UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
DEPARTMENT OF BIOLOGY AND ECOLOGY

INTRINSICALLY DISORDERED PROTEINS OF THE EUROPEAN CORN BORER *OSTRINIA NUBILALIS* (HBN, 1796)

Doctoral dissertation

Advisor:
Prof. Željko D. Popović, PhD

Candidate:
Miloš Avramov, MSc

Novi Sad, 2023

UNIVERSITY OF NOVI SAD **FORM – 5a**
FACULTY OF SCIENCES

KEY WORD DOCUMENTATION[1]

Document type:	Doctoral dissertation
Author:	Miloš Avramov, MSc
Supervisor (title, first name, last name, position, institution)	Prof. Željko D. Popović, PhD, Faculty of Sciences
Thesis title:	Intrinsically disordered proteins of the European corn borer *Ostrinia nubilalis* (Hbn, 1796)
Language of text (script):	English language, latin script
Physical description:	Number of: Pages – 166 Chapters – 9 References – 240 Tables – 28 Illustrations – 6 Graphs – 22 Appendices – 1
Scientific field:	Biology
Scientific subfield (scientific discipline):	Biochemistry
Subject, Key words:	intrinsically disordered proteins; intrinsically disordered protein regions; Os*trinia nubilalis*; cold hardiness; diapause; gene expression
Abstract in English language:	Intrinsically disordered proteins (IDPs) are a specific and diverse group of proteins that, under physiological environmental conditions, do not possess higher levels of structural organization, but are found in a form resembling denatured proteins. When IDPs are exposed to conditions that are denaturing for ordered proteins, such as high and low temperatures or strongly acidic or basic pH of environment, they acquire transient forms of ordered structure. Also, a large number of typically ordered, globular proteins have intrinsically disordered regions (IDRs) in their structure. Contrary to the established view that the function of proteins depends on their ordered structure, research has shown that IDPs possess various molecular functions regardless of the absence of structure. Due to the lack of higher orders of structure, IDPs have the ability to recognize a large number of different biomolecules and interact with them, and during these interactions they partially fold and acquire a more ordered structure. Due to their flexibility and plasticity in intermolecular interactions, IDPs and proteins with intrinsically disordered regions are often involved in cell signaling processes, regulation of transcription and translation, and also function as molecular chaperones. Based on their diverse roles, as well as resistance to denaturation under unfavorable conditions, IDPs probably have a pronounced biological significance in adaptive processes of organisms to abiotic stress factors. In order to

[1] The author of the doctoral dissertation has signed the following Statements:
5б – Statement on the authorship,
5в – Statement that the printed and e-version of the doctoral dissertation are identical and about personal data,
5г – Statement on copyright licenses.
The paper and e-versions of the Statements are held at he faculty and are not included into the printed thesis.

	evaluate their potential importance in the process of adaptation to hypometabolic living conditions during phases of arrested development and developing cold resistance, in this dissertation 5th instar larvae of the European corn borer *Ostrinia nubilalis* were used. In order to survive the cold winter months this species enters a specific resting state called diapause. During diapause, corn borer larvae undergo biochemical and molecular changes that allow them to survive only on stored energy sources and to develop resistance to extremely low winter temperatures. For this reason, the European corn borer is often used as an *in vivo* model system for studying these processes. In this dissertation, a method was established and optimized for enriching the content of IDPs, validation of the enrichment procedure with specific 2D polyacrylamide gel electrophoresis and isolation of disordered proteins from whole-body homogenates of larvae that were subjected to different temperature treatments to develop low temperature resistance. Isolated proteins were identified using liquid chromatography coupled with tandem mass spectrometry. Also, the degree of intrinsic disorder of the identified proteins was determined by *in silico* analyses, quantitative and qualitative parameters of their disorder were evaluated, and the functional characterization of the proteins was performed. The relative expression of selected genes for proteins with a significant degree of intrinsic disorder was also examined in 5th instar larvae that were reared under non-diapausing conditions, as well as diapause-inducing conditions that also lead to the development of cold resistance. The results showed that in this species there is a significant content of IDPs and proteins with disordered regions, and that the proportion of proteins with IDRs is correlated with the acquisition of cold resistance. The expression of most of the analyzed genes is also correlated with this adaptation. The results of this dissertation provide new insights into the molecular basis of processes that enable the European corn borer to adapt to unfavorable living conditions characterized by a reduced metabolic rate, limited energy sources for basal metabolism, as well as the negative effects of exposure to low temperatures on the functional and structural homeostasis of proteins and processes they are involved in.
Accepted by the Scientific Board on:	30th March 2023
Defended on: (Filled by the faculty service)	
Thesis Defense Board: (title, first name, last name, position, institution)	President: Prof. Edward Petri, PhD, Faculty of Sciences Member: dr Ágnes Tantos, Senior Research Fellow, Research Centre for Natural Sciences, “Eötvös Loránd” Research Network, Budapest, Hungary Member (advisor): Prof. Željko D. Popović, PhD, Faculty of Sciences
Note:	

