chromatin fibrils are attached to the lateral elements. Thus, SC acts as a framework and temporarily holds homologous chromosomes together so that each locus is strongly opposite to its counterpart. The space between the lateral elements is known as central space. This harbors recombination nodules, conglomerates of enzymes necessary for DNA recombination [Zickler and Kleckner, 1999].

In spite of the universal general arrangement, the SC ultrastructure differs among ascomycetes, nematodes, insects, mammals, higher plants, and other remote taxa. In particular, the difference concerns the ultrastructure of the lateral elements and the transversal filament pattern in the central space [von Wettstein et al., 1984; Zickler and Kleckner, 1999]. Structural proteins of SC have been studied experimentally in four mammalian species [Meuwissen et al., 1992; Dobson et al., 1994; Liu et al., 1996; Meuwissen et al., 1997] and in yeast Saccharomyces cerevisiae [Tung and Roeder, 1998; Dong and Roeder, 2000] (for a review, see Heyting, 1996 and Penkina et al., 2002). Although functionally analogous, yeast and mammalian SC proteins have no homology. The question arises as to how the functionally equal subcellular structures are assembled from different proteins in meiotic cells of evolutionarily distant organisms? Which features of the nonhomologous proteins are responsible for the similarity of the SC general structure among eukaryotes of different kingdoms?

![Fig. 1. A sketch of the synaptonemal complex. CE — the central element; RN — recombination nodule (the compartment for recombination enzymes).](image-url)
The transversal filament proteins are of special interest in this respect. Both in yeast and in mammals, SC transversal filaments are formed by species-specific proteins. These are Zip1 in *S. cerevisiae* and SCP1 (SYCP1) in mammals. Notwithstanding the evolutionary distance between these organisms, the proteins are structurally similar. Each comprises about 900 amino acid residues and consists of three domains: a central one, which forms a coiled-coil, and two terminal domains of another secondary structure. The C-terminal domain is globular and possesses the DNA-binding activity (see Tung and Roeder, 1998). Both Zip1 and SCP1 are classed with intermediate proteins, such as those of the spindle pole body, nuclear lamina, and nuclear matrix; the myosin heavy chain; and several other similar proteins of yeast, *C. elegans*, *Drosophila*, mouse, and human [Meuwissen et al., 1992; Liu et al., 1996; Dong and Roeder, 2000].

The SC transversal filament resembles two zipper tooth-like halves which are engaged with each other to connect the SC lateral elements. In mammals, a tooth is a dimer of similarly oriented parallel SCP1 molecules (Figure 2). The C-terminal domain of each molecule is attached to the SC lateral element. The N-terminal domain is directed inwards (in the SC central space) to face the N-terminal domains of molecules forming the opposite zipper tooth [Heyting, 1996]. As we deduced from the *S. cerevisiae* Zip1 sequence, the N end (pl 4.8) of the protein may interact with the basic N-subterminal region (residues 200–300, pl 9.6) of the coiled-coil. Possibly, such interaction of the two opposite teeth provides for the formation of the SC central element (CE) (Figure 1), which is clearly seen on electron microscopic images of ultrathin cell sections or surface spreads of isolated SCs of various organisms [von Wettstein et al., 1984]. However, this structure of the central element is probably not universal. As Schmikol and Danchol (1995) have shown by electron microscopic imaging of SC of the beetle *Blaps cribrosa*, “TFs can pass as continuous filaments from one lateral element, through the central element, to the opposite lateral element”.

There is evidence that specific mutations preventing the SC assembly first and foremost disturb the formation of the SC central element. This leads to chromosome synapsis in prophase I and reduces the crossing-over frequency. Such mutations include c(3)G in *Drosophila melanogaster* [Smith and King, 1968], *asy* in wheat *Triticum durum* [LaCour and Wells, 1970] and maize *Zea mays* [Maguire and Riess 1996], *syl* in rye *Secale cereale* [Sosnikhina et al. 1992], *as6* in tomato [Havekes et al., 1994], and *zip1* in *S. cerevisiae* [Sym and Roeder, 1995]. Tung and Roeder, 1998 have shown that *zip1* allelic mutations result in various deletions from Zip1, and that the transversal filament length and SC central space width decrease depending on the size and position of a deletion.

