

Aromatic Di-Alanine Repeats (AdAR) Are Structural Motifs Characteristic of the Soluble *N*-Ethylmaleimide-Sensitive Factor Attachment Protein (SNAP) Family

Franz Bruckert,^{1*} Thomas Casavant,² and Michel Satre¹

¹Laboratoire de Biochimie et Biophysique des Systèmes Intégrés, UMR 5092 Université Joseph Fourier/CEA/CNRS, Département de Biologie Moléculaire et Structurale, CEA-Grenoble, France

²Coordinated Laboratory for Computational Genomics and Parallel Processing Laboratory, Department of Electrical and Computer Engineering, University of Iowa, Iowa City, Iowa

ABSTRACT The aromatic di-alanine repeat is a novel 12-amino acid-long motif constituting alternate small and large hydrophobic residues that mediate the close packing of α -helices. A hidden Markov model profile was constructed from the motifs initially described in Soluble *N*-ethyl maleimide-sensitive factor attachment proteins (SNAP), a family of soluble proteins involved in intracellular membrane fusion. Scanning different sets of protein sequences showed unambiguously that this profile defines a structural motif independent of the tetratricopeptide repeat, another widespread α -helical motif. In addition to SNAP, aromatic di-alanine repeats are found in selective LIM homeodomain binding proteins (SLB) and in proteins from the *Pyrococcus* and *Archaeoglobus* prokaryotes. **Proteins 2001;45:40–46.** © 2001 Wiley-Liss, Inc.

Key words: membrane fusion; α -helix; tetratricopeptide repeat

INTRODUCTION

Intracellular membrane fusion is regulated by a superfamily of specialized membrane proteins called SNAREs.¹ Separate “donor” and “acceptor” membranes carry different SNARE partners. Bilayer fusion is coupled to the transition between a *trans*-SNARE complex, in which the SNAREs are bound to separate bilayers, to a *cis*-SNARE complex, where the SNAREs transmembrane domains are on the same membrane.^{2,3} The dissociation of the *cis*-SNARE complex after fusion to retrieve fusion competent SNARE partners is an energy-demanding process catalyzed by NSF, an ATPase, and protein adapters called SNAPs. These proteins were indeed discovered as obligate and ubiquitous players of vesicular transport. NSF is a homo-hexamers and requires SNAP to bind to the four-helix bundle of the SNARE complex.⁴ Two subfamilies of SNAP have been described; α/β and γ . Except for yeast, all eukaryotes seem to possess one isoform of each subfamily. Mammals have another isoform of the α/β subfamily present in the brain.^{5,6} Despite its ubiquitous presence, γ -SNAP does not appear to play a role in SNARE dissociation. The functional NSF-SNAP-SNARE complexes contain therefore three α -SNAPs, and an optional γ -SNAP molecule.^{7,8}

The structure of the yeast α -SNAP member, Sec17p, has been solved.⁹ The protein is composed of 14 α -helices. Nine of them are folded in an antiparallel manner and form an elongated twisted sheet structure. The remaining ones form a globular C-terminal domain. Biochemical data and electron microscopy imaging of individual SNAP-SNARE complexes showed that the N-terminus of SNAP binds to the SNARE complex, probably near the membrane anchor.^{10,11} The C-terminal globular domain of the protein may also interact with the SNARE complex, but mutagenesis and peptide competition experiments strongly suggested that this domain interacts with NSF, stimulating its ATPase activity.^{12,13} SNAP seem therefore to transmit forces between the NSF hexamer to the *cis*-SNARE complex.¹⁴

The ordered three-dimensional (3D) structure of the SNAP twisted sheet domain is reflected in the primary sequence. Ordway et al.¹⁵ recognized the presence of repeats similar to the tetratricopeptide repeat (TPR), a 34-aa sequence motif present in a great variety of proteins that often mediates interaction with exogenous α -helical domains.^{16,17} However, Weidenhaupt et al.⁸ suggested that the basic repeat unit in SNAP was smaller, corresponding to a new, specific heptad structure. The first heptad position (a) is mostly occupied by aromatic (F, Y, W) or bulky hydrophobic (L, I, M) residues, while alanines or other small aa such as G, C, T, or S are predominantly found at positions (d) and (e) of the heptad. An obvious example of such a motif was the segment 34–48 of *D. discoideum* α -SNAP: YDDAASD YTKAANL F. This heptad repeat was hence named aromatic di-alanine repeats (AdAR). Although no consensus sequence can be defined for TPRs, AdAR, and TPR motifs have in common the

Abbreviations: aa, amino acid; AdAR, aromatic di-alanine repeat; TPR, tetratricopeptide repeat; HMM, hidden Markov chain model; NSF, *N*-ethylmaleimide-sensitive factor; SNAP, soluble *N*-ethylmaleimide-sensitive factor attachment protein; SNARE, SNAP receptor.