УНИВЕРЗИТЕТ У НОВОМ САДУ **ОБРАЗАЦ – 5а**
ПРИРОДНО-МАТЕМАТИЧКИ ФАКУЛТЕТ

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА[2]

Врста рада:	Докторска дисертација
Име и презиме аутора:	MSc Милош Аврамов
Ментор (титула, име, презиме, звање, институција)	др Жељко Д. Поповић, ванредни професор, Природно-математички факултет
Наслов рада:	Структурно-неуређени протеини кукурузног пламенца *Ostrinia nubilalis* (Hbn, 1796)
Језик публикације (писмо):	Енглески језик, латиница
Физички опис рада:	Унети број: Страница – 166 Поглавља – 9 Референци – 240 Табела – 28 Слика – 6 Графикона – 22 Прилога – 1
Научна област:	Биологија
Ужа научна област (научна дисциплина):	Биохемија
Кључне речи / предметна одредница:	структурно-неуређени протеини; структурно-неуређени региони протеина; *Ostrinia nubilalis*; отпорност на хладноћу; дијапауза; експресија гена
Резиме на српском језику:	Структурно-неуређени протеини (*енгл.* intrinsically disordered proteins, IDP-ови) су специфична и разноврсна група протеина који при физиолошким условима средине не поседују више нивое структурне организације, већ се налазе у облику који наликује денатурисаним протеинима. Када су IDP-ови изложени условима који су денатуришући за уређене протеине, као што су високе и ниске температуре или изразито кисела или базна pH средине, они задобијају пролазне облике уређене структуре. Такође, велик број типично уређених, глобуларних протеина у својој структури поседују регионе који могу бити структурно-неуређени (*енгл.* intrinsically disordered regions, IDR-ови). У супротности са устаљеним виђењем да функција протеина зависи од њихове уређене структуре, истраживањима је показано да IDP-ови поседују разноразне молекуларне функције без обзира на одсуство структуре. Услед одсуства виших нивоа структуре, IDP-ови имају способност препознавања великог броја других биомолекула и интераговања са њима, а приликом тих интеракција се делимично савијају и задобијају пролазу уређену структуру. Због своје флексибилности и пластичности у међумолекулским интеракцијама, IDP-ови и протеини са структурно-неуређеним регионима често су укључени у процесе ћелијског сигналинга, регулацију транскрипције и транслације, а поседују и

[2] Аутор докторске дисертације потписао је и приложио следеће Обрасце:
5б – Изјава о ауторству;
5в – Изјава о истоветности штампане и електронске верзије и о личним подацима;
5г – Изјава о коришћењу.
Ове Изјаве се чувају на факултету у штампаном и електронском облику и не кориче се са тезом.

