**Final report – July 2004- July 2008**

**Clp protease complexes in plastids of *Arabidopsis thaliana;* Functional analysis of ClpS1, CpS2, ClpC1 and a potential substrate modulator, ClpT**

Chloroplasts are essential organelles required for plant growth and biomass production. They synthesize many essential secondary metabolites (e.g. hormones, isoprenoids, amino acids, etc.) and house the photosynthetic apparatus needed for conversion of light energy and $CO_2$ into chemical energy [in the form of reduced carbohydrates, ATP and NADPH]. Thus chloroplasts are essential for life on earth and essential for production of bioenergy.

Formation and maintenance of a functional chloroplast requires an extensive investment in the biogenesis and homeostasis apparatus. Protease and proteolysis play a critical role in these processes, with the Clp gene family being particularly central. Proteolysis of proteins and protein complexes in plastids is poorly understood, and is not only critical for biogenesis, adaptation and maintenance but is also important for plant development. Several years ago, the vanWijk lab identified a large and relatively abundant ClpP/R/S complex, along with ClpC1,C2 and ClpD chaperones and a putative Clp affinity modulator in plastids. So far, no substrate recognition mechanism has been determined for any Clp complex in plants. The purpose of this grant was initiate functional analysis of three members of the Clp family.

***Change nomenclature of ClpS,T***

The van Wijk lab was the first to use the name ClpS [1], and we used the name ClpS1, ClpS2 for two chloroplast-specific proteins that we found associated with the 350 kDa ClpPR complex. Subsequently, a Clp substrate modulator was identified in E. coli and was named ClpS; a homologue is also present in plants. To avoid confusion, we renamed *Arabidopsis* ClpS1,S2,T into respectively ClpT1,T2,S, to be consistent with the nomenclature for *E. coli* ClpS; we will used these new names throughout this final report. In this grant, we thus proposed to address the role of ClpS1, ClpS2 and the potential substrate modulator, ClpT.

***Homology modeling, sub-cellular localization of ClpS and relative expression.***

Transient expression of a ClpS-GFP fusion protein in *Arabidopsis* leaves introduced by particle bombardment suggested that ClpS is located in chloroplasts (Patel and Rudella, not shown). Identification of ClpS by mass spectrometry of FPLC-fractionated chloroplast stroma further supported chloroplast localization of ClpS (Olinares, unpublished). Based on high resolution structures of E*. coli* ClpS in association with the isolated N-terminal domain of E. *coli* ClpA (Zeth et al, 2002), we created a model for *Arabidopsis* ClpS to investigate possible interactions with ClpC - unpublished by Daniel Ripoll (Fig. 1).

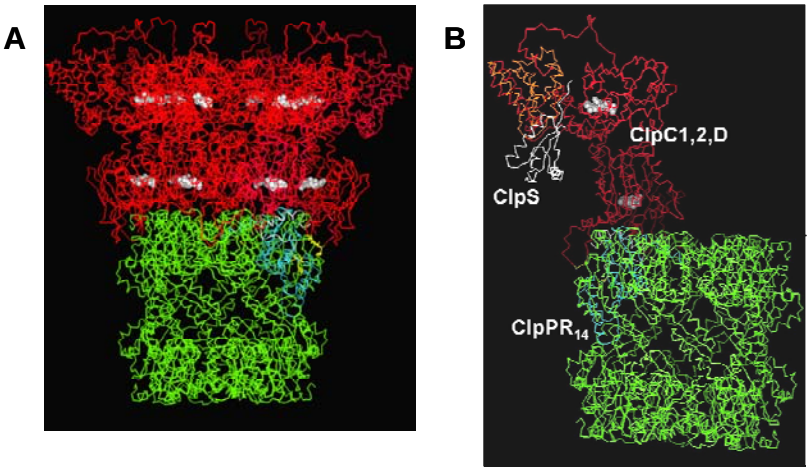



***Figure 1****. Model of ClpS. A. Ribbon structure model of ClpPR core (green), and representative subunit (cyan) and the ClpC hexamer (red). B. Ribbon model of the ClpPR core with a ClpC monomer (red), in which the predicted N-domain of ClpC is in orange, and the ClpS monomer is in white.*

