3. Results

3.1 The Physical Properties of the ScPex13p SH3 Domain

3.1.1 ScPex13p Entry in the Gene Database

The genome of the yeast *Saccharomyces cerevisiae* has been completely sequenced, which can be retrieved from several DNA databanks storing the complete genomic DNA sequence of yeast with a free internet access. In the dababases the gene encoding *S.c.* Pex13p is recognized as PEX13 or YLR191w, which is localized on chromosome XII with 1161 bps deoxyribonucleic acids (ORF contains 386 amino acids).

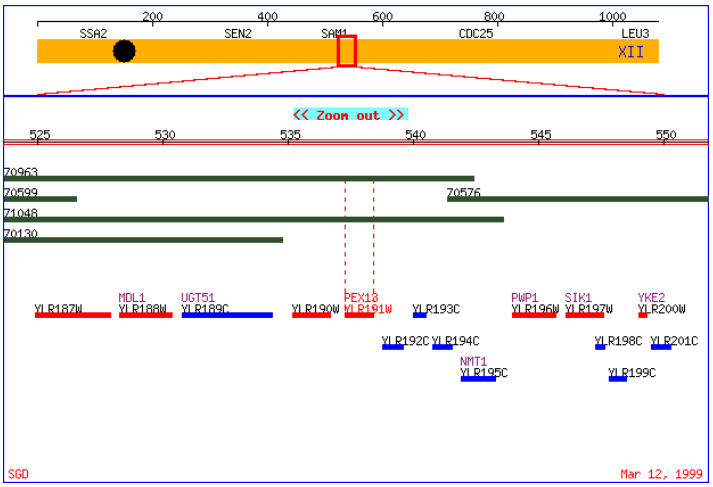


Fig. 5 Gene map of ScPex13p. The gene map of ScPex13p was generated by the search engine with the data in the *S.c.* gene databank maintained by Stanford University.

3.1.2 Determination of the ScPex13p SH3 Protein Sequence

The protein sequence of the ScPex13p SH3 domain (ScPex13pSH3) in this work was determined by the motif search using the program "Pfam HMM" (http://pfam.wustl.edu/hmmsearch.shtml), the protein sequence of ScPex13p used is retrieved from the MIPS databank of protein sequences. This motif search showed that the SH3 domain of ScPex13p protein consists of 61 amino acids comprising amino acid 309 and 370 of ScPex13p (Fig. 6).

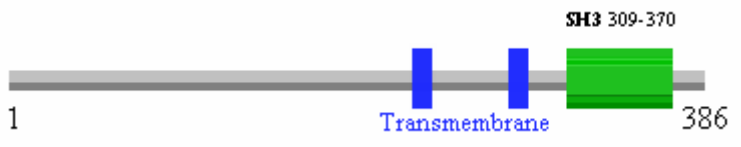


Fig. 6 Schematic illustration of the ScPex13p structure. The locus of SH3 domain in the ScPex13p is illustrated.

This result was also confirmed by the protein sequence alignment of the Scpex13p with those of its homologous proteins by using the alignment program "MultiAlign". As shown in Fig. 7, the region of amino acids from 309 to 370 in ScPex13p protein is highly conserved among these homologous proteins.

Fig. 7 Alignment of various Pex13ps. The sequence alignment of various Pex13p proteins was produced by the program "MultiAlign".

The ScPex13p SH3 protein sequence resulted from the motif search is shown below:

FARAL YDFVP ENPEM EVALK KGDLM AILSK KDPLG RDSDW WKVRT KNGNI GYIPY NYIEI I

3.1.3 Predicted Physical Properties of the ScPex13p SH3 Domain

Since the protein sequence of the ScPex13p SH3 domain is known, it is possible to predict some of its physical properties by using programs in the internet. Of the calculated parameters the following ones are important:

Amino acid number:	61 aa
Molecular weight:	7104.2 Da
Extinction coefficients:	16500 $M^{-1}cm^{-1}$ at 280 nm
Instability index (II):	36.87
pI:	6.37

The calculated instability index (II) (36.87), which comprise less than the criteria 40, classifies the protein into stable groups. This ensures that the ScPex13p SH3 domain could be purified relatively easily and stored at 4 °C for a long time. Practically the purified ScPex13p SH3 domain can be deposited in its storage buffer at 4 °C over several weeks without any obvious degradation detectable by the structure determination by NMR.

3.2 Isolation and Purification of the ScPex13p SH3 Domain by Strategy I

3.2.1 Strategy for the Expression Construct

The 6x His-tag is one of the smallest tags to isolate overexpressed proteins. It does not disturb the structure and the functionality of the heterologously expressed protein at most times. The construct pQE31ScPex13SH3-I coding for the SH3 domain of ScPex13p has been designed. As shown in Fig. 8, the encoded protein also contains a protease site, just at the start of the SH3 protein domain sequence, recognized by Factor Xa, which was introduced by PCR primer RE162. This proteinase site has been introduced so that the coded ScPex13p SH3 domain could be released from the His-tagged fusion protein in the case the short His tag and other amino acids in front of the SH3 domain (see Fig. 8) affects the stereo structure of the SH3 domain.

```
      6x His          Factor Xa
1  MRGSHHHHHH TDPHASSVPS TIEGRFARAL YDFVPENPEM EVALKKGDLM AILSKKDPLG
                                                    SH3 Domain
61 RDSDWWKVRT KNGNIGYIPY NYIEII
      SH3 Domain
```

Fig. 8 Protein sequence of the pQE31ScPex13SH3-I construct. The protein sequence of the pQE31ScPex13SH3 construct is illustrated.

Supposing that the ORF (86 aa) in the construct pQE31ScPex13SH3-I is too short to express protein heterologously in *E. coli* because of degradation of foreign proteins in

the bacteria cells, an alternative construct was also designed at same time, pQE41ScPex13SH3-I, in which ScPex13 SH3 domain was fused to the C-terminus of the DHFR protein (Dehydrofolate reductase) with 6x His at the N-terminus. There is also a factor Xa site, introduced by the PCR primer, between DHFR and SH3 so that the expressed fusion protein can be cleaved and the SH3 domain can be extracted from the overexpressed fusion protein.

```
  1  MRGSHHHHHH ..................DHFR.......................RMTTTSSVEG
        6x His
181  LSEVQEEKGI KYKFEVYEKK GSDLHAVPST IEGRFARALY DFVPENPEME VALKKGDLMA
                                     Factor Xa        SH3 Domain
241  ILSKKDPLGR DSDWWKVRTK NGNIGYIPYN YIEII
       SH3 Domain
```

Fig. 9 Protein sequence of the pQE41ScPex13SH3-I construct. The protein sequence of the pQE41ScPex13SH3 construct is illustrated.

The cloning of the SH3 domain into the vector pQE31 or pQE41 was successful; however, both of these constructs pQE31ScPex13SH3 and pQE41ScPex13SH3 could not be expressed in bacteria cells (*E.coli*) of various strains under different expression conditions.

3.3 Isolation and Purification of the ScPex13p SH3 Domain by Strategy II

3.3.1 Strategy for the Expression Construct

The construct pGEX-4T2-ScPex13SH3-I contained the SH3 domain (309 – 370 aa) of ScPex13 fused to the C-terminus of GST protein with a proteinase site of Factor Xa between them. This fusion protein was named as GST-SH3, the protein sequence of which is illustrated in Fig. 10.

```
  1  MSPILGYWKI ................... GST ..................... DAFPKLVCFK
181  KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LVPRGSPGIP GSTIEGRFAR
                                                            Factor Xa
241  ALYDFVPENP EMEVALKKGD LMAILSKKDP LGRDSDWWKV RTKNGNIGYI PYNYIEII
                              SH3 domain
```

Fig. 10 Protein sequence of GST-SH3 with Factor Xa protease site. The protein sequence of the GST-SH3 domain construct is shown.

Physical properties of the GST-SH3:

Amino acid number:	298 aa
Molecular weight:	34461.9 Da
Extinction coefficients:	57180 $M^{-1}cm^{-1}$ at 280 nm
Instability index (II):	37.82
pI:	6.13

The fusion protein overexpressed in *E. coli* could be isolated from the cell lysate by the GST-binding Matrix. The purified fusion protein was digested with proteinase Factor Xa on the matrix or in solution to free the ScPex13p SH3 domain, which was purified further by removing protein GST and the proteinase factor Xa. (see Fig. 11)

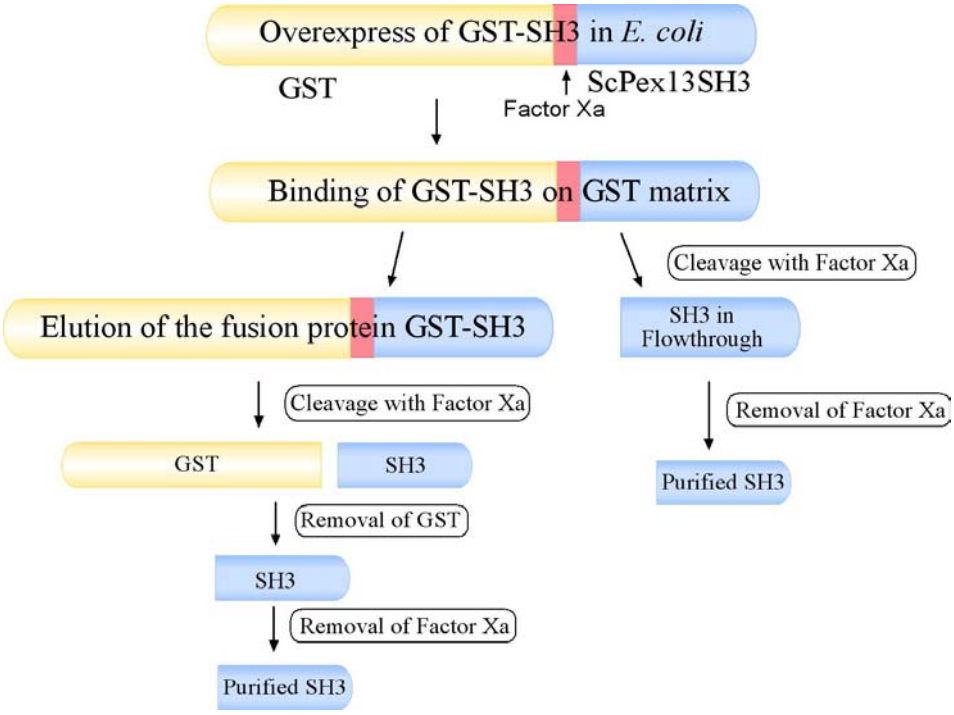


Fig. 11 Schematic illustration of the isolation pathway of SH3

3.3.3 Optimization of Expression and Purification

Optimization of the overexpression of GST-SH3 in LB medium was performed by variation of the following factors:

- Cell growth phase for induction: the induction was carried out at the OD_{600}: 0.3, 0.4 0.5 and 0.6 of the cell culture.
- 0.05 to 0.1 mM IPTG end concentrations was used to induce the bacteria cells to overexpress the fusion protein GST-SH3.

- The induction of the bacteria cells to express the fusion protein was carried out at 25 °C (room temperature), 30 °C and 37 °C.

All these variations of the expressing conditions made little difference in the heterologous expression of the fusion protein GST-SH3 in bacteria cells. The bacteria strains (BL21(DE3), TG1) tested in this work showed almost the same expression level. In contrast, the different conditions for cleavage of the fusion protein GST-SH3 with the protease Factor Xa resulted in very different yields of purified SH3 (see below). Therefore the optimization of the purification of SH3 protein was focused on the cleavage of the fusion protein with the Factor Xa and the following purification steps.

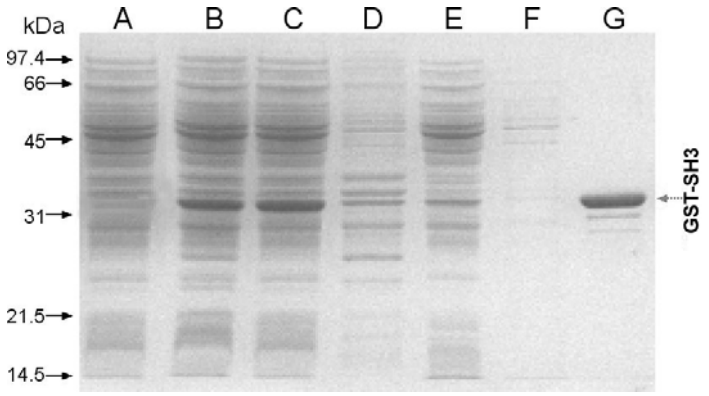


Fig. 12 Pilot isolation of GST-SH3 (Factor Xa). BL21(DE3) bacteria cells transformed with the plasmid pGEX-4T2-ScPex13-SH3-I was induced by 0.1 mM IPTG, the heterologously expressed fusion protein was purified by a GST column. A: not induced; B: total lysate of the induced cells; C: supernatant of the lysate after 10 min. centrifugation at 10 000 g; D: pellet of the lysate after centrifugation; E: flow-through of the GST column; F: Wash; G: eluate, purified GST-SH3.

Fig. 12 shows a pilot isolation of GST-SH3 out of 200 ml bacteria cell culture, which was induced with 0.1 mM IPTG end concentration at 37 °C with the OD_{600} of 0.6 for 4 hours.

3.3.3.1 Cleavage of the GST Fusion Protein in Solution

As illustrated in Fig. 11, this strategy of cleavage of the fusion protein consists of the following steps:

- Elution of GST fused protein GST-SH3 in the elution buffer;
- Dialyses of the eluate containing GST-SH3 fusion protein with the cleavage buffer, in which Factor Xa can work effectively;

- Cleavage of the fusion protein GST-SH3 in solution by Factor Xa;
- Removal of the GST protein (also some rest of uncleaved fusion protein) and Factor Xa individually, or removal of these proteins in one step.

Glutathione had to be removed as it can deactivate Factor Xa by disrupting its disulphide bridges. Glutathione in the eluate containing GSH-SH3 was easily and rapidly removed from the protein solution by centrifugation through a Centricon membrane filter with a 10 kDa cut-off. This filtering procedure also maintained an appropriate concentration (0.1 ~ 20 mg/ml) of the protein and pH value of the solution. In this work it was found that 17 mg of a target protein GST-SH3 with a molecular weight of ca 35 kDa could be cleaved with the highest efficiency at pH 8.0 in a total volume of 200 ml or less.

3.3.3.2 Isolation of SH3 from the Cleaved Mixture

Various methods exploiting the difference in molecular weight or fusion tag were tested to isolate ScPex13p SH3 from the cleaved mixture.