Grant sponsor: Université Joseph Fourier-Grenoble; Grant sponsor: Centre National de la Recherche Scientifique; Grant sponsor: Commissariat à l’Energie Atomique.

*Correspondence to: Franz Bruckert, DBMS/BBSI (UMR5092 CNRS), CEA-Grenoble, 17 rue des Martyrs, 38054 Grenoble Cedex 09, France. E-mail: fbruckert@cea.fr

Received 27 February 2001; Accepted 14 May 2001

characteristic presence of alternating large and small hydrophobic residues.¹⁶ It could therefore be possible that the antiparallel α -helical repeats of SNAP represent an insignificant variation within the large TPR-containing protein family.

We therefore undertook a quantitative sequence analysis study to answer the following questions: Are SNAP true members of the TPR family? How close are the TPR and AdAR motifs? Is the AdAR motif characteristic of SNAP? For this purpose, we defined different protein sequence collections and sorted them according to their scores in Hidden Markov Chain Models (HMMs) trained on either AdAR or TPR motifs. HMMs are well suited to this task because the scores calculated, once calibrated on the size of the data set, are directly related to a meaningful likelihood, i.e., the probability that the motif under consideration did not occur by chance. Furthermore, the HMM technique is more flexible and allows uncovering motif occurrences not contained in the initial training set. An optimal HMM may therefore be constructed by an iterative cycle of training and searching cycles, exploring most of the motif space.

HMM-based search tools¹⁸ differ fundamentally from sequence similarity search tools such as BLAST¹⁹ in that they do not use information from a “generic” substitution matrix. Such matrices are normally constructed from very general characteristics of aa substitution properties, or based on assumptions about the evolutionary frequency of occurrence of point mutations. At the core of an HMM are Bayesian statistics gathered from a known multiple alignment for a motif of interest. A Markov model is then constructed in which the state probabilities and state transition probabilities are derived from the specific attributes of the multiple alignment being considered as representative of the motif. Once such a model has been constructed, it can be used to search a database of sequences by threading the subject sequences through the model and generating a match “score” based on the likelihood that the model fits the subject. It should be noted that HMMs are neither generally inferior, nor superior, to similarity-based searches such as BLAST. HMMs do perform a different function fundamentally, though. It is possible to “BLAST” search a database with only a single subject sequence, or with a clear consensus sequence. However, with HMMs, it is possible to consider a motif as represented by a collection of related sequences (though nonidentical), as a template for searching. The power of HMMs is that at each residue position, the likelihood of observing a particular amino acid or nucleotide can be considered in the context of a set of training data. BLAST searches cannot adapt to a specific set of related sequences as the basis for a search. Therefore, in the case of searching for SNAP and TPR domains, the choice of the HMM approach is an obvious one.

METHODS

Generation of the Protein Sequence Sets Used for HMM Comparison

The aa sequences of the SNAP used in this study are available at the EMBL, GenBank or SwissProt databases

under the following accession numbers: *Saccharomyces cerevisiae* (M93104), *Pichia pastoris* (AF216957), *Dictyostelium discoideum* (AU038530 and AF157836), *Drosophila melanogaster* (U09374 and AF177988), *Arabidopsis thaliana* (AF177989 and AF177990), *Vitis vinifera* (AB001375), *Coprinus cinereus* (U79186), *Loligo pealei* (X82847), *Mus musculus* (X61455, P28663), *Rattus norvegicus* (X9968), *Bos taurus* (P81125, P81126, P81127), and *Homo sapiens* (U39412, U78107). The set of TPR proteins derived from the SMART alignment²⁰ was retrieved at the PFAM²¹ databank under the accession number PF00515. The 38th release of SwissProt was downloaded directly from the ExpASy site.

Analysis of Sec17p Structure

The crystal structure of Sec17p obtained by Rice et al.⁹ (PDB accession number 1QQE) was analyzed using Swiss-PdbViewer^{22,23} (software version 3.5, freely available at <http://www.expasy.ch/spdbv/>) run on an iMac Apple computer. The water accessibility of each amino acid was computed by using the corresponding function of the software. Briefly, accessibility is defined as the ratio of the exposed surface area of an aa X in the actual structure to the maximum possible exposed surface area of the same aa. This maximum surface area is the surface accessible to water of X in a pentapeptide GGXGG in extended conformation. For the calculation, 1.4 Å is added to the radius of every atom, to take into account the size of a water molecule. The distance between two aa residues was defined as the minimum distance between two side-chain atoms of each aa.

RESULTS AND DISCUSSION

Constructing the Hidden Markov Model Recognizing Aromatic di-Alanine Repeats (AdAR) in SNAP Proteins

Initially, 61 12–20-aa-long AdAR motifs were defined by inspection in 14 full-length SNAP sequences.⁸ These motifs were aligned using the ClustalW program²⁴ and trimmed of extra aa extending on both sides of a twelve aa core. These 61 aligned sequences of 12 aa were used as input for the HMMER 2.1.1 suite of software from Sean Eddy’s group at the University of Washington in St. Louis¹⁸ (<http://hmmmer.wustl.edu>) to build the first of our HMM AdAR models (AdAR model I). The program *hmm-build* was used to construct this first model, and *hmmcalibrate* was used to calibrate *E*-value scores.