Gene expression levels of *CLPS* and potential interacting

*CLPC1,D* chaperones, as well as plant specific *CLPT1,2* were examined after mining the public dataset from AtGenExpress (http://www.weigelworld.org/resources/microarray/AtGenExpress/). Across development, and using mean normalized expression, *CLPS* shows a similar expression pattern to *CLPC1*, but is less similar to *CLPT1/T2* (Fig. 2). *CLPD*, but not *CLPS* expression, is strongly induced in senescing leaves and late stage flowers. *CLPS* expression is highest at the 1st node stage, and at emergence of the first true leaves, suggesting a particularly high activity during leaf and/or chloroplast development. The co-expression pattern of *CLPS* and *CLPC* is consistent with a functional interaction.

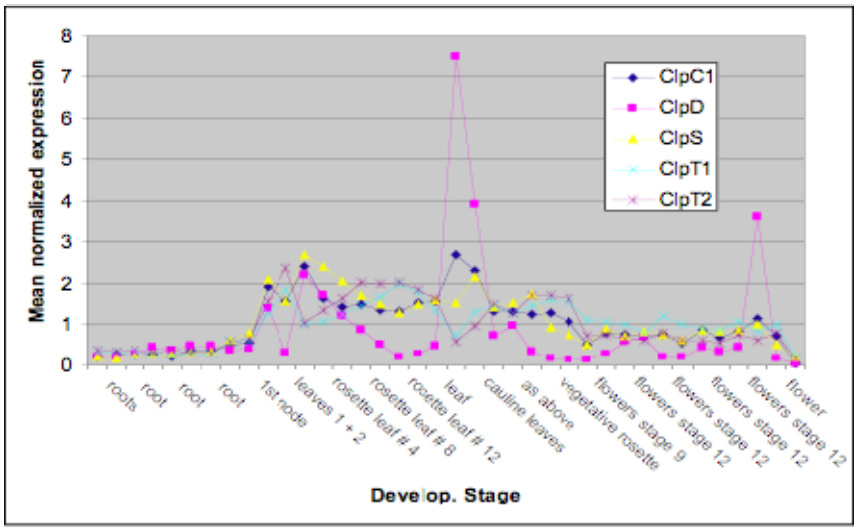



***Figure 2.*** *Normalized transcript accumulation of CLPS, CLPT1,2, and chaperones CLPC and CLPD, across development in various plant organs. Data were downloaded from http://www.weigelworld.org/resources/microarray/AtGenExpress/.*

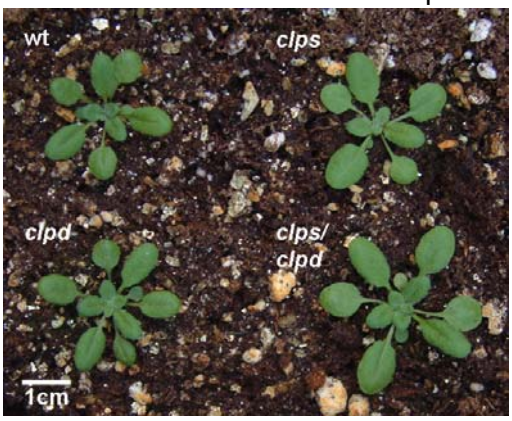
***Characterization of a CLPS null mutant: clps-1***