![Fig. 2. Scheme of the central space of synaptonemal complex in mammals, and yeast as it follows from Heyting, 1996 (Model 1, see text). LE, lateral element; CE, transversal filament formed of two SCP1 molecules; C and N, designate C- and N-terminal domains, respectively. Central, rod-shaped domains of the protein molecules are in gray; dotted lines show presumable border of the axial cores of sister chromatids.](image-url)
We analyzed the data reported by Tung and Roeder (1998) along with the data on mammalian SCP1 [Liu et al., 1996; Meuwissen et al., 1992; Meuwissen et al., 1997; Dobson et al., 1994] and found that the central space width correlates well with the size (in amino acid residues) of the protein molecule forming a transversal filament (r = 0.85; p < 0.001) (Figure 3). The correlation was even better with the size of their central road-shaped coiled-coil domain (r = 0.90, p << 0.001) (Figure 3). A rather high coefficient of correlation means that the SC central space width strongly depends on the length of rod-shaped central domain of the proteins constituting SC transversal filaments [Grishaeva et al., 2002; Bogdanov et al., 2002a].

In D. melanogaster, the c(3)G (crossover suppressor on 3 of Gown) mutation causes the same ultrastructural alterations of SC as zip1 does in S. cerevisiae [Smith and King, 1968; King, 1970]. Hence the virtual protein encoded by c(3)G could be considered as a candidate transversal-filament protein. This role of the c(3)G product has been demonstrated by two methods.

In one case, we have identified in silico c(3)G as CG17604 (annotated by Celera Genomics, Inc.) on the NCBI molecular map of the D. melanogaster genome [Grishaeva et al., 2001; Bogdanov et al., 2002a], using elements of the procedure for searching for functional analogs (see below). With the Prot-Param Tool program, we showed that, among 78 putative genes present in the c(3)G region, only one, c(3)G, codes for a protein similar in physico-chemical properties to Zip1p and SCP1 as inferred. The C(3)G/CG17604 virtual protein consists of 900 amino acid residues and contains three domains, the central one forming an α-helical coiled coil with the length corresponding to the half width of the SC central space in D. melanogaster. The protein has other properties (isoelectric points of the domains, sites of the ATPase activity in the coiled-coil domain, etc.), which also testify to its structural analogy to Zip1p and SCP1.

In the other case, Page and Hawley (2001) have experimentally established the function of the c(3)G protein product. Recombinant constructs harboring c(3)G fused with the green fluorescence protein (GFP) gene were obtained and used to transform homozygotes for the c(3)G asynaptic mutation. In transformants, chromosomes restored the ability to form bivalents in meiotic cells. Moreover, green fluorescence was observed in the central region of the bivalents, fluorescent foci producing a ladder-like pattern similar to that of SC transversal filaments. These data testified to the role of the c(3)G product in formation of the SC transversal filaments.

Thus, our conclusions based on the computer analysis [Grishaeva et al., 2001] coincided with those independently made by Page and Hawley (2001) on the basis of molecular-genetic experiments. Hence, we expected that a similar strategy would be applicable for searching other proteomes for proteins analogous to SCP1, Zip1p, and CG17604. We sought the virtual SC proteins in organisms with completely sequenced genomes, including nematode C. elegans [The C. elegans Sequencing Consortium, 1998] and plant A. thaliana [The Arabidopsis Genome Initiative, 2000]. As the result of our search, one A. thaliana and two C. elegans proteins were found. In the case of C. elegans, the presence of the two such proteins was shown to agree with the electron microscopic data on the SC structure.

However, one C. elegans protein, which is undoubtedly involved in the organization of the SC central space and probably forms transversal filaments [MacQueen et al., 2002], escaped detection by our method. Possible causes are discussed below. This allowed us to evaluate the extent to which the utility of our software is limited. Yet the general search strategy seems to be practicable.