This model was used to scan the 38th release of SwissProt protein databank using *hmmsearch*. Top scores distinguished SNAP and 4–6 motifs were predicted per SNAP. To scale the results, proteins similar to bovine α -SNAP were searched in the same databank, using BLAST. All proteins recognized as SNAP in this way were also found by the HMM model. A more comprehensive BLAST search of SNAP was then conducted within the nr (not redundant) division of Genbank, and these SNAP sequences were examined with HMM AdAR model I. Positive scores were obtained with all sequences, each of which contained more than four motif occurrences. At this stage, the number of

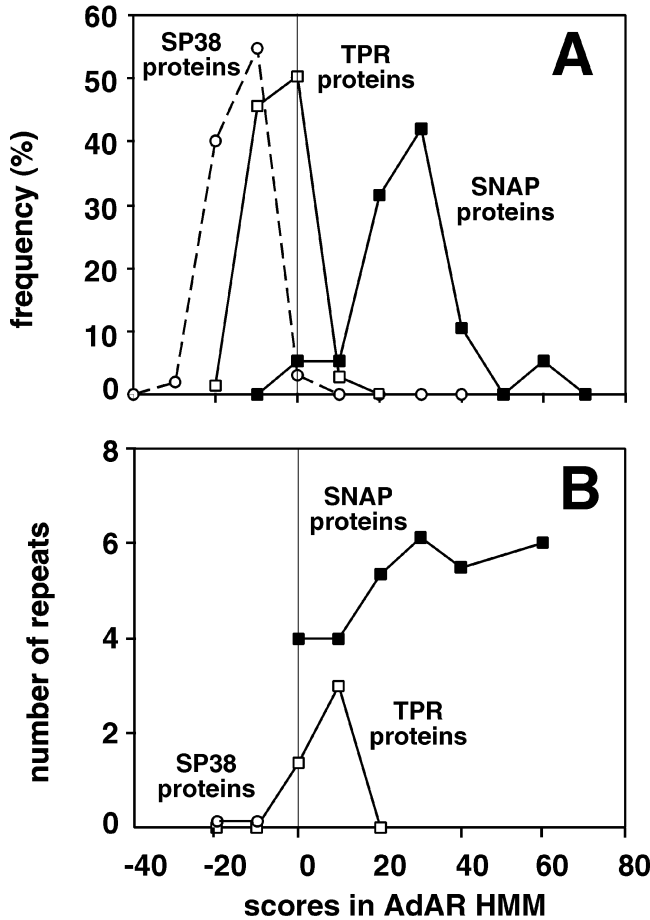


Fig. 1. Results of the aromatic di-alanine repeats (AdAR) hidden Markov chain model (HMM) model. The AdAR HMM model was built as described in the text and was applied to SwissProt 38th release (80,000 sequences), a set of tetratricopeptide repeat (TPR)-containing proteins (460 sequences) and 19 soluble N-ethylmaleimide sensitive factor attachment protein (SNAP) sequences. **A:** Frequency histogram of the scores obtained. Each class corresponds to a score interval of 10. **B:** Mean number of repeats per protein sequence within each class, as a function of the class score.

AdAR motifs amounted to 87 in 19 full-length or partial SNAP sequences.

A second HMM model (AdAR model II) was generated with these 87 motifs using the HMMER 2.1.1 program. This model detected 106 motifs in the 19 GenBank SNAP sequences. In the 38th release of the SwissProt protein database, the 8 SNAP ranked at top scores, totaling 4–7 motifs per protein sequence. A third iteration of model generation did not increase significantly the number of AdAR motifs detected in these sets of protein sequences. HMM AdAR model II was therefore used for all subsequent analyses and is referred to simply as AdAR HMM.

Comparison Between HMM Models for Aromatic Di-Alanine Repeats (AdAR) and Tetratricopeptide Repeats

Among the 80,000 proteins of the 38th SwissProt release, a set of about 30 non-SNAP gave highly significant scores (superior to 9). One-third of these proteins were