3.3.3.2.1 Gel filtration Chromatography Method

The Gel filtration chromatography was tested with an aliquot of 1 ml of the digested mixture. It turned out that this method is not suitable for the isolation of the SH3 domain because the very small SH3 came out as the last fraction and was diluted in the column buffer almost 100 times, which is difficult to concentrate again without introducing buffers other than PBS for later NMR studies. A suitable method is to centrifuge the diluted sample through a filter memberane such as Centricon, but the loss of the target protein through the concentrating procedure is noticeable high.

3.3.3.2.2 Isolation of SH3 out of the Digested Mixture by Centricon Membrane

Based on that the SH3 domain to be isolated has a much smaller molecular weight than all the other contaminants in the mixture, the Centricon filter membrane with a cut-off of 30 kDa was used for seperation.

The SH3 domain was in the filtrate while all the other proteins stayed in the concentrate above the membrane filter. The target protein SH3 in the filtrate from the first filtration was concentrated by filtration through the Centricon-3 membrane filter

with a cut-off of 3 kDa. Any small molecules with molecular weights less than 3 kDa were centrifuged through the filter while the target protein SH3 domain stayed above the membrane in the concentrate. During this filtration the SH3 was also transferred into its storage buffer which is the best buffer condition to take NMR spectrum.

3.3.3.2.3 Improved Filter Membrane Method

The Centricon membranes were successfully deployed to the isolation of the SH3 domain. However, the amount of SH3 domain protein is just a very small portion of the protein mixture digested by factor Xa, therefore the filter membrane was very susceptible to be blocked with the huge amount of these big proteins consisting of GST and the fusion protein GST-SH3 not cleaved by the protease. In the improved filter membrane method GST-SH3 or GST were remove by affinity chromatography with the GST Sepharose after the reduced glutathione was removed from the mixture. By this method more than 95% GST and GST-SH3 could be removed from the mixture before the advanced isolation using Centricon membrane filters was carried out.

However the total efficiency of the complete isolation procedure was still as low as 20%, even with the improved filter membrane method. A trial experiment of isolation as shown in the following was analyzed in order to locate the unsuccessful step in the isolation procedure for further improvement.

E. coli BL21(DE3) cells transfected with the plasmid pGEX-4T2-ScPex13-SH3-I coding for the ScPex13 SH3 were cultured in 3 liter M9 medium. Ca. 50 mg GST fusion protein GST-SH3 could be eluted with reduced glutathione from the GST Sepharose column loaded with the cleared lysate of the bacterial. The GST fusion protein in the eluate was then transferred into the buffer for digestion by using Centricon-10 (with cut-off of 10 kDa) while the end concentration of the fusion protein in the buffer for digestion with protease factor Xa was maintained at about 1.5 mg/ml in order to give the best efficiency of digestion. This step of changing buffer and concentration had an efficiency of 91.3%. After the fusion protein (45 mg) was digested by factor Xa. Removal of non-digested GST-SH3 protein and GST protein reached an efficiency of 87.1%. The flow-through of the GST-column was then filtered with the Centricon Y30. The protease factor Xa and the other big proteins,

like GST and GST-SH3, stayed in the concentrate on the membrane filter, while the SH3 domain protein was found in the filtrate. Unfortunately the efficiency of this step of purification was as low as 32.4%. At last the filtrate containing the purified SH3 domain was concentrated and transferred in the storage buffer by centrifuged through the filter Centricon Y3 with a cut-off of 3 kDa, with an efficiency of 75%. Aliquots of the sample from every step of the whole purification procedure from the lysis of the bacterial cells having overexpressed the fusion protein GST-SH3 to the purified SH3 domain protein were analyzed by SDS page (Fig. 13).

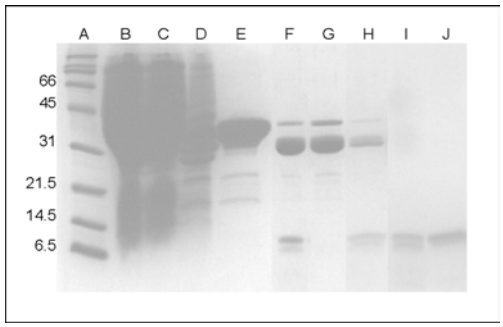


Fig. 13 Illustration of the complete procedure of isolation of the SH3 domain, exploiting Factor Xa and Centricon. A: low range Biorad protein marker; B: total lysate of the bacterial cells; C: supernatant of the cleared lysate by centrifugation at 10 000 g; D: pellet of lysate centrifuged at 10 000 g; E: eluate of the GST column; F: eluate digested with factor Xa; G: proteins captured out of the digested mixture by a second GST column; H: flow-through of the second GST column; I: filtrate of the flow-through of a Centron-30 membrane; J: concentrate of the filtrate from the last step by Centron-3 membrane.

3.3.3.3 Cleave GST Fusion Protein on the GST Sepharose

Alternative to the digestion of the GST fusion protein in solution, it was also investigated to carry out the digestion directly on the GST Sepharose in the column. The advantages of this method are obvious: the GST-removing step after the digestion by protease would be not necessary any more; it would be also easier to suspend the GST Sepharose in the cleavage buffer for factor Xa than to transfer the GST fusion protein in the eluate solution from one buffer to another. With the reduced steps of the purification procedure, it was expected that the over all efficiency of the purification procedure would be increased.

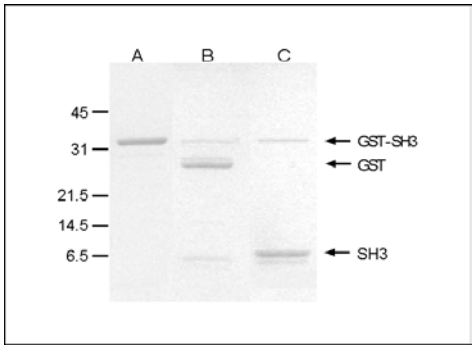


Fig. 14 Cleavage of GST-SH3 on the GST matrix with factor Xa. A: GST-SH3 fusion protein; B: proteins remaining on the GST resin after the GST-SH3 fusion protein has been digested with protease factor Xa; C: SH3 protein released into the flow-through solution by cleavage with protease factor Xa. A little GST-SH3 fusion protein was also translocated into solution.

As already tested in the solution, factor Xa can digest the GST fusion protein at room temperature; this temperature was also chosen for the digestion of the GST fusion protein on the GST Sepharose. After a series of pilot experiments to optimize the digestion condition of GST fusion protein on the resin it was found that the optimal amount of factor Xa should be at least 180 Unit for 0.5 ml solid volume of the resin and that the digestion time should be 16 hours (not more than 24 hours at room temperature). Under this optimized condition 6.5 mg GST fusion SH3 domain bound on the GST-column was digested by the factor Xa with an efficiency of 72%. However, more than 600 units of factor Xa were needed to produce 3.5 mg SH3 protein (1mM, 0.5 ml solution) needed for NMR structure research.

3.3.3.3.1 Removal of Factor Xa with Xarrest Kit

Theoretically, after incubation with factor Xa there are just the cleaved SH3 domain and factor Xa in the flow-through of the GST Sepharose resin. Besides the methods described above, the Xarrest kit provided by Novagen can also be used to quantitatively remove the factor Xa.

Following cleavage of the GST fusion SH3 protein, the factor Xa is bound batch-wise to Xarrest Agarose and the target protein is recovered by spin filtration. However, it must be pointed out that 50 µl solid volume of the Xarrest agarose was needed to remove just 4 Unit factor Xa, thus, as in the example above, for 6.5 mg GST fusion SH3 domain (binding on 0.5 ml GST Sepharose) 9 ml Xarrest resin would be needed to capture the protease after digestion. However this could just produce 0.2 mg of

purified SH3 protein. In order to get the required 3.5 mg of the purified SH3 (1 mM 0.5 ml solution) 160 ml Xarrest resin would be needed.

3.4 Isolation and Purification of ScPex13p SH3 by Strategy III

3.4.1 Strategy for the Expression Construct

So far none of the various methods investigated to isolate the SH3 domain did yield a satisfying efficiency to fulfill our need. The His tag fused protein is relative easy to isolate while the short His tag sticking at one end of the target protein has usually no influence on the structure folding and function of the targeting protein; but both of such a expression constructs coding the SH3 domain did not work. The GST fusion protein was easily heterologously overexpressed in *E. coli* and it was easily purified with an affinity column of GST Sepharose. However, the GST tag could not be removed with high efficiency to purify the target protein SH3 domain.

A fusion protein was constructed which should exploit the advantages of both of His- and GST-tag fusion proteins. The fusion protein named GST-6xHis-SH3 was constructed based on the pGEX-4T2 vector, in which the GST tag could be remove from the purified fusion protein by the protease thrombin, and the remaining His-SH3 domain could be further purified by the Ni-NTA affinity chromatography. In such a strategy it was assumed that the short peptide GSHHHHHH should not affect the folding and the functional properties of the SH3 domain, which was verified later in the NMR spectrum of the purified SH3 domain.

Another advantage of this strategy is the high cleavage efficiency of the protease thrombin with a very low mass ratio of enzyme to target proteins.

The protein sequence of the fusion protein GST-6xHis-SH3 is shown in Fig. 15

```
  1  MSPILGYWKI .......... .......... ....GST... .......... DAFPKLVCFK
181  KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LVPRGS HHHHHH FARALYDF
                                                 Thrombin  His
                                                                    SH3
241  VPENPEMEVA LKKGDLMAIL SKKDPLGRDS DWWKVRTKNG NIGYIPYNYI EII
                           SH3
```

Fig. 15 Protein sequence of GST-His-SH3 construct. The protein sequence of the GST-His-SH3 construct is shown.

Physical properties:

Amino acid number:	293 aa
Molecular weight:	34219.5 Da
Extinction coefficients:	57420 $M^{-1}cm^{-1}$ at 280 nm
Instability index (II):	35.33
pI:	6.49

Physical Properties of the SH3 cleaved from GST-6xHis-SH3 by thrombin:

Amino acid number:	69 aa
Molecular weight:	8701.1 Da
Extinction coefficients:	16500 $M^{-1}cm^{-1}$ at 280 nm
Instability index (II):	32.25
pI:	7.20

3.4.3 Overexpression and Purification of the SH3 Domain

3.4.3.1 Optimization of the Expression in *E. coli*

As the new construct is very similar to pGEX-4T2-ScPex13pSH3-I coding the GST fusion protein GST-SH3, the optimization of overexpression of the fusion protein GST-6xHis-SH3 started with the optimized condition for the construct pGEX-4T2-ScPex13pSH3-I coding the fusion protein GST-SH3.

The optimization of overexpression of the fusion protein GST-6xHis-SH3 was carried out in LB medium. The *E. coli* BL21(DE3) bacterial cells transformed with the plasmid DNA pGEX-4T2-ScPex13pSH3-II grew at 37 °C till the OD_{600} of the culture reached 0.6 and then the cells were induced with 1 mM IPTG for 4 hours. Under this condition ca. 2 g cells out of 1 liter cell culture could produce ca. 9 – 12 mg fusion protein GST-6xHis-SH3. As in the case of the fusion protein GST-SH3, in which the expression level of the fusion protein GST-SH3 was similar in various conditions, the condition tested here for the fusion protein GST-6xHis-SH3 was not further optimized.

3.4.3.2 Optimization of the Purification of the SH3

The purification schema is illustrated in Fig. 16 (see below). The fusion protein GST-6xHis-SH3 overexpressed in bacteria *E. coli* cells was extracted from bacteria cell lysates by the GST affinity chromatography. The isolated fusion protein GST-6xHis-SH3 was cleaved either on the GST Sepharose resin or in solution by the protease thrombin. As expected it seemed be better to carry out the cleavage directly while the

fusion protein still bound on the GST Sepharose, providing this could be done with a high efficiency. In such way, GST or GST-fusion protein remained bound to the resin, while the SH3 domain protein was released into the solution, which could be easily separated from the GST agarose resin. Alternatively, the cleavage of the fusion protein could be carried out in solution so that an high efficiency of cleavage could be reached; In this case, the target protein SH3 domain fused with 6x His at its N-terminus could also be isolated from the cleaved mixture by a Ni-NTA affinity chromatography.

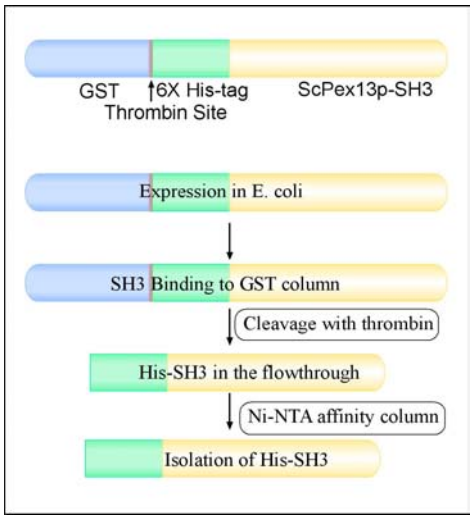


Fig. 16 Illustration of the pathway of the isolation of His-SH3. In the illustration the comprehensive procedure of purification is shown.

3.4.3.2.1 Optimization of Thrombin Digestion

The preferred simple method – cleaving the fusion protein on GST Sepharose by thrombin –worked well for our goal to purify SH3.

The purified fusion protein GST-6xHis-SH3 isolated by affinity chromatography was left on the GST Sepharose resin, and aliquots were suspended in 10 mM reduced glutathione solution to determine the protein concentration on the Sepharose resin.

8 aliquots of 500 µl out of the completely suspended GST matrix slurry were taken for the Thrombin cleavage experiment. These samples are labeled as A, B, C, D, E, F, and G respectively. Each sample contains 200 µg (500 X 0.4) fusion protein GST-6xHis-SH3. Amounts of thrombin used to digest are shown in the table, the ratio of thrombin to the fusion protein range from 1:200 to 1:3400.

As the digestion of the fusion protein GST-6xHis-SH3 by Thrombin at room temperature produced many byproducts shown on the SDS page, the optimized cleavage was carried out at 4 °C, 50 µl slurry aliquots of each sample were taken for analysis after incubation of 6, 12 and 18 hours, which were labeled as A1, A2, A3; B1, … respectively. All these aliquots for assay were analyzed on 15% SDS PAGE. The SDS gels were stained with Coomassie Brilliant Blue, as shown in Fig. 17

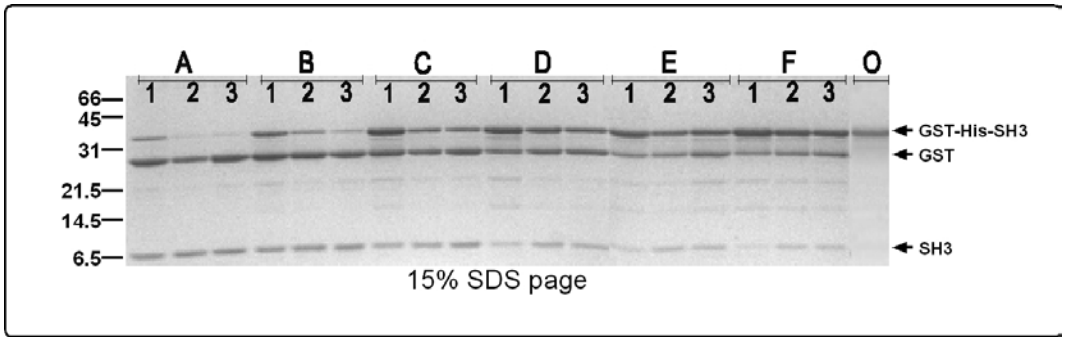


Fig. 17 Optimization of cleavage of GST-His-SH3 by thrombin. The GST-SH3 domain protein was digested with thrombin in different concentration for different time.