METHODS

As resources, we used NCBI databases on the known and putative genes and proteins of S. cerevisiae, D. melanogaster, A. thaliana, and C. elegans (http://www.ncbi.nlm.nih.gov/). In addition, TAIR AGI Information (http://www.arabidopsis.org/home.html) and WormBase (http://www.wormbase.org/) were
Fig. 3. The relationship between the SC central space width (nm) and size of the protein molecules (number of amino acid residues), forming the transverse filaments in SC central space. Solid circles correspond to the yeast protein Zip1 in the wild-type or in mutants. The mutant proteins have internal deletions or duplications of various lengths [Tung and Roeder, 1998]. Solid rectangles correspond to mammalian proteins SC1 from mouse, rat and human [Liu et al., 1996; Heyting, 1996]. The sloping solid line shows the regression for the shown symbols. Dashed lines show the 95% confidence interval of the regression line. a) regression line for the SC central space width versus size of the total protein molecule. Coefficient of correlation, $r = 0.85; p < 0.001$. b) regression line for the SC central space width versus size of the coiled-coil part of the central domain. All the amino acid sequences, which do not participate in the coiled-coil, have been subtracted from the abscissa length. Coefficient of correlation, $r = 0.90; p < 0.001$.

used for *A. thaliana* and *C. elegans*, respectively.

The analysis of protein domain structure and the search for structural analogs were performed using CDART (Conserved Domain Architecture Retrieval Tool, http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=trps). This was supplemented by prediction of the physico-chemical properties of proteins with the use of the ProtParam tool provided by ExPASy Molecular Biology Server (Expert Pro-

In addition, we constructed two calibration plots (Figure 3a and b).

In Figure 3a, the SC width in mammals and yeast was plotted against the size (amino acid residues) of SCP1 and Zip1. The data were taken from published works [Meuwissen et al., 1992; Meuwissen et al., 1997; Dobson et al., 1994; Liu et al., 1996; Dong and Roeder, 2000]. In the case of Zip1p, we analyzed not only the wild-type, but also several products of mutant Zip1 alleles differing in Zip1p size [Tung and Roeder, 1998]. The coefficient of correlation between protein size and SC width was found to be high, $r = 0.85$ at $p < 0.001$.

Another plot (Figure 3b) shows the SC central space width versus the size of the coiled coil within the central domain. In this case, we also used Zip1 mutants with variously sized deletions or duplications within the coiled-coil domain [Tung and Roeder, 1998; Dong and Roeder, 2000]. The correlation coefficient proved to be even higher, $r = 0.90$ ($p < 0.001$). So high a correlation suggests a strong dependence between the SC central space width and the length of the coiled-coil domain.

The computer searching strategy described below was much the same as that employed in the previous search for the virtual D. melanogaster c(3)G+ product [Grishaeva et al., 2001; Bogdanov et al., 2002a]. The only difference was that the entire genome was analyzed in the case of A. thaliana and C. elegans, whereas the search was restricted to the known region of chromosome 3R harboring c(3)G in the case of D. melanogaster.

The A. thaliana and C. elegans genome databases were searched for open reading frames (ORFs) which potentially code for proteins similar in size and in secondary structure to yeast Zip1p, mammalian SCP1, and D. melanogaster CG17604 / C(3)G. Advantage was taken of the fact that the width of the SC central space in A. thaliana and in C. elegans is within the range of its variation among yeast, mammals, and D. melanogaster [von Wettstein et al., 1984].

The analysis included several steps. First, we employed the CDART program to select the deduced proteins with the domain structure similar to that of the mouse SCP1. The proteins selected by the program were analyzed with respect to the following criteria:

1. correspondence of the size of the total protein molecule, and especially of its rod-shaped central domain, to the width of the SC central space;
2. presence of an extended coiled coil in the central domain and of an ATPase site in the coiled coil, which are characteristic of SCP1;
3. absence of a coiled coil from the N- and C-terminal domains;
4. physico-chemical properties of the C-terminal domain;
5. correspondence of the selected protein to the morphological features of SC in the given organism.

The criteria are considered in detail below.

**Criterion 1.** The size (number of amino acid residues) of a protein was considered appropriate when it fell within the 95% confidence interval on the plot of central space width versus whole molecule size of the known transversal-filament proteins (Figure 3a).

**Criteria 2 and 3.** The coiled-coil domain was predicted and its size determined using the ISREC program. We selected the hypothetical proteins with the coiled-coil length falling within the 95% confidence interval on the plot (Figure 3b) which served as a calibration plot at the second step of protein selection. At the third step, the secondary structure of C- and N-terminal domains was analyzed with the ISREC program.