	функције молекуларних шаперона. На основу њихових разноликих улога, као и отпорности на денатурацију у неповољним условима, IDP-ови вероватно имају изражен биолошки значај у процесима адаптација организама на абиотске факторе стреса животне средине. Како би се проценио њихов претпостављени значај у процесу адаптације организма на хипометаболичке услове живота током фазе мировања и развијања отпорности на хладноћу, у овој дисертацији коришћене су гусенице 5. инстара врсте кукурузног пламенца *Ostrinia nubilalis*. За ову врсту је карактеристично да у зимском периоду улази у фазу мировања звану дијапауза, како би преживела хладне зимске месеце. Током дијапаузе, у гусеницама кукурузног пламенца долази до промена на биохемијском и молекуларном нивоу које им омогућавају да преживе само на ускладиштеним изворима енергије као и да развију отпорност на изузетно ниске зимске температуре. Због тога се кукурузни пламенац често користи као *in vivo* модел-систем у истраживањима ових процеса. У овој дисертацији успостављена је и оптимизована метода за обогаћивање садржаја IDP-ова, валидацију поступка обогаћивања специфичном 2Д електрофорезом на полиакриламидном гелу и изолацију неуређених протеина из хомогената целих гусеница које су подвргаване различитим температурним третманима за развој отпорности на ниске температуре. Изоловани протеини су идентификовани помоћу течне хроматографије купловане са тандемском масеном спектрометријом. Такође, *in silico* анализама су утврђени степен структурне неуређености идентификованих протеина, процењени квантитативни и квалитативни параметри њихове неуређености и урађена је функционална карактеризација протеина. Испитана је и релативна експресија одабраних гена за протеине са утврђеним значајним степеном структурне неуређености у гусеницама 5. инстара држаним у условима недијапаузе, као и дијапаузе уз развијање отпорности на хладноћу. Резултати су показали да у овој врсти постоји значајан удео IDP-ова и протеина са неуређеним регионима, као и да је удео протеина са IDR-овима корелисан са стицањем отпорности на хладноћу. Експресија већине анализираних гена је такође корелирана са том адаптацијом. Резултати ове дисертације дају нове увиде у молекуларне основе процеса који омогућавају врсти кукурузног пламенца да се прилагоди неповољним условима живота које карактеришу смањена стопа метаболизма, ограничени извори енергије за потребе базалног метаболизма, као негативне ефекте излагања ниским температурама на структурну и функционалну хомеостазу протеина и процеса у којима су они укључени.
Датум прихватања теме од стране надлежног већа:	30. март 2023. године
Датум одбране: (Попуњава одговарајућа служба)	
Чланови комисије: (титула, име, презиме, звање, институција)	Председник: др Едвард Петри, ванредни професор, Природно-математички факултет Члан: др Агнеш Тонтош, виши научни сарадник, Истраживачки центар за природне науке, Истраживачка мрежа "Етвеш Лоранд", Будимпешта, Мађарска Члан (ментор): др Жељко Д. Поповић, ванредни професор, Природно-математички факултет
Напомена:	

"New kids on the block…"

In memoriam – ***Miloš Avramov (1943 – 2021)***

Acknowledgements

*Having reached this point, people usually ask themselves, "Where do I begin?" That is not the case here. This dissertation is dedicated to my grandparents, **Miloš** and **Nada**, whom I owe the biggest gratitude for becoming the person who I am today. They have been my greatest source of inspiration, as well as my biggest supporters, throughout my entire life. Nothing brought me more joy than seeing them be proud of my accomplishments, as I then knew that their trust and faith in me was repaid. Baba i deda, hvala vam.*

*A special mention goes to my high school biology teacher, **Biserka Nenadović**, who, in a school dedicated to the study of languages and social sciences, made sure that we were prepared for the world of natural sciences, as well. Thank you, **Biso**, for setting me on this path.*

*I must also express my gratitude to my advisor, **Prof. Željko D. Popović**, for guiding me not only throughout my doctoral studies, but also during my undergraduate and master studies, ever since I met him as a 2nd year student in the practical course for „Fundamentals of molecular biology", more than a decade ago. Through his own example, sage advice and constructive criticism, he always pushed me to better myself. How much I owe my grandparents for the person I am today, that much I owe **Željko** for where I am as a researcher and man of science.*