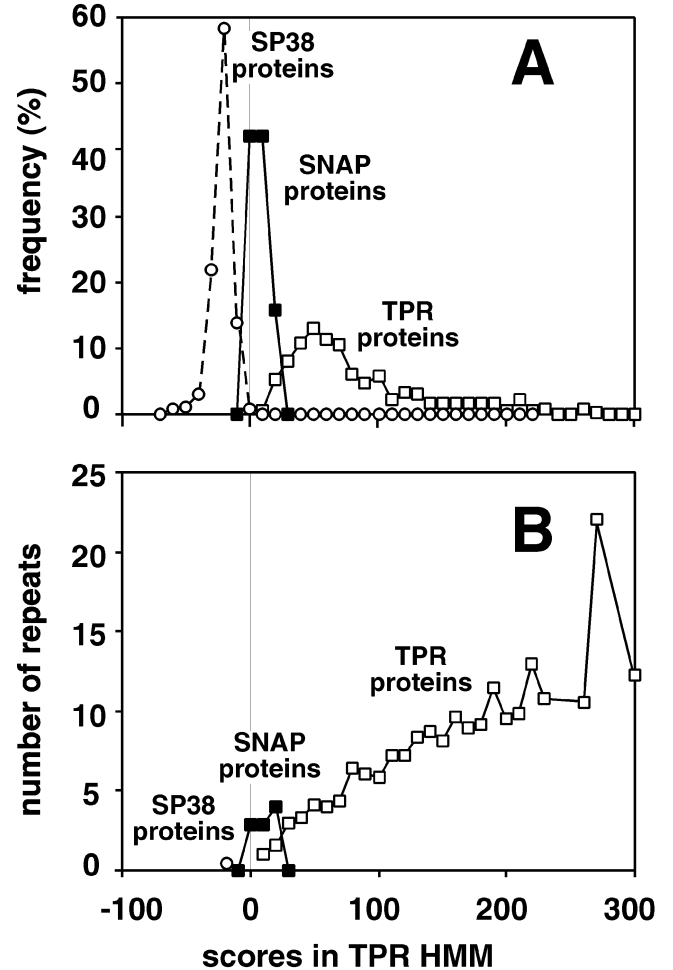


Fig. 2. Results of the tetratricopeptide repeat (TPR) hidden Markov chain model (HMM). A TPR model was built from a collection of TPR-containing proteins as described in the text. The three sets of protein sequences used are defined in Fig.1. **A:** Frequency histogram of the scores obtained. Each class corresponds to a score interval of 10. **B:** Mean number of repeats per protein sequence within each class, as a function of the class score.

members of the tetratricopeptide repeats (TPR) family. Two of them even contained more than 3 AdAR motifs: the rat UDP-*N*-acetylglucosamine-peptide *N*-acetylglucosaminyltransferase (OGT1_RAT, 6 motifs) and the yeast Yhr7p hypothetical protein (YHR7_YEAST, 4 motifs). We therefore undertook a detailed comparison of results obtained by applying AdAR or TPR HMMs on various protein sequence sets. For this purpose, a set of representative TPR proteins was obtained from the PFAM database. Three sets of proteins were therefore analyzed with AdAR and TPR HMMs: 80,000 sequences of SwissProt, 460 TPR-containing protein sequences and 19 SNAP sequences, designated as “SwissProt38”, “TPR proteins” and “SNAP” respectively. The results are shown in Figures 1 and 2.

When analyzed with AdAR HMM, “TPR proteins” gave an average score of -5, close to that of “SwissProt38” (average score = -10) but well below that of “SNAP” (average score = +30) [Fig. 1(A)]. Furthermore, the num-