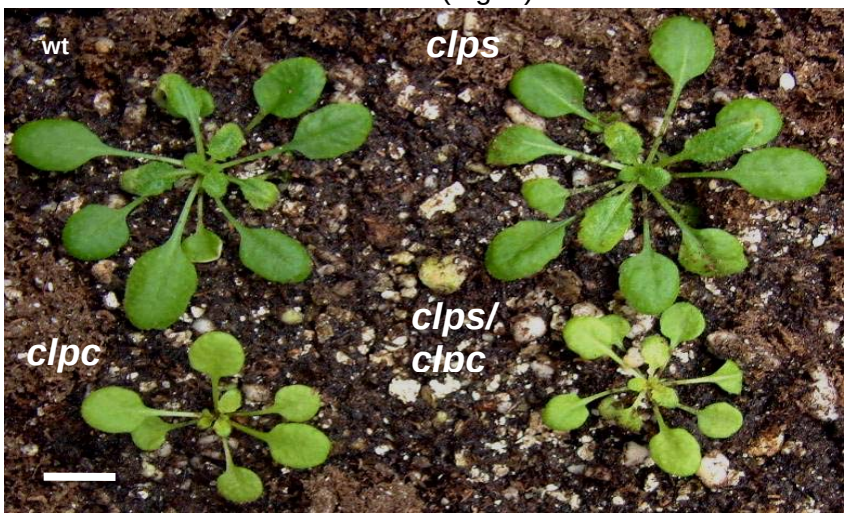
A null mutant for CLPS (At1g68860), *clps*, was isolated from the Gabi-KAT collection (Garlic_326b_G12) (Fig. 3). The T-DNA insertion is in the 3rd exon of the gene. This mutant was genotyped by PCR and DNA sequencing of the border of the T-DNA insertion, and RT-PCR showed no transcript accumulation. The results of these analyses indicate that the *clps* mutant line is a true null mutant for *CLPS* (Fig. 3). *clps* plants have no visible phenotype when grown under normal conditions (120 µmol photons. $m^{-2}$. $s^{-1}$, 22°C, 16h/8h photoperiod) on either ½ X MS plates or soil, nor under various different light regimes (not shown).

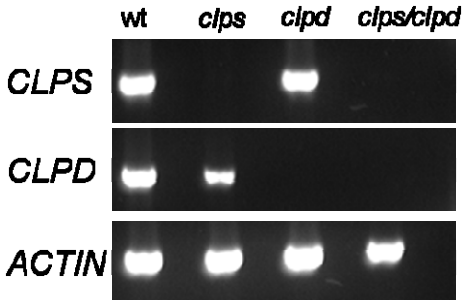
Considering the photoperiod specific "gain of function" phenotype in the *clps* null mutant, we sought to create 35S::*CLPS* lines to determine the effects of over-expression of *CLPS*. *CLPS* cDNA was cloned into the pEARLEYGATE 100 vector (Earley et al., 2006). This vector carries a 1X 35S promoter, no C-terminal tag and is compatible with Gateway cloning strategies. Putative transformants were isolated and genotyped. T1 plants will be screened for a visible phenotype in continuous light as compared to *clps* and wt plants. RT-PCR of *CLPS* will be needed to determine the level of transcripts of each 35S::*CLPS* line.

***Genetic Interaction of ClpS with other Clp Chaperones***

To investigate the potential functional (genetic) interactions of *CLPS* with *CLPC* and *CLPD*, double mutants were generated. The *CLPC* (At5g50920) mutant is from the SALK collection and is denoted *clpc1-1*. This mutant was previously characterized as pale, exhibiting delayed development (Sjögren et al, 2004) and showed reduced protein import rates to the chloroplast (Kovacheva et al, 2005). The function of ClpC as a HSP100 AAA+ chaperone to the ClpPR complex has not yet been studied, and how ClpC delivers substrates for degradation (with/without ClpS) is not well understood. *CLPD* was previously characterized as EARLY RESPONSE TO DEHYDRATION1 (ERD1) (Nakashima et al 1997), a "ClpC-like" gene induced during drought stress, salinity, and senescence (Weaver et al, 1999, Simpson et al, 2003). ClpD shares 53% amino acid sequence homology with ClpC and contains conserved AAA domains shared by ClpC (Weaver et al, 1999). The null mutant of *CLPD* (At5g51070) is denoted *clpd.* Neither the *clpd* and *clps* single mutants, nor the double mutants in *clpd/clps* show a phenotype in normal conditions (22$^{o}$C, 120 µmol photons. m$^{-2}$. s$^{-1}$) when grown under continuous light, or when grown under a photoperiod of 16h/8h light/dark (Fig. 3). RT-PCR analysis of the *clpd/clps* double mutant shows no transcript accumulation for either *CLPS* or *CLPD* (Fig. 3).