**Criterion 4.** The physico-chemical properties of candidate proteins were analyzed using the ProtParam tool. We selected proteins having the C-terminal domain with an isoelectric point of at least 8. This requirement follows from the properties of the C-terminal domain in mammalian SCP1 and in yeast Zip1p.
analyzed with the ProtParam tool. Another reason for selecting proteins with a basic C-terminal domain was that the C-terminal domain of SCP1 binds to DNA [Heyting, 1996] and DNA-binding domains should be basic.

**Criterion 5.** To decide whether a selected protein may be involved in the SC transversal filaments, we considered the reported ultrastructural features of SC in the organisms under study.

The set of computer programs employed had a substantial drawback. At the first step (criterion 1), we needed a program that screens databases and selects proteins by secondary structure of their domains. This was essential because the similarity between SC proteins SCP1, Zip1p, and CG17606/C(3)G is restricted to their domain composition and domain secondary structure. For this purpose the Secondary Structure Alignment Server (http://lutece.rutgers.edu/~piech/SSAS/ss.html) potentially could be used. Yet this program proved to be inapplicable because it could not be employed in long-term Internet sessions, which are necessary for a database search.

Hence, the CDART program was employed at the first step. This program selects proteins by primary structure consensus. As such, we chose the ATPase consensus of the SCP1 central domain.

The results of the analysis are reported in the next section.

**RESULTS AND DISCUSSION**

The results obtained at each step of computer-assisted search and analysis of putative proteins with unknown functions and the domain structure similar to that of the mouse SCP1 are shown in Table 1. It should be noted that the proteins selected included rat and human SCP1, as well as Zip1p, and CG17604 (see Introduction). These proteins and those with known functions are omitted in Table 1, which considers only putative proteins.

<table>
<thead>
<tr>
<th>In silico search step</th>
<th>Number of proteins selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Screening with the use of CDART</td>
<td>Circa 500</td>
</tr>
<tr>
<td>2. Selection by protein size (criterion 1)</td>
<td>70</td>
</tr>
<tr>
<td>3. Analysis of secondary structure (criteria 2 and 3)</td>
<td>10</td>
</tr>
<tr>
<td>4. Selection by physico-chemical properties (criterion 4)</td>
<td>4</td>
</tr>
</tbody>
</table>

The putative proteins selected at the last step included one of imperfect fungus *Dictostelium discoideum*, one *A. thaliana* protein, and two *C. elegans* proteins. The *D. discoideum* protein was excluded from further analysis, because we had no information on meiosis in this organism and because only about 92% of the *D. discoideum* genome had been sequenced by that time. Hence, only the *A. thaliana* and *C. elegans* proteins were considered. Important parameters of these proteins and of SCs of the two organisms are presented in Table 2. In the same table the parameters were compared with those reported for other organisms studied in this respect.

**Synaptic protein of *A. thaliana***

While our search for a specific c(3)G gene sequence in *Drosophila* was restricted to a relatively small genome region and, consequently, a relatively small (78) number of candidate genes, identification of *A. thaliana* genes functionally similar to SCP1 and ZIP1 was complicated by a lack of mutations affecting the SC formation in this plant. Hence, it was necessary to extend the search to the entire *A. thaliana* ge-
nome.
We failed to find any published data on the width of the SC central space in *A. thaliana*. We noted, however, that the SC width is 90–120 nm in the vast majority of phylogenetically distant organisms, in-