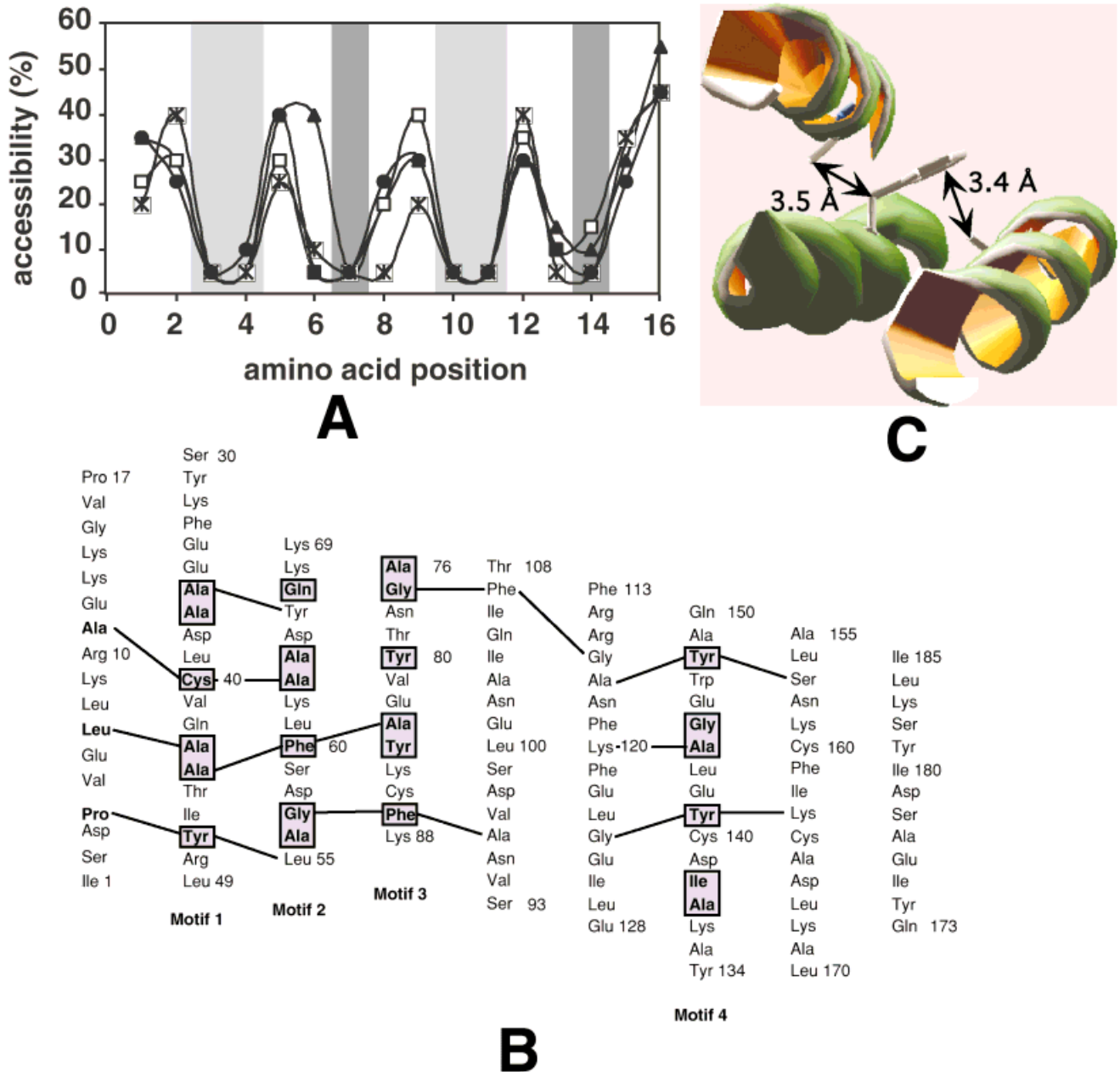


Fig. 3. Topology of the aromatic di-alanine repeats (AdAR) motifs in the yeast α -sensitive factor attachment proteins (SNAP) (Sec17p) structure. **A:** Amino acids of the AdAR motifs are buried in Sec17p interior. The accessibility of each aa was calculated at each position of Sec17p sequence using Swiss-PdbViewer 3.5. The results are shown for the sequences 34–49 (\blacktriangle), 54–69 (\square), 74–89 ($*$), 135–150 (\bullet) that contain the four AdAR motifs detected in Sec17p by the AdAR hidden Markov chain model (HMM) plus a 2-aa extension on both sides. The characteristic positions of the di-alanine and aromatic repeats in the AdAR motifs are boxed in light and dark gray, respectively. **B:** Backbone of connecting junction sites in Sec17p protein. For each aa of the AdAR motifs detected in Sec17p sequence by the AdAR HMM, closest neighbors were searched using Swiss-PdbViewer 3.5. aa having side-chain atoms closer than 3.5 Å are connected by a line. For clarity, aa connections within the same helix are omitted. The aa included in the 9 helices of the N-terminal and twisted sheet domains are written in separated columns. The characteristic aa of the four Sec17p AdAR motifs are in bold and boxed. **C:** Example of a connecting junction site between adjacent Sec17p helices. The portions of the three α -helices corresponding to the AdAR motifs 1, 2, and 3 of Sec17p are represented as ribbons (aa 36–47, 56–67, and 76–87). Motif 1 is on the bottom right. The side-chains of Ala 44, Phe60, and Ala83, which belong to the AdAR motifs 1, 2, and 3, respectively, are shown to exemplify their close proximity.

ber of AdAR motifs per protein sequence never exceeded an average value of 3 for “TPR proteins,” whereas it was higher than 4 for all “SNAP” examined so far [Fig. 1(B)]. Conversely, analyzing with TPR HMM resulted in an average score of +70 for “TPR proteins”, +5 for “SNAP,” and –20 for “SwissProt38” [Fig. 2(A)]. For “TPR proteins,”

the number of TPR motif occurrences increased steadily from 1 to 12 with respect to the HMM score. “SNAP” totaled in average 3–4 TPR motifs per sequence, a significantly higher number than “TPR proteins” of similar scores [Fig. 2(B)]. In SNAP sequences, the number of TPR motifs was therefore about one-half that of AdAR motifs.

This observation is linked to the fact that within most TPR motifs detected in a given SNAP, two AdAR motifs were found, separated by a distance of 1, 7, or 8 aa. This contrasts with the situation in TPR proteins, where the number of AdAR motifs found is always less than the number of TPR motifs.

