doi:10.1016/j.jmb.2007.04.045 J. Mol. Biol. (2007) 370, 170–182

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Amino Acids in the Second Transmembrane Helix of the Lhca4 Subunit Are Important for Formation of Stable Heterodimeric Light-harvesting Complex LHCI-730

Dominik Corbet[1], Thorsten Schweikardt[2] Harald Paulsen[1] and Volkmar H. R. Schmid[1]*

[1]*Institut für Allgemeine Botanik, Johannes Gutenberg-Universität Mainz Müllerweg 6, 55099 Mainz, Germany*

[2]*Institut für Molekulare Biophysik, Johannes Gutenberg-Universität Mainz Jakob Welder Weg 26, 55099 Mainz, Germany*

**Corresponding author*

Photosynthetic light-harvesting complexes (LHCs) are assembled from apoproteins (Lhc proteins) and non-covalently attached pigments. Despite a considerable amino acid sequence identity, these proteins differ in their oligomerization behavior. To identify the amino acid residues determining the heterodimerization of Lhca1 and Lhca4 to form LHCI-730, we mutated the poorly conserved second transmembrane helix of the two subunits. Mutated genes were expressed in *Escherichia coli* and the resultant proteins were refolded *in vitro* and subsequently analyzed by gel electrophoresis. Replacement of the entire second helix in Lhca4 by the one of Lhca3 abolished heterodimerization, whereas it had no effect in Lhca1. Individual replacement of three amino acid clusters in Lhca4 that deviate from the corresponding sequence of Lhca3, demonstrated their contribution to Lhca1-Lhca4 dimerization. Further dissection by mutation of individual amino acid residues in Lhca4 showed the importance of a serine, phenylalanine, and histidine (S88, F95, H99) for LHCI-730 assembly. Alignment of consensus sequences of the Lhc proteins demonstrated that these amino acids are predominantly unique in Lhca4 at the relevant positions. Construction of a homology model based on the high-resolution structure of LHCII and superimposing these models onto the photosystem I structure suggested an orientation of S88, F95, and H99 toward the third transmembrane helix of Lhca1. Since some of the amino acids are too far apart from their nearest neighbors in Lhca1 for a direct interaction, different modes of interaction are discussed. Finally, by quantifying chlorophylls bound to monomeric LHC obtained with the H99 mutant, we identified this amino acid as a further chlorophyll binding site.

Keywords: protein–protein interaction; light-harvesting complex; photosystem I; reconstitution; chlorophyll binding

Introduction

According to crystallographic data available in the year 2004 about 15% of all proteins exist in an oligomeric state, and dimers are most frequent among them,[1] often being the initial state for higher oligomeric associations.[2] Detailed analyses revealed the significance of recurring sequence motifs for protein–protein interactions. Examples for these are the tetratrico peptide repeat sequence representing a degenerate amino acids consensus sequence forming a scaffold for interactions[3] and the frequent GXXXG motif, which promotes interaction by itself or as part of a major interaction sequence between helices of membrane or soluble proteins.[4,5] Due to the availability of an increasing number of high resolution X-ray structures, the identification of amino acids forming the interfaces in oligomeric proteins was possible for a rapidly growing number

Abbreviations used: Chl, chlorophyll; LHC, light-harvesting complex; Lhca and Lhcb, apoproteins of LHCs belonging to photosystem I or II; PG, phosphatidylglycerol; PS, photosystem; TMH, transmembrane helix.

E-mail address of the corresponding author: vschmid@uni-mainz.de

of proteins. In water soluble proteins, approximately 18% of the surfaces are engaged in oligomerization and in addition to M, L, and I, the aromatic amino acids Y, F, and W are the most prevalent amino acids in protein–protein interfaces.[6] In membrane proteins the small amino acids G, A, and S are most frequently found to form contact points and these amino acids are often four or seven amino acids apart from each other,[7] as for example in the aforementioned GxxxG or the Alacoil motif.[8] However, the significance of individual amino acids for complex formation and stabilization can only be determined by mutation analysis.

In the thylakoid membrane of plants various light-harvesting complexes (LHCs) are located at the periphery of both photosystems (PS). LHCs consist of apoprotein(s) (Lhc) with non-covalently attached pigments. Lhc proteins are an interesting class of proteins, since they differ in their oligomerization behavior and their binding site in PS despite a considerable amino acid sequence identity of about 35%.[9] While Lhcb1-3 form the trimeric LHCII, Lhcb4, Lhcb5 and Lhcb6 usually form the monomeric CP29, CP26, and CP24, respectively.[10,11] By contrast, Lhca proteins form dimers. Lhca1+4 compose the LHCI-730, and Lhca2+3 another dimeric LHC, originally called LHCI-680.[12,13] However, because there are no tetratrico peptide repeat sequences in Lhc proteins and only few GXXXG motifs, the above-mentioned sequence motifs for protein–protein interaction are only scarcely present in Lhc proteins. This indicates that oligomerization of Lhc proteins depends on different structure properties.