Sample	A	B	C	D	E	F	O
Thrombin Volume	10	5	2.5	1.25	0.8	0.6	0
Thrombin : Protein	1:200	1:400	1:800	1:1600	1:2500	1:3400	

Unit of volume: µl, Thrombin stock solution concentration: 0.1 U/µl.

The fusion protein showed no degradation without thrombin under the digestion condition; with a ration of thrombin to the fusion protein of more than 1:2500 the digestion would produce some byproducts, which could be avoided when more thrombin (with a ration of fusion protein to thrombin less than 2500) was used; It was also shown that the fusion protein GST-6xHis-SH3 could be completely cleaved with a ration of thrombin to the fusion protein of 1:200 at more than 12 hours. This optimized digestion condition was chosen to cleave the GST-6xHis-SH3 fusion protein in further experiments of this work.

3.4.3.2.2 Optimization of Ni-NTA Chromatography

As the digestion with thrombin was carried out directly on the GST Sepharose resin, most of the SH3 was in the flow-through of the GST affinity column, contaminated

with other proteins including thrombin protease. The next and also very pivotal step, the purification of the SH3 domain out of the digestion became a trivial thing, due to the His-tag at the N-terminus.

It was investigated and confirmed by a pilot isolation that the introduced 6x His tag at the N-terminus of the SH3 domain exposes on the surface of the SH3 tertiary structure. This allowed the isolation of the His-tagged SH3 by affinity-chromatography of the Ni-NTA matrix under native condition. The fact that the thrombin digestion was carried out directly on the Sepharose resin did significantly eliminate contaminating proteins. In addition to a high ionic strength (300 mM NaCl), low concentration of imidazole in wash buffers were used to further minimize nonspecific interactions between proteins and the Ni-NTA resin.

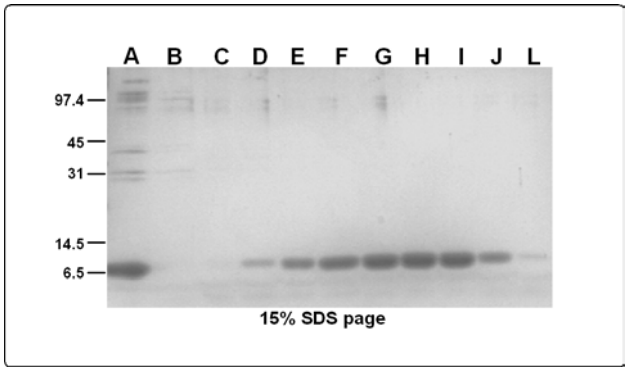


Fig. 18 Optimization of the isolation of the His-SH3 domain. A: the digested mixture loaded onto the His-NTA column; B: flow-through; C: wash with 20 mM imidazole; D: wash with 40 mM imidazole; E: wash with 60 mM imidazole; F: eluate with 80 mM imidazole; G: eluate with 100 mM imidazole; H; eluate with 120 mM imidazole; I: eluate with 140 mM imidazole; J: eluate with 160 mM imidazole; L: eluate with 180 mM imidazole.

The optimization experiments showed that very pure SH3 could be eluted from the Ni-NTA resin by an elution buffer containing imidazole with a concentration from 100 mM to 140 mM. In the following, the elution buffer with 140 mM imidazole was used to elute the target protein SH3 of the His-NTA column.

Subsequently, the purified SH3 domain protein was also successfully transferred to the storage buffer for NMR.

3.5 Large Scale Purification of the SH3 Domain

Three strategies were checked out in this work to find out a suitable method to isolate SH3 domain for illustration of its tertiary structure in solution by NRM technique. After a series of trial and error experiments it has been found out that the third strategy worked at best, in which a His-tagged ScPex13p SH3 domain was fused into the C-terminus of the GST protein.

It could be concluded from the optimization experiments that the best protocol to purify the SH3 domain for NMR structure study was over-expressing the fusion protein GST-6xHis-SH3, cutting the His-SH3 fusion protein of GST-6xHis-SH3 by thrombin and then purifying His-SH3 by Ni-NTA affinity chromatography.

This complete protocol of purifying SH3 domain had a total efficiency of ca. 72%.

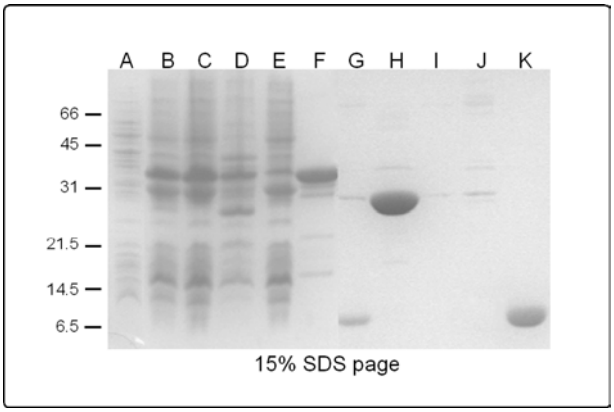


Fig. 19 Illustration of the complete procedure of isolation of ScPex13p SH3 domain. The complete procedure to purify SH3 domain for NMR structure study is illustrated. A: not induced cells (transfected with the plasmid DNA); B: induced cells transformed with the plasmid DNA pGEX-4T2-SH3-II; C: cell lysate cleared by centrifugation; D: pellet fraction of the clearing centrifugation; E: flow-through of the GST Sepharose column; F: GST-His-SH3 fusion protein bound on the GST Sepharose resin; G: His-SH3 released from the GST-column by thrombin cleavage. H: proteins still bound to the GST-column after digestion with thrombin; I: flow-through of the Ni-NTA affinity column; J: wash of the Ni-NTA column; K: eluate of Ni-NTA column containing the purified SH3 protein, which was concentrated by Centricon filter membrane.

3.5.1 Purity of the Isolated SH3

The ScPex13p SH3 isolated in this way was pure enough for NMR structure studies. The purity of the isolated ScPex13 SH3 was analysed in two ways: first the

electrophoretic analysis on 15% SDS PAGE showed only one band corresponding to the pure protein (Fig. 19, lane K); secondly, the MS spectrum which was taken from the ^{15}N labeled purified protein revealed only one major peak.

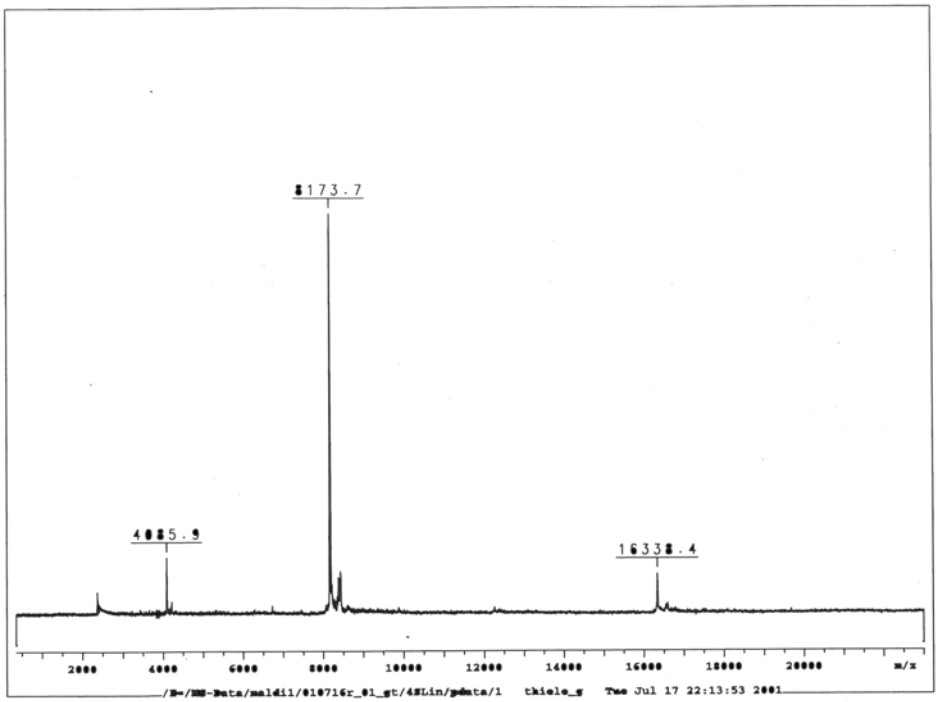


Fig. 20 MS spectrum of the purified His-SH3.

3.5.2 Checking Tertiary Structure of Purified SH3

An NMR spectrum (^{1}H) was taken for the isolated SH3 and all NMR signals of ^{1}H in molecular ScPex13SH3 indicate that the SH3 domain was folded correctly. The result indicated that the protocol developed here can be deployed to isolate ScPex13SH3 for NMR structure study.

3.5.3 Purification of ^{15}N labeled SH3

The bacteria BL21(DE3) cells transformed with the expression vector plasmid pGEX-4T2-ScPex13pSH3-II grew in M9 minimal medium containing ^{15}N marked amino chloride as unique nitrogen source. 3.3 g bacteria cells out of 2 liter culture medium (the output was a little less in M9 medium than in LB medium) which expressed the fusion protein GST-6xHis-SH3 were processed according to the optimized protocol described above. This isolation procedure resulted in 4.1 mg of the target protein ScPex13SH3[^{15}N] with a recovery efficiency of ca. 67.5% out of ca. 25 mg GST fusion protein GST-6xHis-SH3 labeled with ^{15}N. The purified target protein SH3 was concentrated in 0.6 ml solution with concentration of 0.8 mM.

3.5.4 Purification of ^{15}N and ^{13}C Double labeled SH3

The bacteria BL21(DE3) transformed with the expression vector plasmid pGEX-4T2-ScPex13pSH3-II cells were grown in M9 minimal medium containing ^{15}N marked amino chloride as unique nitrogen source and ^{13}C marked glucose as unique carbon source, 2.8 g bacteria cells out of 2 liter culture medium which expressed the fusion protein GST-6xHis-SH3 were processed as described above. This isolation procedure resulted in 3.6 mg of the target protein ScPex13SH3[$^{15}N^{13}C$] with a recovery efficiency of ca. 77% out of ca. 19 mg GST fusion protein GST-6xHis-SH3 double labeled with ^{15}N and ^{13}C. The purified target protein SH3 was concentrated to a final concentration of 0.7 mM in 0.6 ml solution.

3.6 Computer Modelling of the Structure of the SH3 Domain

Before we could illustrate the stereo structure of the ScPex13 SH3 by the NMR spectrum, a computer model of the ScPex13p SH3 domain was calculated by alignment with its homologous proteins with known stereo structures.

3.6.1 Automatic Modelling by the Swiss Modeling Server

An automatic modelling of the tertiary structure of the ScPex13p SH3 by the Swiss protein modeling server was carried out. Three other SH3 domains, namely 1bk2, 1aey, 1SHG, were chosen automatically by the modelling program as templates based on the alignment with the ScPex13 SH3 domain with the highest sequence identity (more than 39%) (see Fig. 21).

The n-Src loop of the ScPex13 SH3 is longer than in the other templates to such an extent that this part of the ScPex13p SH3 could not be modelled automatically.

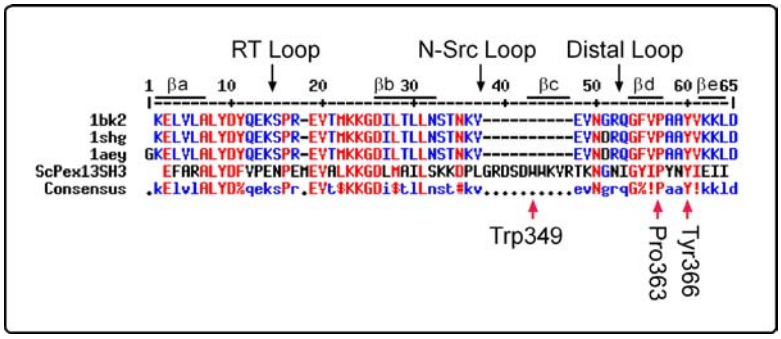


Fig. 21 Alignment of ScPex13p SH3 with other SH3 domains with known structure.

```
PDB ENTRY 1BK2.pdb
    SH3-DOMAIN                                14-JUL-98
    A-SPECTRIN SH3 DOMAIN D48G MUTANT
    X-RAY DIFFRACTION
    RESOLUTION. 2.0  ANGSTROMS.
PDB ENTRY 1SHG.pdb
    CYTOSKELETON                              19-MAY-93
    RESOLUTION. 1.8  ANGSTROMS.
PDB ENTRY 1AEY.pdb
    CYTOSKELETON                              02-MAR-97
    ALPHA-SPECTRIN SRC HOMOLOGY 3 DOMAIN, SOLUTION NMR,
    15 STRUCTURES
```

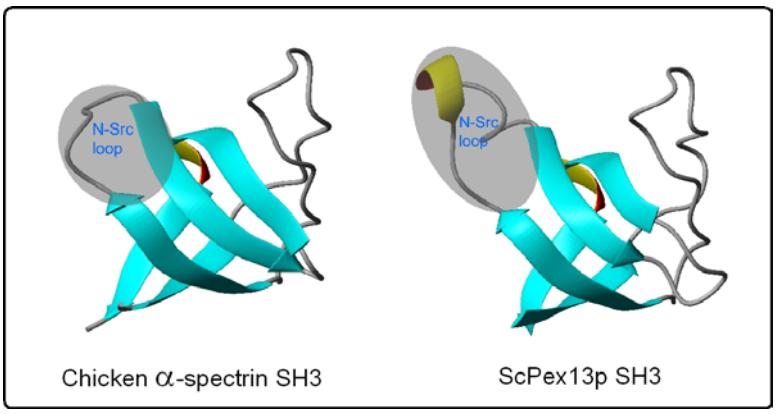


Fig. 22 3D structure of the template (chicken α-spectrin SH3) and the ScPex13p SH3 domain. Left: 3D structure of the modeling template chicken α-spectrin SH3; Right: 3D structure of ScPex13p SH3 automatically modeled by the Swiss Model program. The primary difference between these two SH3 domains lies within the n-Src loops indicated by the gray shadow.