As a consequence, it is possible to distinguish “SNAP sequences” from “TPR-proteins” using the AdAR and the TPR HMMs, at a confidence level better than 95%. First, using the AdAR HMM, a score limit of 10 and a number of motifs strictly higher than 3 is characteristic of most SNAP sequences. For the sequences giving AdAR HMM scores comprised between 1 and 10, the number of motifs per protein sequence again makes the difference, since SNAP always contains more than 4 motifs. However, some bona fide TPR proteins give intermediate scores and a large number of motif occurrences using the AdAR HMM: in the case of the 38th SwissProt release, this corresponds to the two OGT1_RAT and YHR7_YEAST sequences reported above. In this limiting case, clarification will be obtained by comparing the scores obtained by the protein of interest in the AdAR or TPR HMMs, to that of the set of representative proteins examined above. It is apparent from Figures 1 and 2 that SNAP and TPR proteins give poor scores in TPR or AdAR HMMs, respectively. Altogether, this comparison quantitatively supports the idea that SNAP and TPR proteins, albeit sharing some similarities, have distinct structures.

Structural Basis of the Aromatic Di-Alanine Repeats

What can be the rationale for the presence of several AdAR motifs in SNAP? A common observation in biology is that sequence features conserved along species correspond to aa implicated in structure or function of the protein family. The probability matrix at each aa position along the AdAR HMM is deposited at the Pfam database (<http://www.sanger.ac.uk/pfam/>) under the accession number PF02071. As expected from their names, the “di-alanine” positions exhibit high probability for small uncharged aa whereas aromatic, as well as large hydrophobic, aa are more probable at the “aromatic” position. A consensus sequence is therefore SSxxLxxSSxxL where S and L designate small and large uncharged aa. Our structural analysis will therefore focus at these characteristic positions 1, 2, 5, 8, 9, and 12 of the AdAR motif.

In the case of SNAP, the 3D structure of one member of the family, the yeast protein Sec17p,²⁵ has recently been solved. Based on this structure, SNAP comprise three domains: (1) a twisted sheet domain consisting of 8 helices, preceded by (2) a short N-terminal domain, and (3) a globular C-terminal domain.^{8,9} Mapping the AdAR motifs to the aligned SNAP sequences shows that their positions coincide with that of helices of the twisted sheet domain (data not shown). Furthermore, the characteristic aa of the AdAR motif (positions 1, 2, 5, 8, 9, and 12) in Sec17p are not accessible to external water molecules [Fig. 3(A)]. The AdAR motifs may therefore be involved in the hydrophobic packing of the helices together. In order to gain more

insight in the structural role, we therefore examined the geometrical relationship of the four AdAR motifs detected by the AdAR HMM in helices 1, 2, 3, and 6 of the Sec17p twisted sheet domain.

Using Swiss-PdbViewer, we examined each helix of the Sec17p twisted sheet domain and, for each aa characteristic of the AdAR motif (at positions 1, 2, 5, 8, 9, and 12), we determined the closest aa in the structure. The three consecutive AdAR motifs at positions 36–47, 56–67, and 76–87 of helices 1, 2, 3 were first considered. As shown in Figure 3(B) and C, almost every large aa in an AdAR motif was sandwiched between two small aa belonging to two AdAR motifs present on two adjacent helices. This extended the structural significance of the AdAR motif to adjacent helices. Some of the characteristic aa of AdAR motifs 1 and 3 were indeed found to interact closely with a set of aa in the N-terminal helix domain and helix 4 of the twisted sheet domain respectively. The alternate presence of large and small aa capable of hydrophobic interactions at these positions is comparable to the characteristics of the AdAR motif. The fourth AdAR motif in the Sec 17 sequence is quite distant from the three others (aa 137–148, 6th helix) and cannot make any connection with them. However, the same kind of hydrophobic interaction exists with aa present on adjacent helices (5th and 7th). For instance, Tyr148 is very close to Ala117 and Ser147. In this manner, four additional motifs, structurally related to AdAR ones, could be defined in Sec17p sequence: aa 4–11, 96–107, 116–124, and 156–164. The positions of all these motifs coincide with that of AdAR motifs found in other SNAP, where the number of AdAR motifs ranges between 4 to 7 (data not shown). This finding suggests that there are indeed 8 AdAR-like motifs per SNAP and that the AdAR HMM detects most, but not all of them.

Scanning cDNA/EST and Genomic Databases With the AdAR HMM

The strategy exposed above to identify potential SNAP sequences was applied to some currently available cDNA/EST databases. First, scanning *S. cerevisiae*, *D. melanogaster*, *A. thaliana*, and *D. discoideum* databases with AdAR HMM model confirmed that no SNAP but those previously identified are present in these organisms. Since training of the HMMER program had been performed on SNAP sequences present in these organisms, additional searches were conducted on *Caenorhabditis elegans* databases, whose SNAP had not been included in the training. Two candidate sequences were found, corresponding to members of the α -SNAP and γ -SNAP subfamilies previously reported.⁸ Unfortunately, low complexity and repetitive sequences prevent efficiently scanning genomic databases with the HMM profile.