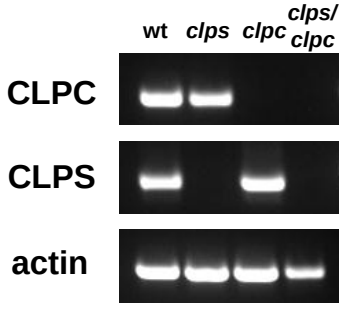




***Figure 3***. *ClpS, ClpD, ClpC single and double mutants. Plants were grown on soil 16H light, 8H dark and a light intensity of ~120 µmol photons. m$^{-2}$. s$^{-1}$. RT-PCR reactions show that all mutants are nulls.*

When *clps* was crossed to *clpc1-1*, double mutants in *clps/clpc* displayed a *clpc*-like phenotype with pale leaves and slower development than wt (Fig. 3) when screened on either ½

X MS media, ½ X MS media supplemented with 1% sucrose, or on soil. RT-PCR analysis of the *clps/clpc* double mutant shows no transcript accumulation for either *CLPS* or *CLPC* (Fig. 3). Therefore, loss of ClpS does not enhance or alter the *clpc1* phenotype. This suggests a level of redundancy in the Clp chaperone system. Perhaps in the absence of ClpC and ClpS, ClpC2 and/or ClpD compensates for loss of ClpC, both as an import chaperone and as a substrate-loading component of the ClpPR core. Therefore, the lower endogenous levels of ClpC2 could be sufficient to compensate for import, and either allow direct substrate binding (in the absence of both ClpC1 and ClpS) to the ClpPR core, or result in less efficient proteolysis resulting in accumulation of proteins that need to be degraded (due to less copies of ClpC2 and therefore less efficient unfolding into the core). Neither scenario's apparently result in a visual phenotype.

***ClpS Antibody and Co-IP***

To determine ClpS accumulation levels in wt and *clp* mutants in different developmental stages and organs, as well as to determine proteins interacting with ClpS, we generated an antibody for Western blot and Co-IP analyses. A mature version of the ClpS protein (matClpS) was over-expressed in *E. coli* for antibody production. In matClpS, the predicted cTP sequence of ClpS, in addition to sequence that is not homologous to *E. coli* ClpS (amino acids 1-45) were omitted. matClpS was cloned into a pGEX5X-1 vector containing an N-terminal GST fusion protein and a Factor $X_a$ cleavage site. Two rounds of protein over-expression were performed and 2 aliquots of ~1.2 and ~1.5 µg of protein were submitted to Cocalico Biologicals (http://www.cocalicobiologicals.com/about.html) for antibody production in 2 individual rabbits.

After receiving final bleeds, the crude sera from both of these rabbits were tested. In both wt total leaf protein, and purified wt stromal protein, a band of expected size at ~17 kDa is detectable in both bleeds, whereas in the *clps* null mutant, no band is detected. An additional band is detected in bleed 303 at 20 kDa that is also not present in the *clps* mutant. Final bleeds were further purified, using the original GST-ClpS fusion protein as bait. After purification, a faint band at the expected size of 17 kDa is detectable in 200 µg purified stroma, but not in total leaf protein for either wt or *clp*s (not shown).

***Investigation of protein interactors and potential substrates of ClpS***

To investigate potential interacting proteins and substrates of ClpS, the pGEX over-expression construct was used as a bait in a preliminary on-column pull-down experiment. After ClpS-GST was overexpressed and immobilized on a GST-column, total stroma from freshly isolated wt chloroplasts of *A. thaliana* was run over the column at a 0.2 ml/min flow rate to maximize binding. After washing with buffer, the ClpS-GST was eluted in 50mM Tris-pH=8.0 and 10mM reduced glutathione. Proteins were collected from 3 independent biological replicates (3 independent ClpS-GST overexpression and 3 independent chloroplast purifications) of this experiment and run on a 12% acrylamide SDS-PAGE (Fig. 4). Loading was based on total volume, and not normalized to total protein for each lane. Lane 1 stroma FT corresponds to the Lane 4 eluate, Lane 2 FT matches Lane 5 eluate and Lane 3 FT corresponds to Lane 6 eluate.

There are 4 possible "subsets" of identified proteins that could result from an experiment of this nature. First, proteins in the eluate could occur due to non-specific binding to the GST column. To control for this, we ran total stroma over the column in the absence of purified ClpS-GST. No proteins could be observed in this lane (not shown). Second, proteins could bind non-specifically to the GST-portion of the fusion protein. Lastly, proteins identified here could represent non-specific binding to ClpS, possibly associating with potential aggregates, or the ClpS specificity could be directly through binding to either substrates or indirectly via ClpC and subsequently to the ClpP/R core. To identify proteins in the ClpS eluate, Lane 5 was cut out and processed for identification by reverse phase nanoLC-ESI-LTQ-Orbitrap. Overall, MS/MS analysis revealed 337 unique proteins identified in the ClpS-GST eluate.