The modeled stereo structure of ScPex13p SH3 domain and the X-ray structure of 1bk2 SH3 domain (chicken alpha-spectrin Src homology 3 domain, X-ray diffraction structure), on which the model of the ScPex13p SH3 domain was based, are shown in Fig. 22 (see below). The n-Src loops in these two structures are definitely different. The n-Src loop in ScPex13p SH3 (especially the region from glycine 344 to arginine 353) could not be properly modelled due to the lack of a known stereo structure base.

This did lead to too many conformations of this part to be optimized by energy minimization.

3.6.2 Manually Modelled Stereo Structure of ScPex13p SH3

3.6.2.1 Modelling ScPex13p SH3 without ScPex14p Peptide Ligand

It was thought that the stereo structures must be the same if the primal sequences of protein are the same or similar. The manual modelling was also based on the alignment of the ScPex13p SH3 and its homologous proteins; however the extreme long n-Src loop in ScPex13p SH3, which exists in just a few SH3 domains but including all Pex13p SH3 domains from different species, was modelled separately.

The protein sequences of the ScPex13p SH3, SPCN_Chick SH3 (1AEY, chicken alpha-spectrin Src homology 3 domain), same protein as 1bk2 but without mutation of D48G and mouse Lck tyrosine kinase SH3 were manually aligned so that the stereo structure of the ScPex13p SH3 domain could be built on the stereo structure of the chicken spectrin SH3 domain.

Base on the alignment with the ScPex13p SH3 domain (Fig. 23), the backbone of the ScPex13p SH3 structure was adopted from the chicken spectrin SH3 domain, the structure of which has been illustrated by X-ray experiments. The residues Glu323 and Asn356 of the ScPex13p SH3 domain, whose counterparts in the chicken spectrin SH3 do not exist, were manually inserted into the adopted stereo structure (see Fig. 24, below).

After the major part of the backbone of the ScPex13p SH3 domain was adopted from the chicken spectrin SH3 domain, the longer extra n-Src loop of ScPex13p SH3 domain was simulated individually by an alignment search in the PDB protein structure database (by PDB web site).

After joining all the blocks of the stereo structure of the ScPex13p SH3 domain, the final structures were reached by minimization of the energy of the complete molecule. There are two ways to insert the extra longer n-Src loop of the ScPex13 SH3 into the

other part of the stereo structure, as shown in Fig. 25. Both structures were optimized by minimization of the molecular energy.

```
SPCN_ChickSH3    TGKELVLALYDYQE.KSPREVTMKKGDILTLLNST.....NKDWWKVEVN..DRQG
SCPEX13_SH3       KLEFARALYDFVPENPEMEVALKKGDLMAILSKKDPLGRDSDWWKVRTK.NGNIG
Lyn_MOUSE_SH3    LQGDIVVALYPYDG.IHPDDLSFKKGEKMKVLE......EHGEWWKAKSLSSKREG

SPCN_ChickSH3    FVPAAYVKKLDP
SCPEX13_SH3      YIPYNYIEIIK
Lyn_MOUSE_SH3    FIPSNYVAKVNT
```

Fig. 23 Alignment of ScPex13p SH3 with other SH3 domains as template for 3D structure modeling.

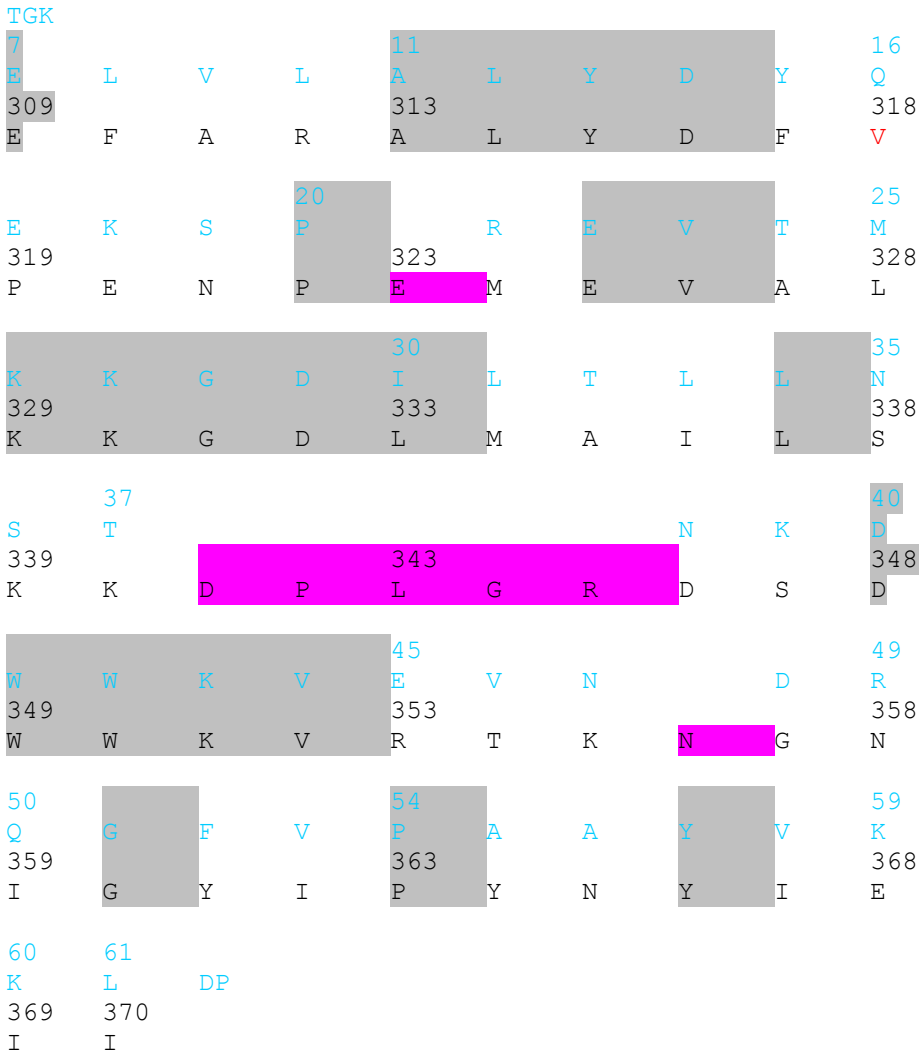


Fig. 24 Fitting of the protein sequence of the ScPex13p SH3 domain to the chicken spectrin SH3 domain. The amino acids in the gray shadows are conserved in structure between the two SH3 domains, this part of the structure SH3 domain was directly adopted from that of the chicken spectrin, while all the other amino acids of ScPex13p SH3 not marked with color have been manually sorted into the structure of

the chicken spectrin SH3; the amino acids of ScPex13p SH3 marked in red color have been fitted manually.

It should be pointed out that the conformation of the binding site in the chicken spectrin SH3 domain for proline rich peptide, namely Trp41, Pro55 and Tyr58, was conserved in ScPex13p SH3 domain, the equivalent amino acids are Trp349, Pro363 and Tyr366.

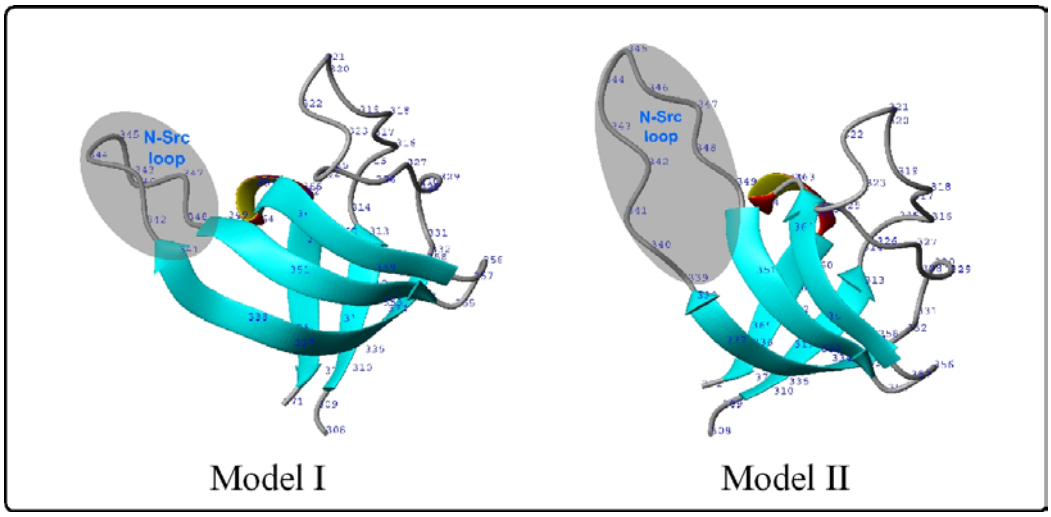


Fig. 25 Modeled ScPex13p SH3 structures. Two alternatives of ScPex13p SH3 domain were generated by manual modelling; the difference between them lies within the fold of the extra long n-Src loops.

3.6.2.2 Modelling of the ScPex13p SH3 Domain with ScPex14p Peptide Ligand

The SH3 domain of Pex13p binds a type II praline rich peptide, which appears in ScPex14p protein. It's well known that the type II binding ligand of SH3 domains adopts an alpha-helix structure upon binding on an SH3 domain; therefore it was possible to model the stereo structure of such a complex of the ScPex13p SH3 domain with its bound ligand ScPex14 peptide by comparison with other peptide-SH3 domain stereo structures, which were very well defined either by X-ray diffraction or by NMR spectrum.

A 16-amino acids peptide YEAMPPTLPHRDWKDY of ScPex14p (82Tyr-97Tyr) containing the proline rich region capable of binding on the ScPex13p SH3 domain was taken to search in the PDB protein database for a SH3 domain type-II ligand,

1ABO and 1AZG ligand peptides were were identified as ligand peptides which aligned very well with the ScPex14p peptide or the peptide of the reversed protein sequence of this ScPex14p peptide.

```
1ABO                    .APTMPPPLPP.
1AZG_pep                ..PPRPLPVAPGSSKT.
ScPex14p_pep            .YEAMPPTLPHRDWKDY
ScPex14p_pep_rev        YDKWDRHPLTPPMAEY.
```

Fig. 26 Alignment of Proline rich peptide of ScPex14p in PDB databank. In the alignment the binding points (PxxP) of the proline-rich peptide were conserved.

Based on the alignment shown in Fig. 26, the ScPex14p binding ligand for ScPex13p domain was modelled as an α-helix, and packed into the complex structure of the ScPex13p SH3 domain. A minimization of the complex energy was carried out afterwards to optimize the modelled complex structure. The optimized SH3 complex structure is shown in Fig. 27.

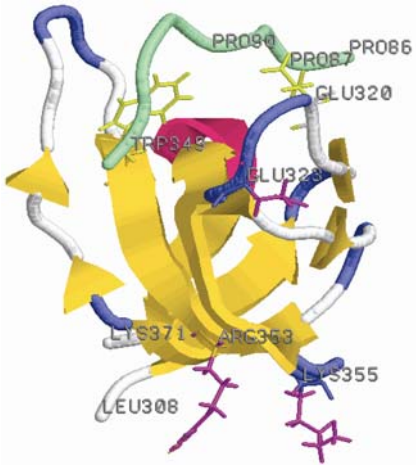


Fig. 27 Modelled structure of the ScPex13p SH3 domain complex with the ScPex14p proline rich ligand. The 3D structure of the ScPex13p SH3 domain complex with the ScPex14p ligand was modelled, and illustrated with the program RasMol.

3.7 ScPex13p SH3 Stereo Structure Generated by NMR

The solution structure of the Pex13p SH3 domain was determined by restrained simulated annealing using NMR-derived restraints. From a 3D ^{15}N NOESY-HSQC (150 ms mixing time), a 3D ^{13}C NOESY-HMQC (100 ms) , and two 2D NOESY (150 ms) in H_2O and D_2O a total of 1045 distance restraints were derived, of which 47 are medium- and 274 long-range. From ^{1}H and ^{13}C chemical shifts analysis further 40 phi and psi dihedrals restraints were derived and also 19 hydrogen-bond mimics based on

the NOE patterns indicative of secondary structure and H_2O/D_2O exchange experiments were included in the calculations.

As shown in Fig. 28, the 2D ^{15}N HSQC signals captured from the ^{15}N labeled ScPex13p SH3 domain sample were sorted automatically, subsequently were checked manually. The proline residues have no signal; all the histidine residues have just one signal; very tryptophan has 2 signals and very asparagines has 3 signals, every one of all the other residues of the SH3 (ScPex13p) domain has a signal.

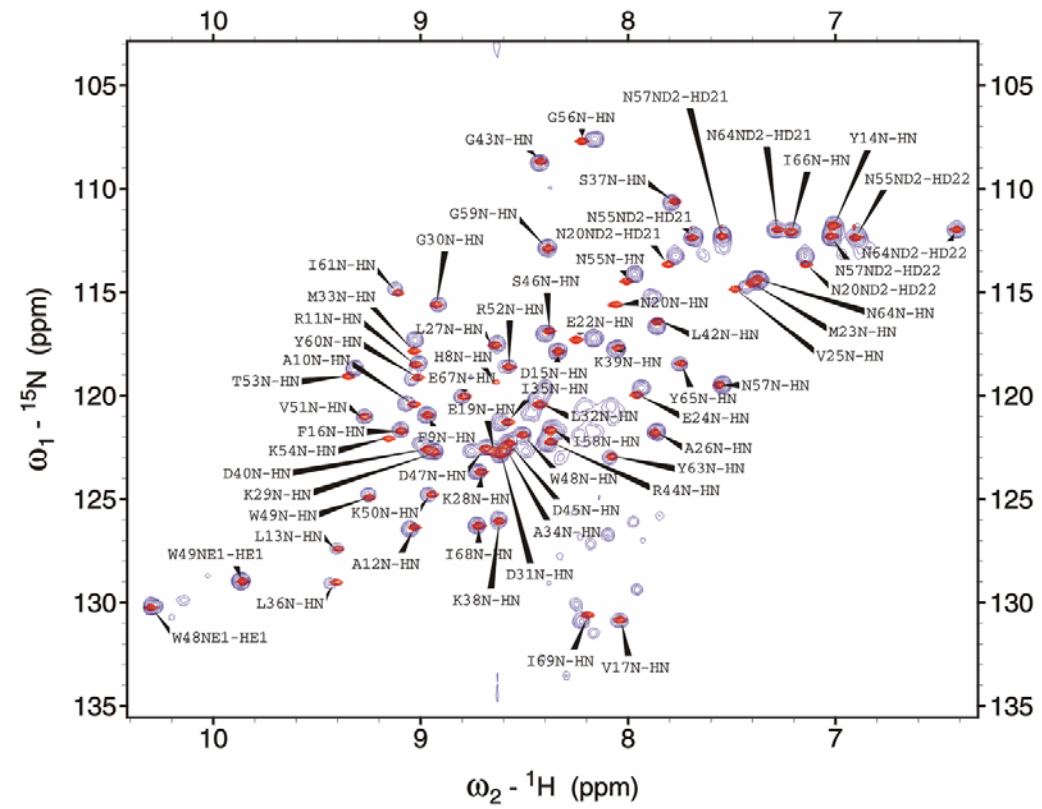


Fig. 28 15N HSQCs spectrum of ScPex13p SH3 domain. All the ^{15}N-^{1}H HSQC signals were sorted automatically or manually.