The AdAR HMM was applied on the TrEMBL database. All 21 SNAP were detected. Six TPR proteins were rejected on the rationale explained above. In addition, four protein sequences obtained a score better than 10 and presented more than three motif occurrences (accession numbers: Q9UXX6, O59542, Q9UG01, Q9JKU3). Two of them corresponded to homologous genes in two *Pyrococcus* species (*P.*

horikoshii and *P. abyssi*), and the others to the rat and human Selective LIM homeodomain-binding proteins (SLB).²⁶ A BLAST search was undertaken to obtain close members of these protein families in GenBank and the AdAR HMM was then run on the sequences obtained. Two prokaryote (*P. abyssi* and *Archeoglobus fulgidus*) and two eukaryote (*C. elegans* and *C. briggsae*) new sequences were examined. In both cases, the position of the AdAR motifs was conserved. The secondary structure prediction algorithm PREDICTPROT²⁷ indicated with confidence that the portion of the proteins containing the AdAR motifs were constituted of α -helices. This suggests that the AdAR motif is widespread in both archae and eukaryote phyla.

CONCLUSIONS

The results presented exemplify the power of HMM models to discriminate between related but distinct protein families. Even if a TPR motif could be envisioned as juxtaposed AdAR motifs, the above analysis shows that this is not the case. Comparison between the HMM scores in a very large protein set such as the SwissProt database clearly defined two subpopulations with limited overlap, the SNAP and the TPR families. The success of such analysis is based on the knowledge of a large initial number of orthologous protein sequences. This generated enough variability to "explore" the huge probability space of a 12-aa-long motif (20^{12} possibilities), at both the determining and nondetermining positions, with a rather small number of training patterns (106).

Using the AdAR HMM model, we have been able to specifically detect SNAP in large databases. We have shown that the AdAR motifs are elements of a set of buried aa that make close contacts between adjacent helices of the SNAP. This set extends along the whole twisted sheet structure. The AdAR repeats are therefore suggestive of the presence of junction sites connecting α -helices. At least two to three atoms are involved in such helix-helix side contact, and each helix makes four contacts on average [Fig. 3(C)]. Using a rough value of 2 kT for the energy gained per contact, a stabilization energy of about 10 kT per helix results from the association of adjacent AdAR motifs. Of course, additional interhelical bonds will contribute to the overall stability of the structure. Nevertheless, this set of close contacts is probably critical in determining the SNAP structure, in the sense that not only favorable hydrophobic and van der Waals interactions stabilize the helix packing, but steric repulsions also prevent the helices from interacting more closely with each other.

One may imagine these junction sites as flexible. Even very little sliding of the helices with respect to each other will give rise to very large changes in the relative orientation of the short N-terminal and the globular C-terminal domains. These domains interact directly with the SNARE complex and NSF, respectively.¹⁰ Evidence for the existence of such a conformational change is well known. SNAP indeed bind reversibly to hydrophobic surfaces, and upon binding become able to recruit NSF²⁸ and modulate its ATPase activity.²⁹ This argues in favor of two conformational states of close energy. We speculate that a slight

rotation of the helices around these junction sites may be the molecular basis of SNAP function. Rotation of α -helices one relative to another has already been proposed to explain the opening of the Na⁺ channel of the synaptic acetylcholine receptor.³⁰ Flexibility in SNAP would allow storing the energy gathered by multiple NSF ATPase cycles as strain in the SNAP structure. Release of this strain would be coupled to individual SNARE dissociation from the SNARE complex by a peeling process.

In addition to SNAP, AdAR motifs were found at conserved places in proteins present in different archaeobacteria. The presence of AdAR motifs in SLB is quite interesting, since some of the positions are conserved from *C. elegans* to humans. Furthermore, the AdAR motifs in rat SLB (aa 809–819, 944–955, 1232–1243, 1311–1322, and 1283–1293) surrounds the minimal LIM interacting domain (aa 1213–1265).²⁶ Altogether this suggests the presence of repeated motifs identical or similar to AdAR in non-SNAP, that by analogy may play important structural roles.

Finally, it should be noted that SNAP of the α -subfamily give consistently better scores than proteins of the γ -subfamily (not shown). Interestingly, a glycine residue present in all α -SNAP sequences (G62 in the human sequence) that belongs to the well-conserved second AdAR motif, is replaced by a lysine in all γ -SNAP sequences (K55 in the human sequence). This is indicative of a potentially important structural difference between the two protein subfamilies.

ACKNOWLEDGMENT

During his sabbatical in Grenoble, Dr. Casavant was a recipient of a Commissariat à l'Energie Atomique fellowship.