*I would also like to thank **dr Ágnes Tantos** for being an integral part in defining the research scope of this dissertation, as well as providing me with the means to carry out my experimental designs, both material as well as educational. Being hosted in her lab group at the Research Centre for Natural Sciences in Budapest as part of a joint project is one of the highlights of my life and it is a period that I will always fondly reminisce on. If only that year did not fly by as quickly as it did.*

*In the same regard I would like to thank **dr Éva Schád** for lending me her bioinformatics expertise in matters of in silico analyses performed in the dissertation, and **dr Ágnes Révész** and **dr Lilla Turiák** for their assistance with the identification of the incredible number of proteins in this work.*

*To **Prof. Edward Petri** I owe gratitude first for introducing me to the field of structural biology and teaching me the skills which were indispensable during my PhD studies, and, second, for his*

assistance with the preparation of this body of work with his helpful comments, critique and advice on how to best improve the manuscript.

I would also like to thank all my colleagues from the ***Laboratory for biochemistry and molecular biology*** *for the many wonderful years we have spent together – from the long days in the lab to our outings when we put the work behind us, talk about anything and everything, and enjoy each other's company.*

When I doubted myself the most, and would worry about every single issue and obstacle that came my way, one person was by me to reassure me and help me push forward at all times. Thank you, ***Teodora****, for your love, support and for always finding the right words and making me believe in myself, time after time.*

Finally, I want to thank my family – my mother ***Nada****, father* ***Srđan*** *and brothers* ***Nestor (Avram)*** *and* ***Mihailo (also Avram)*** *who, along with my closest friends, supported me in countless ways throughout my entire formal education which has come to an end after 24 long years. Again, thank you all for everything.*

Miloš (Prime Avram)