Table 2

<table>
<thead>
<tr>
<th>No.</th>
<th>Species, wild-type or mutant protein</th>
<th>SC central space width, nm</th>
<th>Protein (domain) size, amino acid residues</th>
<th>Isoelectric point (pI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total protein</td>
<td>Coiled-coil</td>
</tr>
<tr>
<td>1</td>
<td><em>M. musculus</em>, SCP1</td>
<td>100</td>
<td>993</td>
<td>713</td>
</tr>
<tr>
<td>2</td>
<td><em>H. sapiens</em>, SCP1</td>
<td>100</td>
<td>973</td>
<td>677</td>
</tr>
<tr>
<td>3</td>
<td><em>R. norvegicus</em>, SCP1</td>
<td>100</td>
<td>946</td>
<td>717</td>
</tr>
<tr>
<td>4</td>
<td><em>S. cerevisiae</em>, Zip1</td>
<td>115</td>
<td>875</td>
<td>632</td>
</tr>
<tr>
<td>5</td>
<td>Zip1-m2*</td>
<td>63</td>
<td>583</td>
<td>285</td>
</tr>
<tr>
<td>6</td>
<td>Zip1-mc1*</td>
<td>49</td>
<td>484</td>
<td>170</td>
</tr>
<tr>
<td>7</td>
<td>Zip1-mc2*</td>
<td>101</td>
<td>776</td>
<td>634</td>
</tr>
<tr>
<td>8</td>
<td>Zip1-n1*</td>
<td>118</td>
<td>732</td>
<td>578</td>
</tr>
<tr>
<td>9</td>
<td>Zip1-nm1*</td>
<td>118</td>
<td>767</td>
<td>512</td>
</tr>
<tr>
<td>10</td>
<td>Zip1-2XH2**</td>
<td>153</td>
<td>1012</td>
<td>799</td>
</tr>
<tr>
<td>11</td>
<td>Zip1-3XH2**</td>
<td>189</td>
<td>1280</td>
<td>1067</td>
</tr>
<tr>
<td>12</td>
<td><em>D. melanogaster</em>, CG17604/C(3)G</td>
<td>109</td>
<td>744</td>
<td>495</td>
</tr>
<tr>
<td>13</td>
<td><em>A. thaliana</em>, AAD10695</td>
<td>100–120</td>
<td>923</td>
<td>476</td>
</tr>
<tr>
<td>14</td>
<td><em>C. elegans</em>, Q11102***</td>
<td>70–85</td>
<td>1132</td>
<td>938</td>
</tr>
<tr>
<td>15</td>
<td><em>C. elegans</em>, T26844</td>
<td>70–85</td>
<td>1083</td>
<td>536</td>
</tr>
<tr>
<td>16</td>
<td><em>C. elegans</em>, T27907</td>
<td>70–85</td>
<td>772</td>
<td>460</td>
</tr>
<tr>
<td>17</td>
<td><em>C. elegans</em>, Z81586***</td>
<td>70–85</td>
<td>484</td>
<td>460</td>
</tr>
<tr>
<td>18</td>
<td><em>C. elegans</em>, WP:CE17456 / SYP-2</td>
<td>70–85</td>
<td>213</td>
<td>50</td>
</tr>
<tr>
<td>19</td>
<td><em>C. elegans</em>, F26D2.2 / SYP-1</td>
<td>70–85</td>
<td>489</td>
<td>355</td>
</tr>
</tbody>
</table>

Synaptonemal complex width and molecular parameters of proteins which were experimentally shown (see text for references) or assumed (this work) to form the SC transversal filaments. Deletion (*) and duplication (**) derivatives of yeast Zip1 are indicated. Two *C. elegans* virtual proteins selected in this work are marked with ***. SYP-1 has been reported by MacQueen et al., 2002.

Including higher plants. Taking this value for criterion 1 (see Methods), we searched the total *A. thaliana* proteome and found the virtual protein AAD10695 which is similar in size and in domain structure to Zip1p and to SCP1 (Table 2, Figure 4). As already mentioned, the basic character (pI > 8) of the C-terminal domain is an essential feature of the SC transversal-filament proteins of yeast, mammals, and *Drosophila*. The *A. thaliana* AAD10695 protein met this requirement.

Thus, like *D. melanogaster*, *A. thaliana* has only one putative functional analog of mammalian SCP1 or yeast Zip1p. This finding is consistent with the major data on comparative proteomics of highly spe-
Specific Features of the SC Genes and Proteins in *C. elegans*

In this section, we discuss two sets of data, one obtained by our *in silico* search and another reported elsewhere.