An ensemble of the 10 best structures (Fig. 29) were chosen from a total of 100 structures calculated by lowest total energy and smallest rmsd to average structure. The β-strands are colored in red, 3_{10} helix in green and loops are colored in cyan. Structures were superimposed by the α-carbons of the residues 310-318; 327-336; 349-354; 358-363 and 367-370 corresponding to the β-strand regions and part of the RT loop.

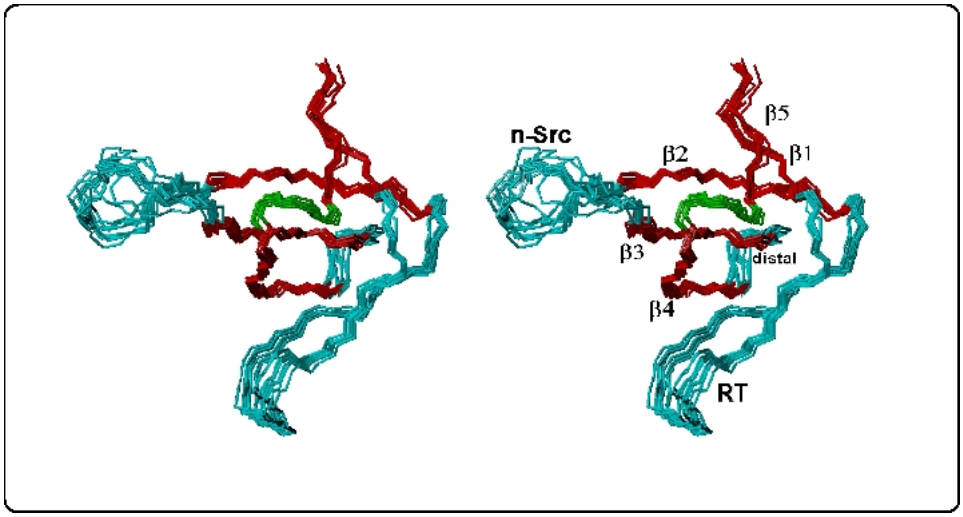


Fig. 29 Stereo illustration of NMR structure of the ScPex13p SH3 domain

For the region used in the alignment, the pair-wise rmsd obtained for this ensemble was 0.49 Å for the backbone and 1.12 Å for all heavy atoms. If the loop regions are also considered we obtain rmsds values of 0.79 Å and 1.46 Å, respectively. These values indicate a good convergence of the ensemble. A summary of the structure statistics is presented in table 3.

Table 3. Structure Statistics

A. Number of Restraints	
Intra-residual	507
Sequential	193
Medium-range	47
Long-range	274
Hydrogen Bonds (two restraints each)	19
Dihedral angles (ϕ,ψ)	40
Total	1123
B. RMSD from experimental restraints	
NOEs (Å)	0.05 ± 0.01
Dihedral angles	2.1 ± 0.3
C. CNS potential energy (kcal mol^{-1})	
E_{total}	402 ± 36
E_{bonds}	14 ± 2
E_{angles}	92 ± 10
$E_{impropers}$	17 ± 2
E_{vdW}	108 ± 10
E_{noe}	158 ± 25

D. Average pair-wise rmsd (Å)	
Excluding loop residues	
Backbone	0.49 ± 0.13
All non-H	1.12 ± 0.20
Including loop residues	
Backbone	0.79 ± 0.25
All non-H	1.46 ± 0.30

The mean structure out of the 10 best ScPex13p SH3 domain calculated above was obtained by averaging these chosen structures.

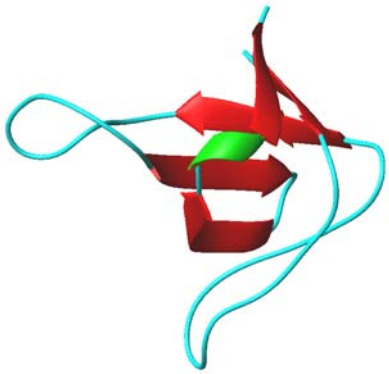


Fig. 30 optimized NMR structure of the ScPex13p SH3 domain.

3.8 Comparison of the Computer-derived Model and the NMR Structure

The NMR structure of the ScPex13p SH3 domain generated differed from the computer-derived models in several aspects. One notable difference between them, as expected, is the conformation of the n-Src loop. In both of the computer models of the ScPex13p SH3 structure this loop extends away from the Trp349 region, which is part of the Pex14p PXXP motif binding pocket. However, the NMR studies pointed out that part of the n-Src loop is in proximity to the Trp349 region. The fact that the chemical shifts (^{15}N-HSQC signal) of amino acids 339K, 340K, 341D and 345G of the n-Src loop changed if the N^{15} labeled SH3 domain solution was added with the ScPex14p peptide containing PxxP motif, indicates that the n-Src loop is also involved in the binding with the ScPex14p PxxP motif.

This difference regarding the n-Src loop of the ScPex13p SH3 structure in the computer model is caused by the lack of an appropriate template for modelling (see Fig. 24).

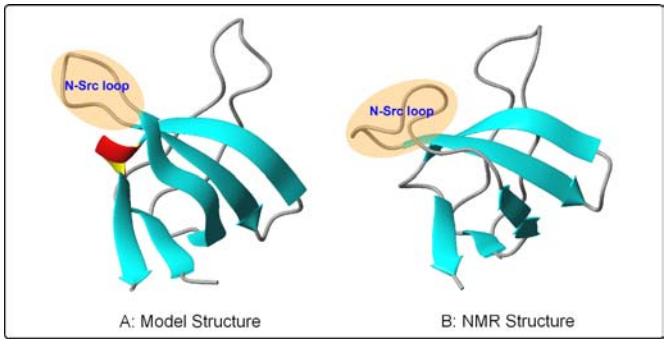


Fig. 31 Comparing of SH3 structures by modelling and by NMR. The major difference of both structures lies in n-Src loop, marked in brown shadow.

3.9 Interaction between ScPex13p and ScPex14p

In yeast, Pex14p is a peroxisomal peripheral membrane protein involved in peroxisome biogenesis and translocation of peroxisome matrix proteins. In *Saccharomyces cerevisiae* ScPex14p forms together with ScPex13p the docking complex at the peroxisomal membrane for import of peroxisomal matrix proteins.

3.9.1 Mapping the ScPex14p Binding Site for the ScPex13p SH3 Domain

Sequence analysis of the ScPex14p protein shows that the protein contains a classic PXXP motif, which locates between Pro86 and Arg92 (PPTLPHR)

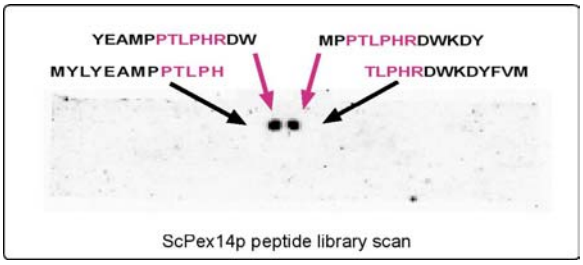


Fig. 32 Localization of the ScPex13p SH3 domain binding site in ScPex14p. Two spots on the ligand membrane were shown to be capable to interact with the GST fused SH3 domain protein. Detection was performed by ligand blot analysis using a monoclonal antibody against GST. The peptide sequences of these two spots are YEAMPPTLPHRDW and MPPTLPHRDWKDY respectively with a consensus sequence accommodating the PxxPxR motif (type II).

ScPex14p derived 12-mer synthetic peptides covering the entire protein sequence, were tested for their interaction with the SH3-domain of Pex13p by ligand blot analysis. The ScPex14p peptide library was immobilized on a cellulose membrane and incubated with GST-tagged ScPex13p SH3 domain. The interaction between

peptides and the SH3 domain was detected immunologically with a monoclonal antibody against GST. As shown in Figure 33, a proline-rich motif of Pex14p representing a typical type II SH3-ligand motif was identified as the SH3 domain binding site in Pex14p. The NMR studies later also confirmed that this proline-rich peptide of Pex14p binds to SH3 domain of Pex13p in vitro.

3.9.2 Location of the ScPex14p Binding Site on the ScPex13p SH3 Domain

The binding sites of the ScPex14p proline rich PxxP motif on the ScPex13p SH3 domain were located by ^{15}N-HSQC NMR spectroscopy. The ^{15}N-HSQC signals contain a "fingerprint" of the protein, i.e., a pair of ($^{1}H^{N}$, ^{15}N) frequencies for each residue in the protein. First a ^{15}N-HSQC NMR spectrum, colored red in Fig. 33, was taken with a sample of 0.7 mM ^{15}N labeled ScPex13p SH3 domain, which was purified by the protocol described above. Then the ScPex14p peptide Ac-MPPTLPHRDW-NH2 presenting the PxxP motif was added to the ScPex13p SH3 sample in a molar ratio of 1:2. The chemical shifts of the ^{15}N-HSQC signals, caused by the changes of the chemical environment of the residues in ScPex13p SH3 domain, indicate the residues involved in binding of the ScPex14p peptide. Strong shifts ($\Delta\delta_{total}$) of the amide signals of the ScPex13p SH3 domain observed in this experiment were Trp349, Tyr361, Ile362, Asn365 and Ile367 in the range $0.39 \leq \Delta\delta_{total} < 1.1$ ppm, other peaks showed moderate shifts in the range $0.1 \leq \Delta\delta_{total} < 0.39$ ppm like Ala313, Leu314, Tyr315, Phe317, Val318, Asn321, Glu323, Met324, Val325, Ala326, Leu327, Lys330, Lys339, kys340, Asp341, Arg345, Asp348, Trp350, Lys351, Tyr366 and Glu368. During the addition of the ScPex14p peptide into the ^{15}N labeled ScPex13p SH3 domain some cross-peaks also appeared, due to intermediate exchange (eg. indole NH of W349). These disappeared again after the stable ligand-domain complex was formed. Such an experiment indicated that the ScPex14p PxxP motif binds with the ScPex13p SH3 domain through the hydrophobic pocket formed by hydrophobic amino acids such as Trp349 on the ScPex13p SH3 domain.

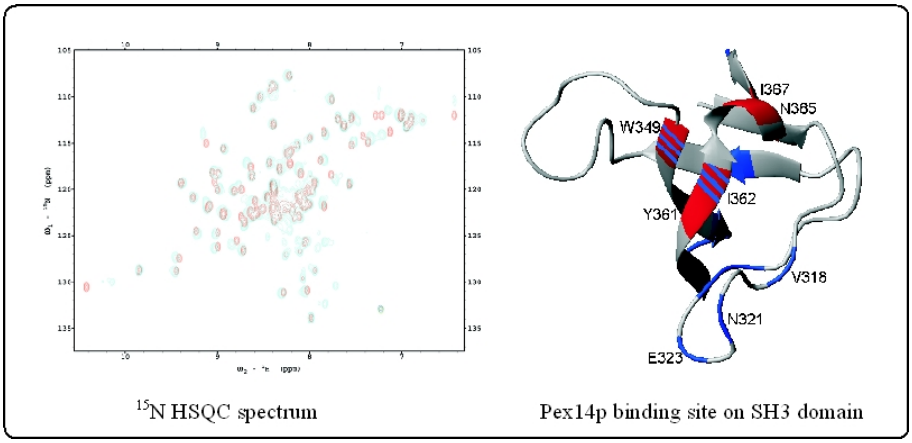


Fig. 33 Chemical shift of amino acids of the SH3 domain upon addition of ScPex14p peptide. The ^{15}N-HSQC NMR spectrum of ^{15}N marked SH3 domain protein alone is shown in red, while its ^{15}N-HSQC NMR spectrum in condition with the ScPex14p ligand peptide is presented in cyan.

3.9.3 Modelling of the SH3-ScPex14 Peptide Complex Based on the SH3 NMR Structure

Many of the complex structures of SH3 domains and their typical ligands were well characterized by either X-ray crystallography or NMR spectroscopy. Due to technical reasons a NMR structure of the ScPex13p SH3 domain with its ScPex14p binding ligand, one of the type II ligand for SH3 domains, was not feasible.

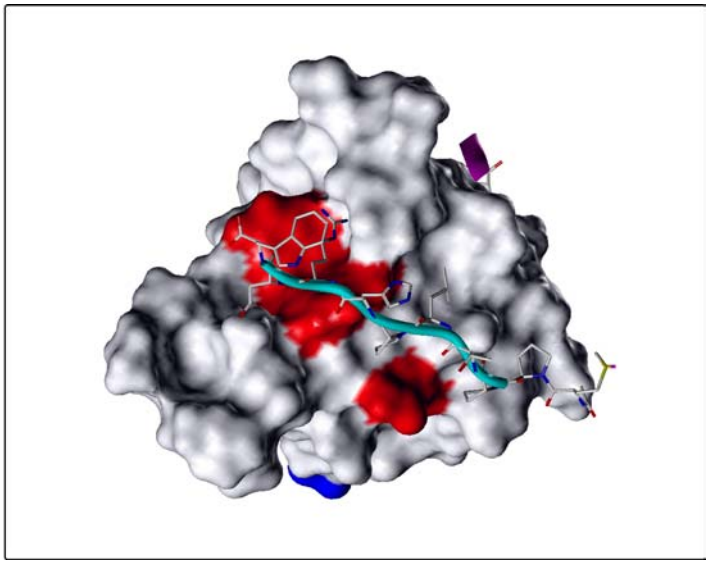


Fig. 34 Model of the SH3(ScPex13p)-ScPex14p peptide complex based on SH3 NMR. The structure of the SH3(ScPex13p)-ScPex14 peptide complex was modelled and illustrated. The binding region of the ScPex14p peptide on the ScPex13p SH3 domain was shown in red and the ScPex14p peptide was shown in cyan.

However, this limitation could be compensated by computer modelling of the ligand-domain complex of ScPex13p SH3 domain and its ligand PxxP peptide of the ScPex14p protein, which could be easily established based on the well optimized NMR structure of the ScPex13p SH3 domain alone and the well defined conformation of a type II ligand in the SH3-ligand complex structure.