List of abbreviations

A

ACB – Asian corn borer

ACN – acetonitrile

ANOVA – analysis of variance

APS – ammonium persulfate

ATP – adenosine triphosphate

C

CID – collision-induced dissociation

cryo-EM – cryogenic electron microscopy

Ct – cycle threshold

D

D(15) – diapausing larvae acclimated to 15°C for two weeks

D(5) – diapausing larvae acclimated to 5°C for two weeks

D(–3) – diapausing larvae acclimated to –3°C for two weeks

D(–16) – diapausing larvae acclimated to –16°C for two weeks

Dca – cold-acclimated diapausing group

DEPC – diethyl pyrocarbonate

Dnca – non-cold-acclimated diapausing group

dNTP – deoxynucleoside triphosphate

DTE – dithioerythreitol

DTT – dithiothreitol

E

ECB – European corn borer

EDTA – ethylenediaminetetraacetic acid

ELMs – eukaryotic linear motifs

ERM – ezrin, radixin, moesin

ETC – electron transport chain

G

Goi – gene of interest

H

HREs – heat shock response elements

hsc70 – heat shock cognate 70 kDa protein gene

HSFs – heat shock factors

HSP – heat shock protein

hsp20.1 – small heat shock protein 20.1 gene

hsp20.4 – small heat shock protein 20.4 gene

hsp70 – heat shock protein 68-like gene

hsp90 – heat shock protein 90 gene

I

ID – intrinsic disorder

IDPs – intrinsically disordered proteins

IDRs – intrinsically disordered regions

L

LC-MS/MS – liquid chromatography tandem mass spectrometry

M

MDPs – mostly disordered proteins

moe – moesin/ezrin/radixin homolog 1 gene

MoRFs – molecular recognition features

N

NAD^+ – oxidized nicotinamide adenine dinucleotide

$NADP^+$ – oxidized nicotinamide adenine dinucleotide phosphate

ND – non-diapausing group

NH_4CO_3 – ammonium carbonate

NH_5CO_3 – ammonium bicarbonate

NOPs – nearly ordered proteins

O

OD – optical density

OPs – ordered proteins

P

PAGE – polyacrylamide gel electrophoresis

PCA – perchloric acid

PDPs – partially disordered proteins

pH – potential of hydrogen

PPII – polyproline type II

PTM – post-translational modification

Q

qPCR – quantitative polymerase chain reaction

R

Ref – reference gene

RIN – RNA integrity number

rps3 – ribosomal protein s03 gene

RT-PCR – reverse transcription polymerase chain reaction

S

sHSPs – small heat shock proteins

SLiMs – short linear motifs

T

TAE – tris-acetate-EDTA

TBE – tris-borate-EDTA

TCA – trichloracetic acid

TCA cycle – tricarboxylic acid cycle

TEMED – tetramethylethylenediamine

TFA – trifluoroacetic acid

thym – thymosin beta gene

tnt – troponin T gene

tropmy2 – tropomyosin-2 gene

Amino acid codes:

amino acid	three-letter	one-letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table of figures

Figure 1. Life stages of *O. nubilalis* 14

Figure 2. Life cycle of bivoltine *O. nubilalis* specimens. 15

Figure 3. Experimental setup and cold acclimation regiment in the Pilot. 24

Figure 4. Setup and cold acclimation regiment in the Main Experiment. 26

Figure 5. Results of in-house 2D PAGE for detecting intrinsically disordered proteins. 42

Figure 6. Total unique and common proteins isolated from the (A) Pilot and (B) Main experimental setups 43

Figure 7. Effect of sample heating on the number of identified proteins in the (A) Pilot and (B) Main experimental setups 45

Figure 8. Total number of proteins with varying degrees of intrinsic disorder in the Pilot (A) and Main (B) experimental setups 47

Figure 9. Percental distribution of proteins with varying degrees of intrinsic disorder in the Pilot (A) and Main (B) experimental setups 48

Figure 10. Effect of heat treatment on the percental distribution of proteins with varying degrees of intrinsic disorder in the two sample types of Pilot setup experimental groups 50

Figure 11. Effect of heat treatment on the percental distribution of proteins with varying degrees of intrinsic disorder in the two sample types of Main setup experimental groups 52

Figure 12. Distribution of long intrinsically disordered regions in proteins with various degrees of intrinsic disorder in the Pilot setup 54

Figure 13. Distribution of long intrinsically disordered regions in proteins with various degrees of intrinsic disorder in the Main experimental setup 56

Figure 14. Ratios of disorder-promoting and order-promoting amino acids in MDPs, PDPs, NOPs and OPs 58

Figure 15. Ratios of individual disorder-(A) and order-promoting amino acids (B) in MDPs, PDPs, NOPs and OPs 59

Figure 16. Biological processes and molecular functions of intrinsically disordered proteins and proteins containing long IDRs 60

Figure 17. Biological processes and molecular functions of intrinsically disordered proteins and proteins containing long IDRs in the Other category 61

Figure 18. Ct values of reference genes – *rps3* and *actin* 62
Figure 19. Relative expression of *hsp90* gene 63
Figure 20. Relative expression of *hsc70* gene.. 65
Figure 21. Relative expression of *hsp70* gene. 67
Figure 22. Relative expression of *hsp20.4* gene 69
Figure 23. Relative expression of *hsp20.1* gene 71
Figure 24. Relative expression of *tropmy2* gene.. 73
Figure 25. Relative expression of *tnt* gene.. 75
Figure 26. Relative expression of *thym* gene 77
Figure 27. Relative expression of *moe* gene 79
Figure 28. Proposed mechanism of proteomic changes induced by the decrease in ambient temperature and progression of a hypometabolic resting state over time.................................. 108