First, we noted the unusual SC morphology in *C. elegans*. On some electron microphotographs, *C. elegans* SC differs in ultrastructure from yeast or mammalian SC. The distinctive feature of *C. elegans* SC is a lack of a clearly defined central element. While the central element is seen as an electron-dense filament

![Fig. 4. The relationship between the SC central space width and size of the central domain of protein molecules forming SC transverse filaments. Wild-type or mutant yeast Zip1 studied experimentally by Tung and Roeder (1998) (solid circles); virtual proteins assumed by Grishaeva et al. (2001) for *D. melanogaster* (open circle), and in this work for *A. thaliana* (open rectangle) and *C. elegans* (ellipsoids). One of the solid ellipsoids corresponds to the length of the central domain of Z81586 virtual protein, another to 0.5 length of the central domain of Q11102 virtual protein, and open ellipsoid cooresponds to the central domain of F26D2.2 / SYP-1 protein (see text). Other designations are as in Figure 3.](image_url)

located in the central space in many organisms [Zickler and Kleckner, 1999], in *C. elegans*, the transversal filaments do not cross the central element on their way from one lateral element to another [Dernburg et al., 1998]. This suggests a different type of molecular organization for *C. elegans* transversal filaments. If we take as Model 1 the organization of transversal filaments in yeast, mammals, and *Drosophila* 1, then another model, Model 2, describes this structure in *C. elegans*. Probably, a long molecule, rather than two molecules with a zipper-like connection, extends from one to another lateral element (Model 2, see Figure 5). Schmekel et al., 1993 and Schmekel and Daneholt (1995) proposed this model for *B. cibrosa* transversal elements. Electron microphotographs of SC allow a similar interpretation of the TF morphology in *C. elegans*. If this is the case, both terminal domains of the protein forming TFs in *C. elegans* should be alkaline (basic) to associate with DNA in the SC lateral elements, and the coiled coil must be twice longer than in SCP1 or Zip1.
In view of this, we searched the *C. elegans* proteome for proteins fitting either model. Three ORFs were found to code for potential SCP1 and Zip1 analogs T26844, T27907, and Q11102 (Table 2). Each of these virtual proteins has a coiled coil suiting either Model 1 or 2. However, the C-terminal domain of T26844 or T27907 has low pI (5.3–5.7) and, therefore, cannot interact with DNA. Hence these proteins were rejected. The third one, Q11102, suits Model 2, half size of its coiled-coil falling within the 95% confidence interval on the regression plot (Figure 4). Thus, the full-length coiled-coil corresponds to the complete central-space width of *C. elegans* SC. Both terminal domains of Q11102 are basic (see Table 2), meeting the other requirement of Model 2.

Another set of data was obtained from the WormBase and published works. Virtual protein products of several *C. elegans* ORFs have been annotated as SC proteins in the WormBase. Basing on the WormBase data, protein WP:CE17456 encoded by *syp2* was identified as a component of SC central space in *C. elegans* [Colaiacovo et al., 2001]. However, our analysis of the protein with the ProtParam program showed that the deduced sizes of the entire protein (213 amino acid residues) and of its coiled-coil central domain (50 residues only) are insufficient for the protein to form the SC transversal filaments (Table 2), assuming that these are organized as in Figure 1. Hence, we think that WP:CE17456 is not a transversal-filament protein.

In addition, Proteome Inc. has reported that the *C. elegans* protein Z81586 is analogous to Zip1p. Using all criteria described in the Methods section, we found that Z81586 may indeed be a functional analog of Zip1p as judged from its total size, domain structure, size of the central coiled-coil domain, and physico-chemical properties of the C-terminal domain (see Table 2). The protein Z81586 fits Model 1 (Figure 5).

To decide between the two proteins, we analyzed the genetic surroundings of the two candidate genes with the WormBase software. The region was $-4 \pm 1$ on the genetic map of chromosome 1 in the case of the Z81586 gene and $-10 \pm 3.5$ of the X chromosome in the case of the Q11102 gene. Neither region proved to contain a gene annotated in WormBase as responsible for normal meiosis. Thus, the decision is
still unclear. Based on the unusual morphology of the SC central space in *C. elegans*, we believe that Q11102 is the *C. elegans* functional analog of SCP1 and Zip1p. We assume that long molecules of this protein extend over the SC central space and form the transversal filaments in *C. elegans*, whereas yeast and mammals have transversal filaments formed by two protein dimers facing each other.