Compared with the model before, the structure of the Pex13p SH3 domain, especially the conformation of the long n-Src loop, was well defined by NMR. Its ligand – the Pex14p peptide – was modelled as a typical type II α-helix ligand for SH3 domains. The binding sites of its ligand – the Pex14p peptide – were determined by the chemical shift of the residues in the Pex13p SH3 domain by acquisition of several ^{15}N HSQC spectra upon titration with the Pex14p-derived peptides. The orientation of the Pex14p peptide on the Pex13p SH3 domain was adapted by the typical orientation of a type II ligand for SH3 domain and further confirmed and optimized by minimization of the molecular energy of the complex. The resulting structure is shown in Fig. 34.

3.10 Interaction between ScPex13p and ScPex5p

An interesting unique character of the ScPex13p SH3 domain is that it can bind two distinct ligands. The first one is Pex14p, which binds to the ScPex13p SH3 domain via its proline rich SH3 ligand motif of type II. However, the SH3 domains of Pex13p is also able to interact with a ScPex5p peptide, a non-PxxP motif peptide.

3.10.1 Mapping the Binding site of ScPex13p SH3 in ScPex5p

Yeast two-hybrid technique and peptide library screening were used to reveal the binding motif of ScPex13p SH3 in ScPex5p protein.

3.10.1.1 Mapping the Binding Site of ScPex13p SH3 in ScPex5p by two-the hybrid Methodology

Various truncated forms of Pex5p were investigated for their ablity to interact with the SH3-domain of Pex13p in the yeast two-hybrid system. It was expected that the SH3 binding region in ScPex5p could be delimited to a shorter peptide.

For the two-hybrid analysis, the following fragments of ScPex5p: ScPex5pF1 (1-312aa), ScPex5pF2 (1-180aa), ScPex5pF3 (181-312aa), ScPex5pF4 (181-240aa),

ScPex5pF5 (241-312aa), ScPex5pF6 (201-226aa), cloned into pPC86 (GAL4 activation domain fusion vector), and the ScPex13p fragments ScPex13pSH3 (309-370aa) and ScPex13pSH3F (285-386aa), cloned into pPC97 (DNA binding domain fusion vector), were investigated for their interaction by two-hybrid analyses. In this two-hybrid system, the shortest ScPex5p fragment showing strong interaction with ScPex13p SH3 domain is ScPex5pF3, while the shorter fragment ScPex5pF4 showed very weak interaction. Therefore, it could be concluded that the SH3-binding region in ScPex5p is limited to a protein fragment comprising amino acids 180-312 (Fig. 35).

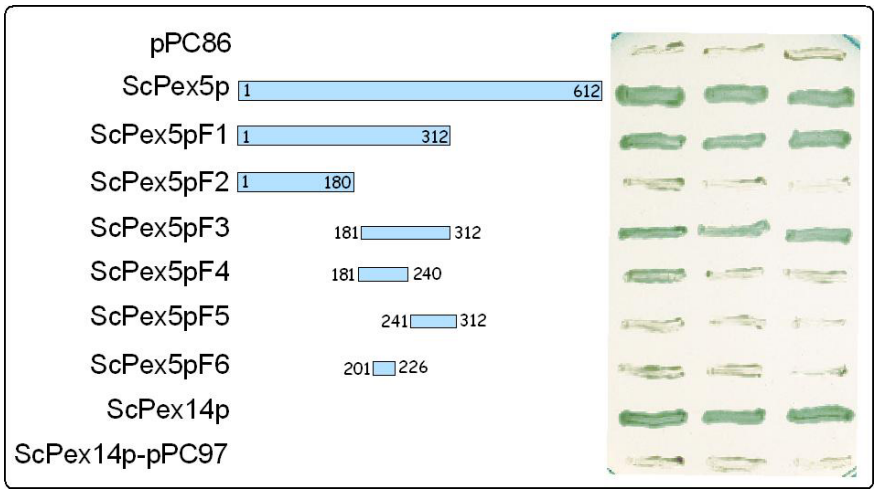


Fig. 35 Two-hybrid analysis of interaction between ScPex5p and ScPex13p SH3 domain. The interactions between the ScPex13p SH3 domain and ScPex5p fragments or ScPex14p were investigated by the two-hybrid analyses; the interaction was detected by the β-galactosidase filter assay. The complete ScPex5p and ScPex14p showed very strong interaction with the ScPex13p SH3 domain. The shortest fragment of ScPex5p showing interaction with the ScPex13p SH3 domain is ScPex5pF4.

3.10.1.2 Mapping the Binding Site by Ligand Blot Analysis

Since the ScPex5p fragments shorter than ScPec5pF4 showed no interaction with ScPex13p SH3 domain in two-hybrid system, it could not be excluded that the short ScPex5p fragment expressed from the two-hybrid vector folded correctly in yeast cell. Therefore, an alternative in vitro experiment was exploited to check whether the ScPex13p SH3 domain binding site in ScPex5p locates to an even smaller region than the one contained in ScPex5pF4. To analyze the binding sites of the SH3-ligands Pex5p in more detail, Pex5p derived overlapping 26-mer synthetic peptides covering the putative binding region were immobilized on a cellulose membrane. The membrane was incubated with GST fused ScPex13p SH3 domain purified from the

E.coli cells. The GST fused SH3 domain bound on the peptides on the membrane was detected immunologically with monoclonal GST antibodies. The SH3-binding site for Pex5p was unambiguously mapped to the non-PXXP sequence QPWTDQFEKLEKEV that represents amino acids 202-215 of Pex5p.

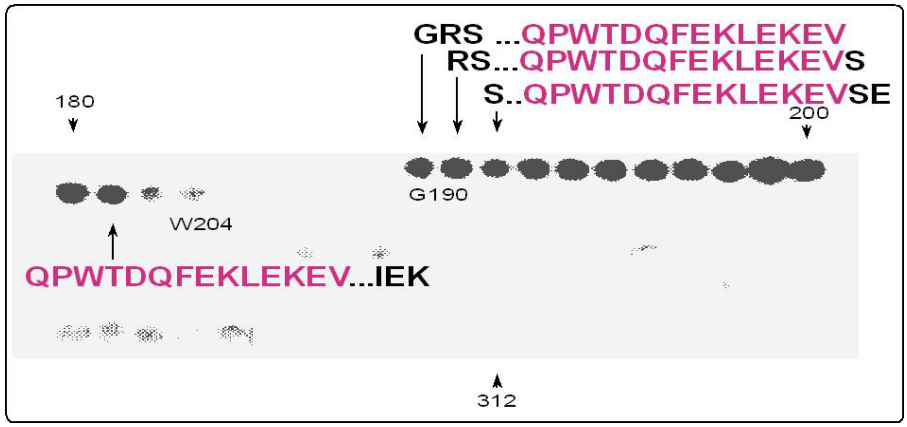


Fig. 36 Map of the binding site of ScPex5p for ScPex13p SH3 domain. Cellulose membranes decorated with 26-mer peptides covering amino acids 181-312 of ScPex5p with 1-amino acid shifts between neighboring peptides were incubated with the recombinant GST-tagged SH3 domain of ScPex13p. Peptide-SH3-complexes were identified by the peptide spot overlay assay using antibodies against GST.

3.10.1.3 Substitution and Truncation Analyses by Ligand Blot

Substitution and truncation experiments were carried out with the peptide libraries based on the sequence VNEQEQQPWTDQFEKLEKEVSENLDI (26 aa) which represents amino acids 195 to 221 of Pex5p. The truncation analysis revealed that the first eight amino acids of the peptide can be replaced by any other amino acid without significant loss of binding capacity while Trp204 and and Phe208 are indispensable for binding (Fig. 37). Furthermore, no amino acid of the central region of 13 amino acids (WTDQFEKLEKEVS) could be functionally replaced by prolin, which would interfere with the proposed alpha-helical conformation of this region. Also glycin cannot substitute for any of the amino acids of the central region.

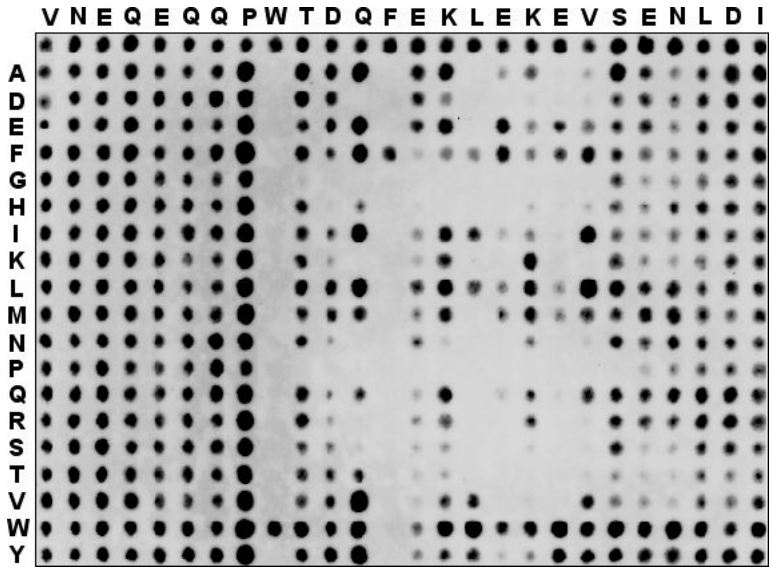
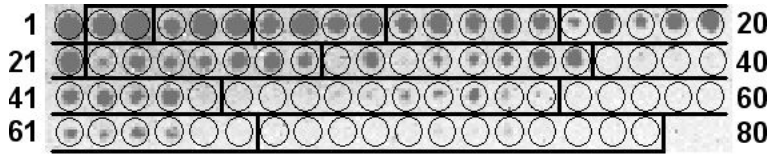


Fig. 37 Substitution analysis. The first row of spots represents the original peptide, representing amino acids 195 to 221 of Pex5p (VNEQEQQPWTDQFEKLEKEVSENLDI, 26 aa), and the spots at any row following peptides with single substitutions in the respective positions.

A truncation analysis based on the 26-peptide (ScPex5p 195-221 aa) was also carried out. Shorter peptides truncated from right or left were still functional as long as the central region was not affected (Fig. 38). The shortest peptide which still did interact consists of 16 amino acids containing the central region. Binding to the SH3 domain of the short (QQPWTDQFEKLEKEVS; 16aa) and a long peptide (QQPWTDQFEKLEKEVSENLDI; 21aa) were further investigated by NMR studies.



Sequence of the spots

1. VNEQEQQPWTDQFEKLEKEVSENLDI
2. VNEQEQQPWTDQFEKLEKEVSENLD
3. NEQEQQPWTDQFEKLEKEVSENLDI
4. VNEQEQQPWTDQFEKLEKEVSENL
5. NEQEQQPWTDQFEKLEKEVSENLD
6. EQEQQPWTDQFEKLEKEVSENLDI
7. VNEQEQQPWTDQFEKLEKEVSEN
8. NEQEQQPWTDQFEKLEKEVSENL
9. EQEQQPWTDQFEKLEKEVSENLD
10. QEQQPWTDQFEKLEKEVSENLDI
11.VNEQEQQPWTDQFEKLEKEVSE
12. NEQEQQPWTDQFEKLEKEVSEN
13. EQEQQPWTDQFEKLEKEVSENL
14. QEQQPWTDQFEKLEKEVSENLD
15. EQQPWTDQFEKLEKEVSENLDI
16.VNEQEQQPWTDQFEKLEKEVS
17. NEQEQQPWTDQFEKLEKEVSE
18. EQEQQPWTDQFEKLEKEVSEN
19. QEQQPWTDQFEKLEKEVSENL
20. EQQPWTDQFEKLEKEVSENLD
21. QQPWTDQFEKLEKEVSENLDI
22.VNEQEQQPWTDQFEKLEKEV
23. NEQEQQPWTDQFEKLEKEVS
24. EQEQQPWTDQFEKLEKEVSE
25. QEQQPWTDQFEKLEKEVSEN
26. EQQPWTDQFEKLEKEVSENL
27. QQPWTDQFEKLEKEVSENLD
28. QPWTDQFEKLEKEVSENLDI
1. VNEQEQQPWTDQFEKLEKEVSENLDI
29.VNEQEQQPWTDQFEKLEKE
30. NEQEQQPWTDQFEKLEKEV
31. EQEQQPWTDQFEKLEKEVS
32. QEQQPWTDQFEKLEKEVSE
33. EQQPWTDQFEKLEKEVSEN
34. QQPWTDQFEKLEKEVSENL
35. QPWTDQFEKLEKEVSENLD
36. PWTDQFEKLEKEVSENLDI
37.VNEQEQQPWTDQFEKLEK
38. NEQEQQPWTDQFEKLEKE
39. EQEQQPWTDQFEKLEKEV
40. QEQQPWTDQFEKLEKEVS
41. EQQPWTDQFEKLEKEVSE
42. QQPWTDQFEKLEKEVSEN
43. QPWTDQFEKLEKEVSENL
44. PWTDQFEKLEKEVSENLD
45. WTDQFEKLEKEVSENLDI
46.VNEQEQQPWTDQFEKLE
47. NEQEQQPWTDQFEKLEK
48. EQEQQPWTDQFEKLEKE
49. QEQQPWTDQFEKLEKEV
50. EQQPWTDQFEKLEKEVS
51. QQPWTDQFEKLEKEVSE
52. QPWTDQFEKLEKEVSEN
53. PWTDQFEKLEKEVSENL
54. WTDQFEKLEKEVSENLD
55. TDQFEKLEKEVSENLDI
1. VNEQEQQPWTDQFEKLEKEVSENLDI
56.VNEQEQQPWTDQFEKL
57. NEQEQQPWTDQFEKLE
58. EQEQQPWTDQFEKLEK
59. QEQQPWTDQFEKLEKE
60. EQQPWTDQFEKLEKEV
61. QQPWTDQFEKLEKEVS
62. QPWTDQFEKLEKEVSE
63. PWTDQFEKLEKEVSEN
64. WTDQFEKLEKEVSENL
65. TDQFEKLEKEVSENLD
66. DQFEKLEKEVSENLDI
67.VNEQEQQPWTDQFEK
68. NEQEQQPWTDQFEKL
69. EQEQQPWTDQFEKLE
70. QEQQPWTDQFEKLEK
71. EQQPWTDQFEKLEKE
72. QQPWTDQFEKLEKEV
73. QPWTDQFEKLEKEVS
74. PWTDQFEKLEKEVSE
75. WTDQFEKLEKEVSEN
76. TDQFEKLEKEVSENL
77. DQFEKLEKEVSENLD
78. QFEKLEKEVSENLDI

Fig. 38 Truncation Analysis. Possible truncations by 1-11 amino acids from right or left of the original peptide were examined for their ability to interact with the SH3

domain. The peptide sequences corresponding to the spot number are given below. Peptides 21 and 61 were further investigated by NMR spectroscopy and determined binding constants were 640µM and 2mM, respectively.