List of tables

Table 1. Proteinogenic disorder- and order-promoting amino acids 8
Table 2. Scientific classification of the European corn borer, *O. nubilalis* 12
Table 3. Composition of insect feed used for maintaining diapausing and rearing non-diapausing larvae 25
Table 4. Composition of gels prepared for first dimension PAGE separation 28
Table 5. Composition of second dimension PAGE separating gel 28
Table 6. Composition of individual reaction mixtures for RT-PCR cDNA synthesis 35
Table 7. Thermal cycler settings for RT-PCR cDNA synthesis 35
Table 8. Composition of individual reaction mixtures for primer specificity analysis 37
Table 9. Thermal cycler settings for primer specificity analysis 37
Table 10. Thermal cycler settings for primer efficiency determination 38
Table 11. Total, unique and shared proteins identified in groups of Pilot and Main experimental setups 44
Table 12. Total numbers of unique and shared proteins identified in different sample types of Pilot and Main experimental setup groups 46
Table 13. Total numbers of proteins with varying degrees of intrinsic disorder identified in different experimental groups of Pilot and Main setups 49
Table 14. Total numbers of proteins with varying degrees of intrinsic disorder identified in the two sample types of Pilot setup experimental groups 51
Table 15. Total numbers of proteins with varying degrees of intrinsic disorder identified in the two sample types of Main setup experimental groups 53
Table 16. Long IDR content of proteins with various degrees of intrinsic disorder in the experimental groups of Pilot setup 55
Table 17. Long IDR content of proteins with various degrees of intrinsic disorder in the experimental groups of Main setup 57
Table 18. Relative expression of *hsp90* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 64
Table 19. Relative expression of *hsc70* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 66

Table 20. Relative expression of *hsp70* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 68
Table 21. Relative expression of *hsp20.4* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 70
Table 22. Relative expression of *hsp20.1* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 72
Table 23. Relative expression of *tropmy2* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 74
Table 24. Relative expression of *tnt* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 76
Table 25. Relative expression of *thym* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 78
Table 26. Relative expression of *moe* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 80
Table 27. Primer pairs used for qPCR analysis. 140
Table 28. Results of gene expression analysis in *O. nubilalis* larvae 141

Table of contents

Abstract........ 1

Извод........ 3

1. Introduction........ 5

1.1. Intrinsically disordered proteins – IDPs........ 6

1.1.1. The code for protein intrinsic disorder 7

1.1.2. Localized intrinsic disorder – intrinsically disordered regions 8

1.1.3. Environmental hardiness of IDPs........ 9

1.1.4. Natural abundance of intrinsic disorder........ 10

1.1.5. Intrinsic disorder and protein function 11

1.2. The European corn borer, *Ostrinia nubilalis* (Hbn., 1796)........ 12

1.2.1. History, distribution and ecology 12

1.2.2. Insect morphology........ 13

1.2.3. Life cycle of *Ostrinia nubilalis* 14

1.2.4. *Ostrinia nubilalis* as a model system for diapause and cold adaptation........ 16

1.2.4.1. Diapause as an overwintering strategy 16

1.2.4.2. Effects of low and subzero temperatures on biological systems 17

1.2.4.3. Molecular adaptations in O. nubilalis during diapause and cold hardening........ 19

2. Thesis hypothesis and aims 21

3. Materials and methods 23

3.1. Experimental design of the study........ 24

3.1.1. Pilot sampling and experimental design........ 24

3.1.2. Main experiment sampling and design........ 25

3.2. Proteomic and bioinformatics analyses........ 27

3.2.1. Sample preparation for proteomic analyses........ 27

3.2.2. 2D PAGE enrichment validation 27
3.2.3. LC-MS/MS protein identification 29
3.2.4. Bioinformatic analyses 31
3.3. Gene expression analysis of selected IDPs 33
3.3.1. Isolation of total RNA 33
3.3.2. RNA concentration and quality assessment 34
3.3.3. RT-PCR cDNA synthesis 35
3.3.4. Primer design 35
3.3.5. Primer specificity analysis 37
3.3.6. Primer efficiency determination 38
3.3.7. Quantitative PCR of genes of interest 39
3.3.8. Statistical analysis of qPCR results 40
4. Results 41
4.1. Proteomic analyses 42
4.1.1. 2D PAGE confirmation of IDP enrichment procedure 42
4.1.2. LC-MS/MS identification of proteins 43
4.1.3. Effect of sample heating on protein identification 45
4.2. Bioinformatic analyses 47
4.2.1. Determination of protein disorder 47
4.2.2. Effect of sample heating on protein disorder distribution 50
4.2.2.1. Pilot experimental setup 50
4.2.2.2. Main experimental setup 52
4.2.3. Determination of long intrinsically disordered regions (long IDRs) 54
4.2.3.1. Pilot experimental setup 54
4.2.3.2. Main experimental setup 56