REFERENCES

1. Rothman JE. Intracellular membrane fusion. *Adv Second Messenger Phosphoprotein Res* 1994;29:81–96.
2. Weber T, Zemelman BV, McNew JA, Westermann B, Gmachl M, Parlati F, Söllner TH, Rothman JE. SNAREpins: Minimal machinery for membrane fusion. *Cell* 1998;92:759–772.
3. McNew JA, Weber T, Parlati F, Johnston RJ, Melia TJ, Söllner TH, Rothman JE. Close is not enough: SNARE-dependent membrane fusion requires an active mechanism that transduces force to membrane anchors. *J Cell Biol* 2000;150:105–117.
4. Wilson DW, Whiteheart SW, Wiedmann M, Brunner M, Rothman JE. A multisubunit particle implicated in membrane fusion. *J Cell Biol* 1992;117:531–538.
5. Whiteheart SW, Griff IC, Brunner M, Clary DO, Mayer T, Buhrow SA, Rothman JE. SNAP family of NSF attachment proteins includes a brain-specific isoform. *Nature* 1993;362:353–355.
6. Stenbeck G. Soluble NSF-attachment proteins. *Int J Biochem Cell Biol* 1998;30:573–577.
7. Rossi G, Salminen A, Rice LM, Brunger AT, Brennwald P. Analysis of a yeast SNARE complex reveals remarkable similarity to the neuronal SNARE complex and a novel function for the C terminus of the SNAP-25 homolog, Sec9. *J Biol Chem* 1997;272:16610–16617.
8. Weidenhaupt M, Bruckert F, Louwagie M, Garin J, Satre M. Functional and molecular identification of novel members of the ubiquitous membrane fusion proteins alpha- and gamma-SNAP (soluble N-ethylmaleimide-sensitive factor-attachment proteins) families in *Dictyostelium discoideum*. *Eur J Biochem* 2000;267:2062–2070.
9. Rice LM, Brunger AT. Crystal structure of the vesicular transport

- protein Sec17: Implications for SNAP function in SNARE complex disassembly. *Mol Cell* 1999;4:85–95.
10. Barnard RJ, Morgan A, Burgoyne RD. Domains of alpha-SNAP required for the stimulation of exocytosis and for *N*-ethylmaleimide-sensitive fusion protein (NSF) binding and activation. *Mol Biol Cell* 1996;7:693–701.
 11. Hohl TM, Parlati F, Wimmer C, Rothman JE, Söllner TH, Engelhardt H. Arrangement of subunits in 20 S particles consisting of NSF, SNAPs, and SNARE complexes. *Mol Cell* 1998;2:539–548.
 12. Barnard RJ, Morgan A, Burgoyne RD. Stimulation of NSF ATPase activity by alpha-SNAP is required for SNARE complex disassembly and exocytosis. *J Cell Biol* 1997;139:875–883.
 13. DeBello WM, O'Connor V, Dresbach T, Whiteheart SW, Wang SS, Schweizer FE, Betz H, Rothman JE, Augustine GJ. SNAP-mediated protein–protein interactions essential for neurotransmitter release. *Nature* 1995;373:626–630.
 14. Yu RC, Jahn R, Brunger AT. NSF N-terminal domain crystal structure: models of NSF function. *Mol Cell* 1999;4:97–107.
 15. Ordway RW, Pallanck L, Ganetzky B. A TPR domain in the SNAP secretory proteins. *Trends Biochem Sci* 1994;19:530–531.
 16. Lamb JR, Tugendreich S, Hieter P. Tetratricopeptide repeat interactions: To TPR or not to TPR? *Trends Biochem Sci* 1995;20:257–259.
 17. Blatch GL, Lassle M. The tetratricopeptide repeat: A structural motif mediating protein–protein interactions. *BioEssays* 1999;21:932–939.
 18. Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998;14:755–763.
 19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
 20. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci USA* 1998;95:5857–5864.
 21. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein families database. *Nucleic Acids Res* 2000;28:263–266.
 22. Guex N, Diemand A, Peitsch MC. Protein modelling for all. *Trends Biochem Sci* 1999;24:364–367.
 23. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 1997;18:2714–2723.
 24. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
 25. Griff IC, Schekman R, Rothman JE, Kaiser CA. The yeast SEC17 gene product is functionally equivalent to mammalian alpha-SNAP protein. *J Biol Chem* 1992;267:12106–12115.
 26. Howard PW, Maurer RA. Identification of a conserved protein that interacts with specific LIM homeodomain transcription factors. *J Biol Chem* 2000;275:13336–13342.
 27. Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 1994;19:55–72.
 28. Clary DO, Griff IC, Rothman JE. SNAPs, a family of NSF attachment proteins involved in intracellular membrane fusion in animals and yeast. *Cell* 1990;61:709–721.
 29. Morgan A, Dimaline R, Burgoyne RD. The ATPase activity of *N*-ethylmaleimide-sensitive fusion protein (NSF) is regulated by soluble NSF attachment proteins. *J Biol Chem* 1994;269:29347–29350.
 30. Unwin N. The nicotinic acetylcholine receptor of the *Torpedo* electric ray. *J Struct Biol* 1998;121:181–190.