3.10.2 Localization of the Binding Site on ScPex13p SH3 for ScPex5p

The binding site of ScPex5p peptide on the ScPex13p SH3 was located also by analyzing the chemical shift of the amide groups of a ^{15}N labeled SH3 domain acquired by NMR spectrum. Changes in the chemical shifts of the amide groups of a ^{15}N labeled SH3 domain construct were followed by acquisition of several ^{15}N HSQC spectra (Kay, 1995) upon titration with the Pex5p-derived peptides according to the result above. The strongest signal shifts observed by addition of Pex5p in a molar ratio 1:2 were in the range of $0.1 < \Delta\delta_{total} < 0.3$ ppm, which covers the amino acids Val318, Asn321, Glu323, Glu325, Val326, Met334, Trp349, Thr354, Lys355, Asn356 and Ile362. Moderate shifts in the range of $0.07 \leq \Delta\delta_{total} \leq 0.1$ ppm were observed for the amide groups Ala311, Ala313, Ala327, Leu333, Ser347, Asp348, Gly357 and Tyr361. According to these chemical shifts of the amino acids on the SH3 domain, the binding site for the Pex5p peptide on the Pex13p SH3 domain was mapped and located.

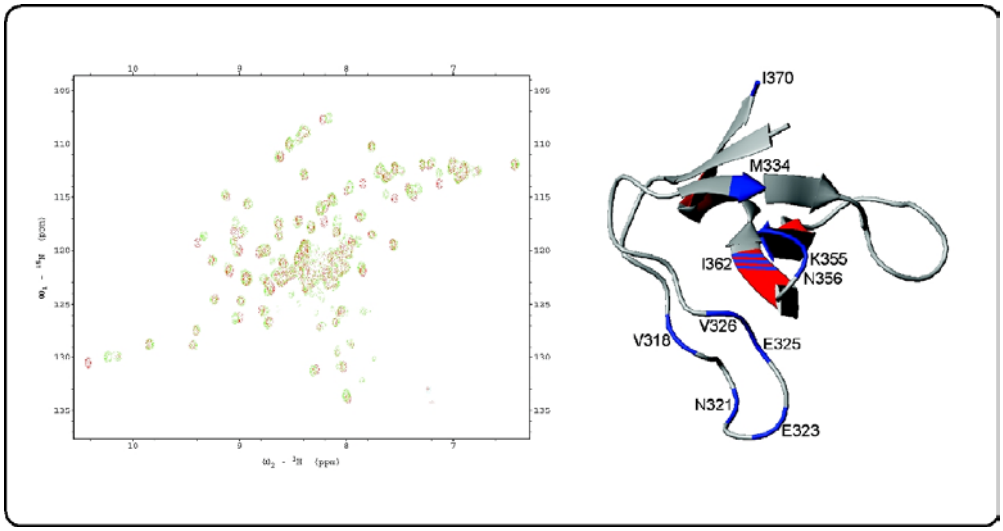


Fig. 39 Binding site of ScPex5p on the ScPex13p SH3 domain. The left side shows a ^{15}N HSQC spectra of ScPex13p SH3: The ^{15}N HSQC signals of all the amino acids of ScPex13 SH3 without addition of the ScPex5p peptide are colored with red; while the signals obtained upon binding of the ScPex5p peptide are colored with cyan. At the right side, a ribbon illustration of the stereo structure of the ScPex13p SH3 is shown, where the binding sites for the ScPex5p peptide were labeled in red color.

3.10.3 Modelling of the SH3-ScPex5p Peptide Complex Based on the SH3 NMR Structure

The Pex5p peptide has the potential to build an α-helix. It's shown by CD-spectrum analysis that this peptide really forms an α-helix while bound onto the Pex13p SH3 domain, however it did not form an α-helix in the absence of the ScPex13p SH3 domain. The complex structure of the Pex13p SH3 domain and its α-helix ligand ScPex5p peptide could be illustrated by computer modelling, which could not be calculated by NMR spectroscopy because of the weak binding constants between the Pex5p peptide and the SH3 domain. The structure of the ScPex13p SH3 domain in this ligand-domain complex was adopted from its NMR structure, the ligand in this complex – the Pex5p peptide – was modelled as an α-helix and the binding sites of the ligand on the SH3 domain were determined by the chemical shift of the residues in the Pex13p SH3 domain by acquisition of several ^{15}N HSQC spectra upon titration with the Pex5p-derived peptides. The orientation of the Pex5p peptide on the Pex13p SH3 domain was determined and optimized by minimization of the complex molecular energy. The resulting structure is shown in Fig 41.

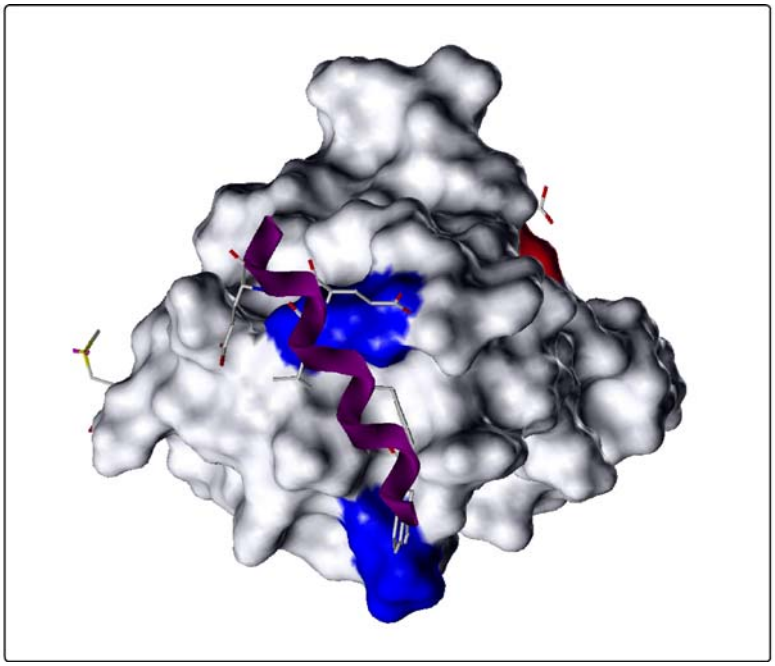


Fig. 40 Complex structure of the ScPex13p SH3 domain and the bound ScPex5p peptide. In this model, the ScPex13p SH3 domain structure is based on the NMR structure and its ScPex5p ligand peptide was modelled as an α-helix on the SH3 surface.

3.10.4 Localization of the Second Binding Site in ScPex13p for ScPex5p

The Binding site in ScPex13p for ScPex5p was also looked by in vitro peptide library screen. A peptide library of 20 aa peptides covering the ScPex13p protein sequence with a 2-amino-acids shift was constructed on a cellulose membrane. The cellulose membrane with the ScPex13p peptide library was incubated with the GST fused ScPex5p protein, subsequently the attached GST-ScPex5p was detected immunologically with monoclonal antibodies against GST. This experiment turned out that a 22-amino-acid region in front of the N-terminus of the SH3 in ScPex13p could also interact with ScPex5p providing a second binding site besides the SH3 domain in the ScPex13p protein for ScPex5p. It should be pointed out that non peptide of SH3 domain showed interaction with GST-ScPex5p, because the binding site in the ScPex13p SH3 domain for ScPex5p can only be functional when the SH3 domain folds properly.

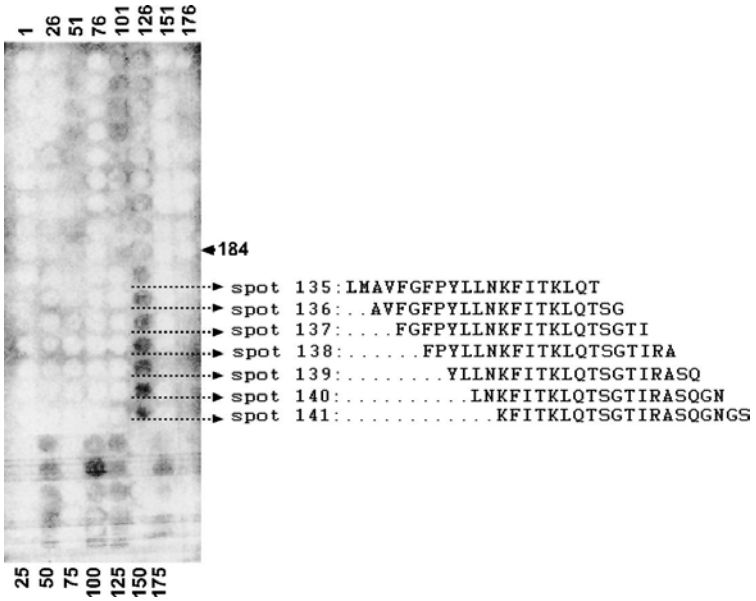
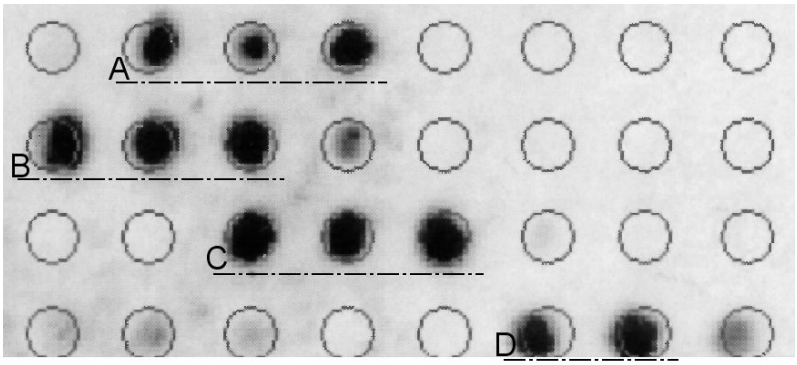


Fig. 41 Peptide library screen in the ScPex13p protein sequence for binding sites for ScPex5p. A peptide library of 20 aa peptides covering the ScPex13p protein sequence with a 2-amino-acids shift was constructed on a cellulose membrane, interaction with GST-ScPex5p was investigated by immunoblot with monoclonal antibodies against GST. The black spots, sequences of which are listed at the right side, presented the peptides being able to interact with GST-ScPex5p. The isolated black spot at the bottom of the membrane is a fault positive.

Based on the results of the screen of the ScPex13p peptide library, the 22-amino-acid peptide (ScPex13p: 269-290 aa) was used as a start point to be truncated from left or right side to find out the shortest peptide still being able to interact with GST-ScPex5p. A library of all possible peptides resulted from the truncation was immobilized on a cellulose membrane, and the interaction was immunologically detected by monoclonal antibodies against GST. It showed that the shortest peptides to interact with GST-ScPex5p are 10-aa peptides TKLQTSGTIR and KLQTSGTIRA.



A: ITKLQTSGTIRAS
TKLQTSGTIRASQ
KLQTSGTIRASQG

B: FITKLQTSGTIR
ITKLQTSGTIRA
TKLQTSGTIRAS

C: ITKLQTSGTIR
TKLQTSGTIRA
KLQTSGTIRAS

D: TKLQTSGTIR
KLQTSGTIRA

Fig. 42 Truncation experiment of the ScPex13p protein sequence to locate the binding site for ScPex5p. A peptide library of all possible peptides resulted from the truncation of the 22-aa peptide was immobilized on a cellulose membrane; interaction with GST-ScPex5p was investigated by immunoblot with monoclonal antibodies against GST. The shortest peptides still being able to bind to GST-ScPex5p were located in "D" block.

3.11 Competition between ScPex5p and ScPex14p upon Binding to the SH3 Domain

3.11.1 In vitro Binding Assays

The histidine fusion proteins His-ScPex5p and His-ScPex14p were heterologously overexpressed in the bacteria *E. coli* BL21(DE3) cells, and subsequently purified by Ni-NTA affinity chromatography. The GST fusion protein GST-SH3 domain was also expressed in the bacterial *E. coli* BL21(DE3) cells and subsequently purified by GST Sepharose affinity chromatography. GST-SH3 was bound on the GST Sepharose resin, and purified His-ScPex14p was added. After washing, the column was eluted with 10

mM reduced glutathione buffer. His-ScPex14p was found together with GST-SH3 domain in the eluate indicating the in vitro interaction of both proteins. After it was verified that His-ScPex14p can bind on GST-SH3 domain, another experiment was carried out, in which GST-SH3 fusion protein was bound on GST Sepharose resin, His-ScPex14p was transferred into this GST-column; and after washing, 1.5 and 5 times (molar concentration) His-ScPex5p was added sequentially. His-ScPex14p appeared the flow-through of 5 times His-ScPex5p; mean while in the eluate with 10 mM reduced glutathione, no His-ScPex14p but GST-SH3 fusion protein and His-ScPex5p was found.

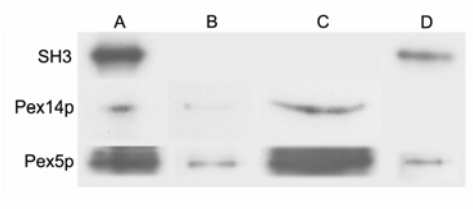


Fig. 43 In vitro binding of ScPex5p, ScPex14p to the ScPex13 SH3 domain. A: positive control of ScPex5p, ScPex14p and ScPex13 SH3 domain; B: flow-through of 1.5 times ScPex5 flowing through GST Sepharose column bound with GST fusion ScPex13 SH3 domain; C: flow-through of 5 times ScPex5p flowing through GST Sepharose column bound with GST fusion ScPex13 SH3 domain; D: eluate of the GST Sepharose column with 10 mM reduced glutathione.

3.11.2 Binding Competition Assays by NMR Experiments

3.11.2.1 Chemical Shift of SH3 Domain Amino Acids upon Sequential Binding of ScPex14p and ScPex5p Peptides

While acquisition of ^{15}N HSQC spectra of ^{15}N labeled ScPex13p SH3 domain (0.7 mM solution) strong chemical shift of NMR signal of amino acids N365 and I362 were observed if the PxxP rich peptide derived from ScPex14p was added into the ScPex13p SH3 solution. This indicates that the ScPex14p derived peptide bound on the ScPex13p SH3 domain surface, which led to changes of the chemical environment of Asn365 and Ile362. It was very interesting that these shifted NMR signal of N365 and I362 of the ScPex13p SH3 domain in presentation of the ScPex14p derived peptide tended to shift back if excessive amount of the ScPex5p derived peptide was introduced into the mixture; in other words the ScPex14p derived peptide was

released from the surface of the ScPex13p SH3 domain if the ScPex5p derived peptide was added excessively.