4.2.4. Amino acid compositional analyses ... 58

4.2.4.1. Total amino acid content ... 58

4.2.4.2. Individual amino acid content ... 59

4.2.5. Functional analysis of disordered proteins ... 60

4.3. Relative gene expression analyses ... 62

4.3.1. Ct values of reference genes ... 62

4.3.2. Relative expression of *hsp90* gene ... 63

4.3.3. Relative expression of *hsc70* gene ... 65

4.3.4. Relative expression of *hsp70* gene ... 67

4.3.5. Relative expression of *hsp20.4* gene ... 69

4.3.6. Relative expression of *hsp20.1* gene ... 71

4.3.7. Relative expression of *tropmy2* gene ... 73

4.3.8. Relative expression of *tnt* gene ... 75

4.3.9. Relative expression of *thym* gene ... 77

4.3.10. Relative expression of *moe* gene ... 79

5. Discussion ... 81

5.1. Proteomic analyses ... 82

5.1.1. Enrichment of IDP content and enrichment validation ... 82

5.1.2. Identification of proteins by LC-MS/MS ... 84

5.1.3. Effect of sample heating on protein identification ... 85

5.2. Bioinformatical analyses ... 86

5.2.1. Determination of intrinsic disorder ... 86

5.2.2. Effect of sample heating on protein disorder distribution ... 87

5.2.3. Long intrinsically disordered regions (long IDRs) ... 87

5.2.4. Amino acid compositional analyses ... 89

5.2.4.1. Glutamic acid, lysine, glutamine and proline 90

5.2.4.2. Compositional bias 92

5.2.5. Functional analysis of disordered proteins 92

5.3. Ecophysiological aspects of proteins with intrinsic disorder in cold hardiness 94

5.3.1. IDP content in *O. nubilalis* cold hardiness 94

5.3.2. Gene expression analyses of disorder-containing proteins 97

5.2.4.1. Heat shock protein genes 97

5.2.4.2. Structural protein genes 102

5.2.4.3. Thymosin beta and moesin genes 105

5.3.3. IDPs as proposed mediators of *O. nubilalis* cold adaptation 107

6. Conclusions 109

7. Literature 112

8. Supplementary materials 139

9. Extended abstract in Serbian 142

Author biography 165

Abstract

Intrinsically disordered proteins (IDPs) are a specific and diverse group of proteins that, under physiological environmental conditions, do not possess higher levels of structural organization, but are found in a form resembling denatured proteins. When IDPs are exposed to conditions that are denaturing for ordered proteins, such as high and low temperatures or strongly acidic or basic pH of environment, they acquire transient forms of ordered structure. Also, a large number of typically ordered, globular proteins have intrinsically disordered regions (IDRs) in their structure. Contrary to the established view that the function of proteins depends on their ordered structure, research has shown that IDPs possess various molecular functions regardless of the absence of structure. Due to the lack of higher orders of structure, IDPs have the ability to recognize a large number of different biomolecules and interact with them, and during these interactions they partially fold and acquire a more ordered structure. Due to their flexibility and plasticity in intermolecular interactions, IDPs and proteins with intrinsically disordered regions are often involved in cell signaling processes, regulation of transcription and translation, and also function as molecular chaperones. Based on their diverse roles, as well as resistance to denaturation under unfavorable conditions, IDPs probably have a pronounced biological significance in adaptive processes of organisms to abiotic stress factors. In order to evaluate their potential importance in the process of adaptation to hypometabolic living conditions during phases of arrested development and developing cold resistance, in this dissertation 5th instar larvae of the European corn borer *Ostrinia nubilalis* were used. In order to survive the cold winter months this species enters a specific resting state called diapause. During diapause, corn borer larvae undergo biochemical and molecular changes that allow them to survive only on stored energy sources and to develop resistance to extremely low winter temperatures. For this reason, the European corn borer is often used as an *in vivo* model system for studying these processes. In this dissertation, a method was established and optimized for enriching the content of IDPs, validation of the enrichment procedure with specific 2D polyacrylamide gel electrophoresis and isolation of disordered proteins from whole-body homogenates of larvae that were subjected to different temperature treatments to develop low temperature resistance. Isolated proteins were identified using liquid chromatography coupled with tandem mass spectrometry. Also, the degree of intrinsic disorder of the identified proteins was determined by *in silico* analyses, quantitative and qualitative parameters of their disorder were evaluated, and the functional characterization of the proteins was