After the above part of our work was completed and presented at the 3rd International BGRS Conference (July 14–20, Novosibirsk) [Bogdanov et al., 2002b], a novel information appeared. MacQueen et al. (2002) have experimentally found that the *syp-1* gene of *C. elegans* (chromosome V) codes for protein F26D2.2 of 489 amino acid residues. The size and other features of SYP-1 (GenBank accession no. AF515883) were deduced from the revised gene structure which has been derived from sequencing cDNAs obtained by RT-PCR and 5'RACE and is consistent with the structure suggested by the EST alignment data present in WormBase. The region between residues 48 and 402 of the protein has been predicted to contain α-helical coiled coils. On immune staining, antibodies to N- and C-terminal parts of SYP-1 have been detected at the interface between synapsed homologs along their full length in meiotic prophase-I cells of the worm. This is a decisive evidence for the role of SYP-1 in forming the SC central space in *C. elegans*. MacQueen et al. (2002) noted that SYP-1 is approximately half the size of Zip1 or SCP1. Since the width of *C. elegans* SC is comparable with that of mammals, yeast and *Drosophila*, MacQueen et al. (2002) have assumed “that SYP-1 exhibits an organization within the SC central region that is distinct from that of Zip1, or SCP1, α-helix and/or that additional structural proteins work in conjunction with SYP-1 to form the mature central region of the *C. elegans* SC.”

In view of the above findings, we studied the *C. elegans* SC on electron microphotographs published by several authors, and observed that the ultrastructure of SC central space/region is rather unusual in *C. elegans*. Two types of transversal filaments may be found on a single microphotograph, one representing a continuous connection between the lateral elements (Model 2) and another resembling zipper-like teeth. The former corresponds to the Q11102 protein predicted in this work, while the latter may consist of Z81586 or/and SYP-1. Taking the consideration of MacQueen et al. (2002) into account, both proteins can take part in the construction of transversal filaments to fit Model 1. This statement is true, if the width of the central SC space in *C. elegans* is about 100 nm, as Mac Queen et al. assume. If the width of the central space is within the range of 75–85 nm, a single protein molecule like Z81586 or SYP-1 can cover this distance and fits well to the regression line at Figure 4. Regrettfully, precise estimation of the width of the central space in *C. elegans* is problematic.

The primary aim of this article was to describe the strategy of the *in silico* search for unknown proteins that potentially play a given part, by properties essential for this part to be played. We restricted the search to structural proteins, which must possess certain properties to fill a certain cell compartment and to perform a certain mechanical function. It was clear *a priori* that the result would depend on the quality of software programs available to us. Indeed, with the function specified, we found new candidate proteins in *A. thaliana* and *C. elegans*, but overlooked one *C. elegans* protein which has more recently been revealed in molecular genetic experiments [MacQueen et al., 2002]. We failed to detect this protein because one of the programs employed was poorly suitable for selecting proteins having a coiled-coil domain (step 1). However, we think that the general search strategy is practicable.

**CONCLUSION NOTES**

Complete genome sequencing in three model multicellular organisms (*D. melanogaster*, *C. elegans*, and *A. thaliana*) and a design of specific computer software have allowed not only the detection of new ORFs, but also the prediction of the structural and functional parameters of their protein products. Though not free from drawbacks, the two major methods of searching for orthologs by sequence homology and by structural domains can be used to initially select the candidate proteins. To improve the reliability of iden-
ifying the structural and functional analogs, we propose an additional method for in silico testing the candidate proteins for compliance with the ultrastructural requirements of the relevant organelles. Our method proved to be effective in the case of an intermediate protein of the SC transversal filaments. Such a protein has been identified in D. melanogaster, providing an experimental support to our findings. Further studies will show whether the method is applicable to other proteins forming various cell structures in other organisms.

It would be of interest to analyze the assembly of SC or another cell structure with a strongly specified spatial arrangement in various eukaryotic kingdoms if the relevant proteins lack sequence homology but act as structural and functional analogs. This might be the case with proteins forming spindle pole bodies in fungi, centrioles in animals, kinetochores in all kingdoms, etc.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 02-04-48761).

REFERENCES

2428–2442.