As a consequence of the high degree in amino acid sequence conservation, it was assumed that all Lhc proteins possess a similar tertiary structure like the monomers of trimeric LHCII. Indeed, the advent of the structure of higher plant PSI confirmed this assumption.[14] Determination of the LHCII structure down to a resolution of 2.7 Å[15] and 2.5 Å[16] makes it the best analyzed member of these proteins. LHCII is composed of three transmembrane helices (TMH), which are connected by loop regions, the extrinsic N and C-terminal domains, and two small amphipathic helices in the lumenal loop and at the C terminus. The first and third TMH form a superhelix, whereas the second TMH is located at some distance. Despite this common construction scheme the Lhc proteins differ in their oligomerization state. Since the termini, the loop regions and the second TMH exhibit the lowest degree of amino acid sequence conservation it can be expected that these regions play a decisive role in achieving different oligomerization. In the case of trimeric LHCII and dimeric LHCI-730 oligomerization happens after formation of the monomeric complexes.[17–21] The LHCII and LHCI-730 are also similar in their requirement of phosphatidylglycerol (PG) for oligomerization.[22–24] Additionally, amino acids at the N and C terminus of Lhcb1[25,26] and Lhca1[21] could be identified, whose replacement resulted in a reduction or loss of oligomers upon gel purification, which reflected their significance for formation and/or stabilization of trimeric LHCII and dimeric LHCI-730. Despite these analogies, interaction of the individual Lhc proteins is very specific, since reconstitution experiments demonstrated that they usually can only form those oligomeric complexes, in which they are normally present *in vivo*.[13,27,28] Furthermore, in contrast to formation of trimeric LHCII, formation of LHCI-730 coincides with ligation of additional chlorophyll (b) molecules.[13]

Later on, most of these results regarding formation/stability of oligomeric LHCs could be confirmed by corresponding localizations in the X-ray structures obtained for LHCII[15] and PSI.[14] The high resolution crystal structure of almost the entire LHCII revealed the presence of the N-terminal domain, the C terminus, the lumenally located section of the third TMH of one monomer, amino acids of the adjacent second TMH in the neighbouring monomer, PG and chlorophylls (Chl) at the interface between the monomers being involved in oligomerization.[15] The PSI structure revealed the presence of an LHCI belt on one side of this complex. The Lhca protein being in close contact with the PSI-G subunit was assigned to be Lhca1.[14] This protein faces with its third TMH and C terminus towards the second TMH of the Lhca4 subunit. From these findings, docking experiments with structure models of Lhca1 and Lhca4[29] and modeling of PSI,[30] it was concluded that the C terminus of Lhca1 interacts with the lumenally oriented section of the second TMH of Lhca4 and the adjacent lumenal loop, which was partly supported by experimental evidence obtained by reconstitutions with mutated Lhca1 protein.[21] Since the 4.4 Å resolution of PSI only allowed detection of the Cα atoms, a combination of homology modeling and molecular dynamics simulations was used for refinement of the structure.[30] The resultant model (Protein Data Bank (PDB) accession code 1YO9), which was based on some assumptions like the presence of one PG molecule at the interface between Lhca1 and Lhca4, suggests some amino acids potentially involved in indirect interaction of the two subunits brought about by (gap) Chls and PG located in the space between the two subunits. However, evidence is lacking as to which amino acids in the second TMH of Lhca4 actually are involved in stabilizing LHCI-730.

In order to elucidate the role of the second TMH and its individual amino acid residues in LHCI-730 formation, we first used chimeric Lhca1 and Lhca4 proteins for dimerization experiments. In these proteins the second TMH was replaced by the corresponding segment of Lhca3, because this protein does not form dimers with either Lhca1 or Lhca4.[27] After confirming the involvement of the second TMH of Lhca4 in dimerization, amino acid groups that deviate in the Lhca3 sequence as compared to the Lhca4 sequence, were mutated in wild-type Lhca4. Amino acids of three clusters that were identified to affect dimerization were subsequently mutated, and some amino acids were found to be important for LHCI-730 formation. To validate