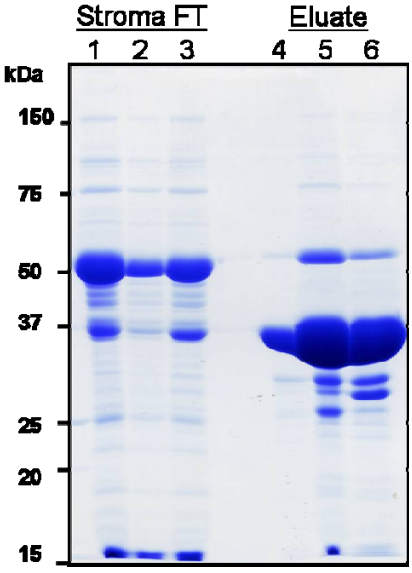



***Figure 4****. ClpS-GST pulldown experiments. Lane 1-3: unbound stroma flow-through from the column from 3 independent biological replicates. Lane 4-6: ClpS-GST eluted with other interacting proteins from 3 independent biological replicates. The large band at ~37 kDa in lanes 4-6 represents ClpS-GST. Only Lane 5 was processed for identification by mass spectrometry.*

For more confidence in the dataset, proteins with less than 2 unique peptides were removed. All of the proteolytic (P1, P3-P6) and non-proteolytic (R1-R4) subunits of the Clp complex were identified in the pulldown. ClpC and ClpT were also identified. Relative protein abundance was calculated based on unique queries (*i.e*. # of times a peptide is identified and matched to a protein with good score), and normalized to the total number of queries in the experiment. These unique queries were compared to the abundance (normalized unique queries) of proteins in a typical wt stroma preparation. 64 were found to be at least two-fold enriched in the ClpS pull-down experiment, compared abundance in intact chloroplasts of wt plants. These 64 proteins are a potential substrate pool. Control experiments where wt stroma is run over a bound GST protein without ClpS will be carried out to determine the significance of the observed enrichment.

***GFP-based degradation reporters to explore degradation signals in wt and mutants***

It is well established *in vitro* that ssrA-tagged GFP can be easily degraded by the purified *E. coli* Clp system, and the loss of fluorescence measured (Farrell et al, 2005 and Erbse et al, 2006). To this end, we have begun to develop a GFP reporter system in *A. thaliana* protoplasts to test whether ssrA-tagged degradation occurs in chloroplasts. We modified a GFP construct that was already shown to target GFP to *Arabidopsis* chloroplasts (Abdel-Ghany et al 2005). Both active (AANDENYALAA) and inactive (AANDENYALDD) tags were added to the C-terminus of GFP, and these constructs were transiently expressed in wt mesophyll protoplasts. After verifying transfection and GFP expression were successful, cycloheximide (15uM) was added to block cytosolic translation and measure the decrease of GFP after 45 minutes. After adding cycloheximide, GFP fluorescence was measured. We determined GFP accumulation levels by quantification with the GFP antibody, and subcellular localization by confocal microscopy. We will seek new funding to continue this line of research. If the GFP-ssrA variants show differential accumulation patterns we will then explore if these levels are different in GFP transformants of *clps, clpc x clps* and/or *clpd x clps* null mutants.

***Search for T-DNA tagged mutants in ClpT1 and ClpT2.***

We extensively screened all available T-DNA collection for ClpT1 and ClpT2 mutants. We ordered several putative lines but after extensive analysis we did not find any homozygous null mutants. Mutants with reduced expression did not show any phenotype (data not shown). Both ClpT genes are very small and therefore the change to find suitable insertion mutants is much reduced as compared to mutants for large genes. We therefore initiated the generation of RNAi suppression mutanst targeting each ClpT gene alone or both ClpT genes at the same time. We did generate the RNAi constructs and we will seek new funding to continue these studies.