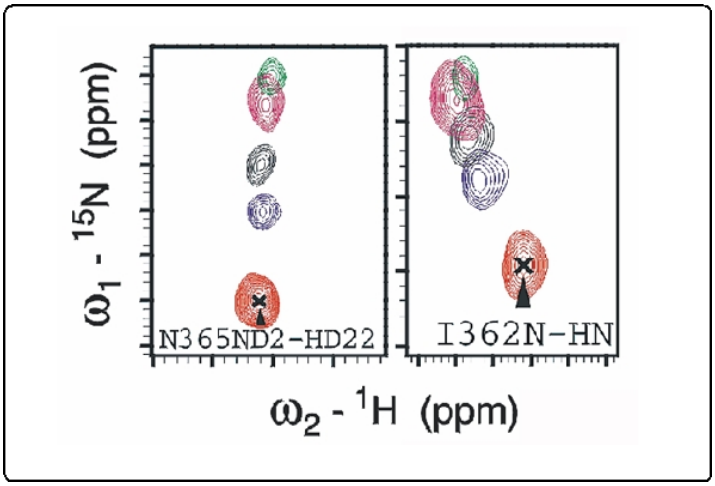


Fig. 44 Chemical shift of N365, I362 of the ScPex13p SH3 domain upon sequentially binding of the ScPex14p and ScPex5p peptides. The NMR signals of N365 and I362 of the ScPex13p SH3 domain labeled with ^{15}N were shifted from the position labeled with red to the one marked in green, when the ScPex14p derived peptide was added in a molar ration of 1:2 of ScPex13p SH3 domain to ScPex14p peptide; these signals tended to shift back if excessive ScPex5p peptide was added into the system with a molar ratio of 1:20 (magenta), 1:30 (black) and 1:40 (blue) of the ScPex13p SH3 domain to ScPex5p peptide.

3.11.2.1 Chemical Shift of SH3 Domain Amino Acids upon Sequential Binding of ScPex5p and ScPex14p Peptides

In a similar experiment like the one described above, this time Pex5p was bound first to the SH3 domain, then Pex14p was added. Upon Pex5p binding strong chemical shifts of NMR signals of amino acids M334 and N356 in the ^{15}N HSQC spectra of ^{15}N labeled ScPex13p SH3 domain were observed. As shown in Fig. 45, the NMR signals of these both amino acids of ScPex13p SH3 domain were shifted from the position labeled with red color to position in green color if the ScPex5p derived peptide was added to the SH3 solution, indicating that the ScPex5p derived peptide bound on the ScPex13p SH3 domain surface, which led to changes of the chemical environment of Met334 and Asn356. However, these shifted NMR signal of M334 and N356 of the ScPex13p SH3 domain did not shift back in the presence of the ScPex14p derived peptide in an excessive amount; that is to say, the binding of ScPex15p derived peptide on the ScPex13p SH3 domain was not disturbed by addition of the ScPex14p derived peptide.

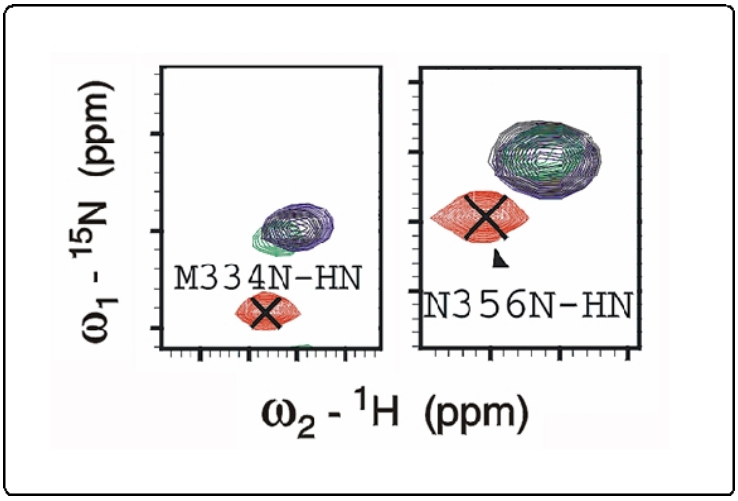


Fig. 45 Chemical shift of M334 and N356 of ScPex13p SH3 domain upon sequential binding of ScPex5p and ScPex14p peptide. The NMR signals of M334 and N356 of the ScPex13p SH3 domain labeled with ^{15}N were shifted from the position labeled with red to the one marked in green when the ScPex5p derived peptide was added in a molar ration of 1:2 of ScPex13p SH3 domain to ScPex5p peptide; these signals did not intend to shift back if excessive ScPex14p peptide was added into the system with a molar ratio of 1:1 (black) and 1:3.5 (blue) of the ScPex13p SH3 domain to ScPex14p peptide.

3.12 In vitro Research on Interactions among Pex5p, Pex14p and the Pex13p SH3 Domain

3.12.1 Interaction between Pex5p and Pex14p

It has been reported earlier that Pex5p and Pex14p interact with each other in vivo. In this work the interaction between Pex5p and Pex14p was further investigated by in vitro methods.

3.12.1.1 Ligand Blot Analysis.

The heterologously expressed human Pex5p (long form HsPex5pl and short form HsPex5ps) and Pex14p (HsPex14p) showed interaction by in vitro ligand blot analysis, in which the purified HsPex5pl and HsPex5ps were immobilized on a nitrocellulose membrane, and then incubated with the bacteria cell lysate containing heterologously expressed HsPex14p. The interactions between HsPex5ps or HsPex5pl and HsPex14p were detected immunochemically with antibodies against HsPex14p.

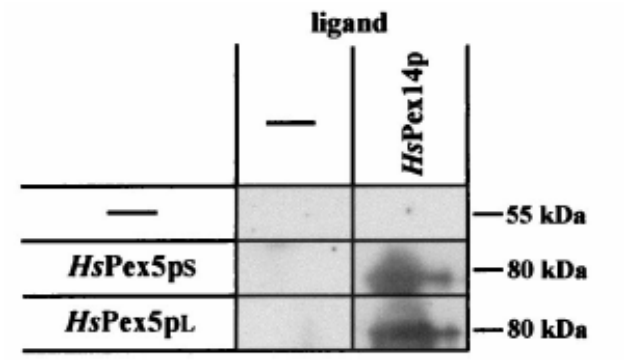
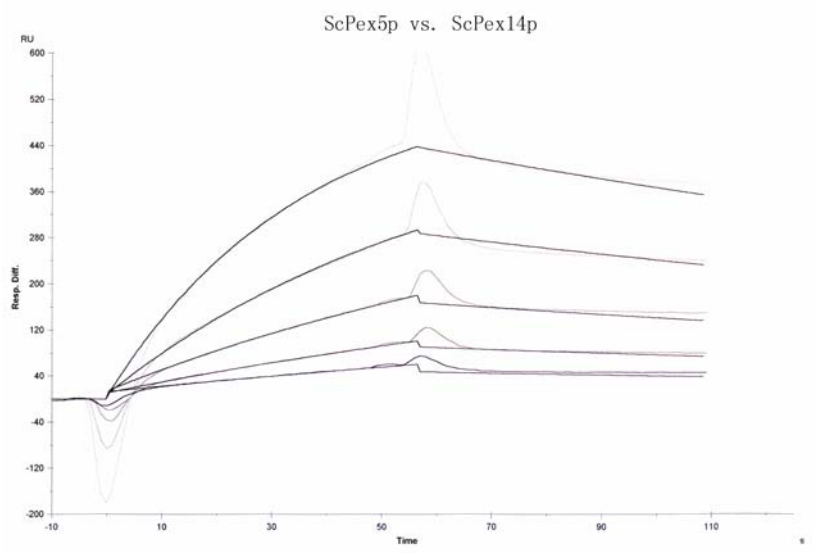


Fig. 46 Ligand blot analysis of the HsPex14p interaction with HsPex5pl and HsPex5ps. Bacterial lysates (20 mg protein) containing HsPex5l, HsPex5ps, or no recombinant protein were subjected to SDS-PAGE and transferred to nitrocellulose. Individual membranes were incubated with buffer containing either purified His6-tagged HsPex14p (1 mg) or no recombinant protein. HsPex14p-containing complexes were visualized by immunoblotting with anti-HsPex14p antibodies.

3.12.1.2 Binding Constants between Pex5p and Pex14p

The purified His-tagged ScPex5p and ScPex14p were used for the determination of the binding constants between the two proteins by BIAcore measurements. Both of His-tagged ScPex5p and ScPex14p were heterologously expressed in bacteria *E. coli* and purified by Ni-NTA affinity chromatography; the concentration of the purified proteins was determined by the Biorad protein assay kit.

To investigate the interaction between Pex5p and Pex14p by BIAcore studies, the purified His-tagged ScPex14p was immobilized on a CM5 chip with negative charged carboxyl groups by electrostatic attraction at pH values below the ligand pI. The amount of Pex14p immobilized on the chip was controlled so that the background RU (Reflection Unit) lies around 500 units. A series of purified Pex5p protein solutions with increasing concentration was introduced into the flow phase, while the changes of the RU were recorded, see Fig. 47. The binding constant between these two proteins, was calculated by the BIAevaluation software. The dissociation binding constant for Pex5p and Pex14p was 33.5 pM,



ScPex5p Concentration (nM)	12.5	25	50	100	200
RU	173	271	380	476	544

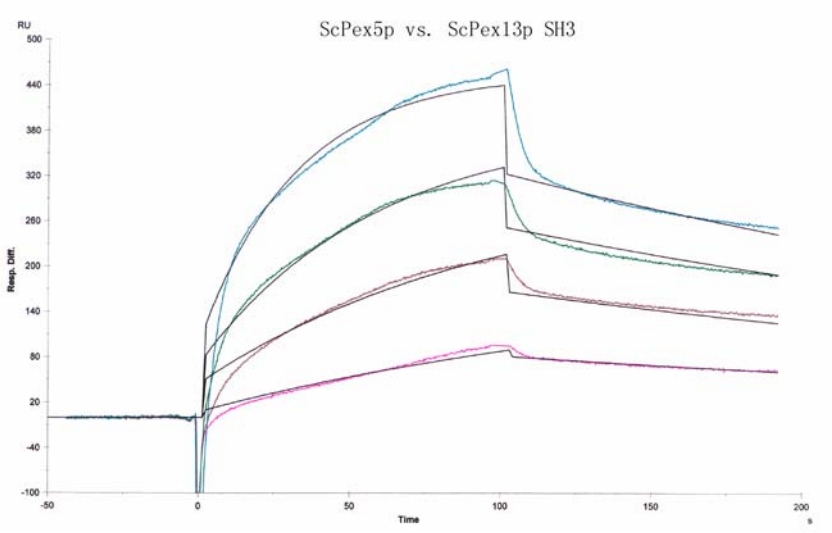
KA (1/M) 2.98×10^{7} KD (M) 3.35×10^{-8}

Fig. 47 BIAcore studies on Pex5p and Pex14p binding. The purified ScPex14p was immobilized on a CM5 chip and a series of the purified ScPex5p samples was introduced into the flow phase. The dissociation binding constants between the two proteins was determined as 33.5 pM.

3.12.2 Interaction between Pex5p and Pex13p SH3 Domain

Similar experiments as above for Pex5p and Pex13p SH3 were also carried out. For binding of the His-tagged ScPex5p to the GST tagged ScPex13p SH3, the purified ScPex13p SH3 domain was immobilized on a CM5 chip. It turned out that the dissociation binding constants between His-ScPex5p and ScPex13p SH3 domain was 55.3 μM.

The dissociation constant, determined upon the changes in the chemical shifts of the amide ^{1}H and ^{15}N atoms of a ^{15}N labeled SH3 domain construct upon addition of the ScPex5p peptide by NMR, were 640μM and 2mM for the long and short peptide, respectively.



ScPex5p Concentration (μM)	50	125.75	251.5	503
RU	178	260	307	337

KA (1/M) 1.8×10^{4} KD (M) 5.54×10^{-5}

Fig. 48 BIAcore studies on ScPex5p and ScPex13p binding. The purified ScPex13p SH3 was immobilized on a CM5 chip and a series of the purified ScPex5p samples was introduced into the flow phase. The dissociation binding constants between the two proteins was determined as 55.4 μM.

3.12.3 Interaction between Pex14p and Pex13p SH3 Domain

Changes in the chemical shifts of the amide groups of the ^{15}N labelled SH3 domain construct were followed by acquisition of several ^{15}N HSQC spectra upon titration with the Pex14p-derived peptides in different concentrations (SH3:ScPex5p-peptide=1:1, 1:2). These experiments yielded a dissociation constant (Kd) of 50 ± 10 μM for the complex Pex13p SH3/Pex14p

3.12 Detection of Concentration of Pex5p, Pex13p and Pex14p in vivo

The purified His-tagged ScPex5p, ScPex14p or GST tagged Pex13p SH3 protein were used as series of standard proteins to estimate the protein concentration of ScPex5p, ScPex13p and ScPex14p in vivo in yeast cells. Comparison of the signals derived by the immunological detection of defined amount of purified protein with the one of signals derived from the analysis of yeast whole cell lysate is expected to reveal the

endogenous concentration of the protein. The wild type yeast UTL7A cells, whose peroxisomes were induced in oleate medium, were detected for their Pex5p, Pex13p and Pex14p by the corresponding antibodies.

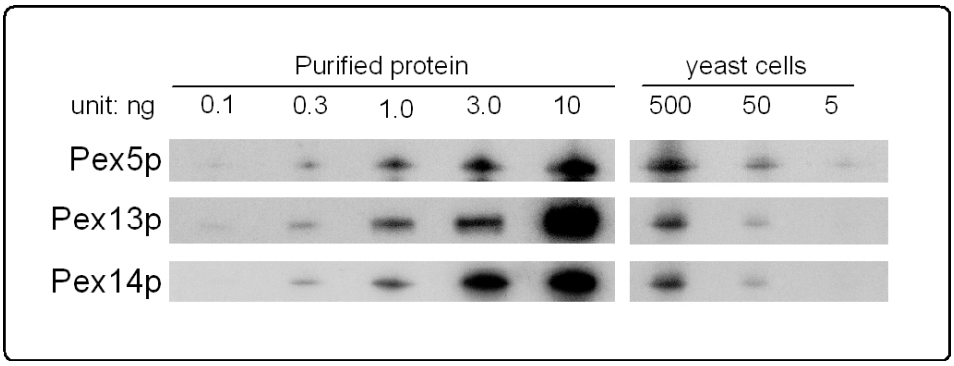


Fig. 49 Assessment of ScPe5p, ScPex14p and ScPex13p concentration in vivo. On the left, recombinant proteins in different amounts detected by immunoblot analysis; On the right, detection of the same protein in yeast cell lysate.

As shown in Fig. 49, the amount of Pex5p in 50 ng yeast cells was estimated to be 0.8 ng; whereas Pex13p in 50 ng cells was estimated approximately 0.2 ng and Pex14p in 50 ng cells to be 0.4 ng. It turned out that 1 g yeast cells contains 16 mg (0.23 µmol) Pex5p, 4 mg (0.09 µmol) Pex13p and 8 mg (0.21 µmol) Pex14p.