***Detection of ClpT1,ClpT2 and ClpS by mass spectrometry and information about their oligomeric state***

Whereas, we identified ClpT1 and ClpT2 by mass spectrometry in association with the ClpPR complex, ClpS was more elusive, most likely due it its small size (<15 kDa). However, after FPLC-based fractionation on non-denatured stroma from Arabidopsis chloroplasts, we did detect ClpS, thus confirming its expression and chloroplast localization. In the same set of experiments we also identified ClpT1,T2 together with the complex ClpPR as expected, but also in a ~120 kDa complex without the presence of ClpPR proteins. In agreement with our homology modeling, this suggest that ClpT1,T2 could form homo- or heterohexamaric complexes. This may be key to understanding the function of these unique Clp members.

**Personnel and training**

Two undergraduate students, 2 graduate students and two postdoctoral fellows have been involved in the project, and some were partially supported by this DOE grant.

*Undergraduate students* Kieren Patel was involved in the genetics analysis of ClpS and ClpT1,2. After completing an honors thesis about the Clp family, he worked for nearly a year in my lab on this project. In 2007, he left my lab to start a PhD at UC Berkeley. Cornell undergraduate Brian Connolly worked under supervision of postdoc Verenice Ramirez on the genetics on several clp mutants. Brian now completed his undergraduate education and is now in a post-graduate training program at NIH. Both undergraduates contributed to this project, and received training in molecular genetics and biochemistry in addition to general lab experience. This facilitated them to find attractive training programs at reputed institutions.

*Graduate Students* Andrea Rudella and Heidi Rutschow both worked in this project. Both have now completed their PhDs. Dr Rudella moved to Europe to become a mass spectrometry expert working for a large analytical company (Waters, Inc). Heidi started a postdoc at the University of Amherst, MA. Both students received extensive multi-disciplinary training that benefited them greatly to find positions of their choice.

*Post-doctoral fellows* Dr Jitea Kim and Dr Verenice Ramirez-Rodriguez both were involved in several aspects of this projects, even if they were not or only partially funded by this DOE grant. Dr Kim is still in my lab continuing on the analysis of the Clp complex, while Verenice found a research position in a national lab in Mexico, her country of origin.

**Dissemination of results**

During the funding period, van Wijk has given more than ten seminars at various institutes and conferences in which I presented published and/or preliminary data on our Clp project. In addition, postdoc Verenice Ramirez presented a poster at the International Arabidopsis meeting in Wisconson in the summer of 2005 about her Clp work, whereas GRA Andrea Rudella presented a poster on the ClpP/R/S complex at the American Society for Mass spectrometry Meeting in Seatle.

**Publications**

So far we published two papers [2, 3] and a third paper was submitted last month [4]. Two additional manuscripts are in preparation and both will be submitted within the next 2 months (to Molecular and Cellular Proteomics and to Plant Cell) [5, 6]. The DOE will be acknowledged for their support in additional future publications regarding ClpT and ClpS.

**Relevant publications and manuscripts by the van Wijk lab**

1. Peltier, J.B., J. Ytterberg, D.A. Liberles, P. Roepstorff, and K.J. van Wijk, Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of Arabidopsis thaliana. J Biol Chem, 2001. **276**(19): p. 16318-27.

2. Rudella, A., G. Friso, J.M. Alonso, J.R. Ecker, and K.J. van Wijk, Downregulation of ClpR2 Leads to Reduced Accumulation of the ClpPRS Protease Complex and Defects in Chloroplast Biogenesis in Arabidopsis. Plant Cell, 2006. **18**(7): p. 1704-21.
3. Adam, Z., A. Rudella, and K.J. van Wijk, Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. Curr Opin Plant Biol, 2006. **9**(3): p. 234-40.
4. Kim, J., Olinares, P.D.B., Rudella, A.,Ramirez Rodriguez, V., Zybailov, B., van Wijk, K.J., Plastid localized ClpR4 is essential for seedling viability and plays a unique role in Clp protease function. submitted, 2009.
5. Zybailov, B., Friso, G., Rudella, A., Kim, J., Ramirez Rodriguez, V., Sun, Q.,van Wijk, K.J., Large scale proteomics of a chloroplast ClpR2 protease mutant indicates folding stress, destabilization of protein homeostasis and feedback regulation on metabolism. in preparation, 2009.
6. Kim, J., Ramirez-Rodriguez, V., Zybailov, B., Rudella, A., van Wijk, K.J., ClpPR subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis and plant development. in preparation, 2009.