performed. The relative expression of selected genes for proteins with a significant degree of intrinsic disorder was also examined in 5th instar larvae that were reared under non-diapausing conditions, as well as diapause-inducing conditions that also lead to the development of cold resistance. The results showed that in this species there is a significant content of IDPs and proteins with disordered regions, and that the proportion of proteins with IDRs is correlated with the acquisition of cold resistance. The expression of most of the analyzed genes is also correlated with this adaptation. The results of this dissertation provide new insights into the molecular basis of processes that enable the European corn borer to adapt to unfavorable living conditions characterized by a reduced metabolic rate, limited energy sources for basal metabolism, as well as the negative effects of exposure to low temperatures on the functional and structural homeostasis of proteins and processes they are involved in.

key words: intrinsically disordered proteins; intrinsically disordered protein regions; *Ostrinia nubilalis*; cold hardiness; diapause; gene expression

Извод

Структурно-неуређени протеини (*енгл.* intrinsically disordered proteins, IDP-ови) су специфична и разноврсна група протеина који при физиолошким условима средине не поседују више нивое структурне организације, већ се налазе у облику који наликује денатурисаним протеинима. Када су IDP-ови изложени условима који су денатуришући за уређене протеине, као што су високе и ниске температуре или изразито кисела или базна pH средине, они задобијају пролазне облике уређене структуре. Такође, велик број типично уређених, глобуларних протеина у својој структури поседују регионе који могу бити структурно-неуређени (*енгл.* intrinsically disordered regions, IDR-ови). У супротности са устаљеним виђењем да функција протеина зависи од њихове уређене структуре, истраживањима је показано да IDP-ови поседују разноразне молекуларне функције без обзира на одсуство структуре. Услед одсуства виших нивоа структуре, IDP-ови имају способност препознавања великог броја других биомолекула и интераговања са њима, а приликом тих интеракција се делимично савијају и задобијају пролазу уређену структуру. Због своје флексибилности и пластичности у међумолекулским интеракцијама, IDP-ови и протеини са структурно-неуређеним регионима често су укључени у процесе ћелијског сигналинга, регулацију транскрипције и транслације, а поседују и функције молекуларних шаперона. На основу њихових разноликих улога, као и отпорности на денатурацију у неповољним условима, IDP-ови вероватно имају изражен биолошки значај у процесима адаптација организама на абиотске факторе стреса животне средине. Како би се проценио њихов претпостављени значај у процесу адаптације организма на хипометаболичке услове живота током фазе мировања и развијања отпорности на хладноћу, у овој дисертацији коришћене су гусенице 5. инстара врсте кукурузног пламенца *Ostrinia nubilalis*. За ову врсту је карактеристично да у зимском периоду улази у фазу мировања звану дијапауза, како би преживела хладне зимске месеце. Током дијапаузе, у гусеницама кукурузног пламенца долази до промена на биохемијском и молекуларном нивоу које им омогућавају да преживе само на ускладиштеним изворима енергије као и да развију отпорност на изузетно ниске зимске температуре. Због тога се кукурузни пламенац често користи као *in vivo* модел-систем у истраживањима ових процеса. У овој дисертацији успостављена је и оптимизована метода за обогаћивање садржаја IDP-ова, валидацију поступка обогаћивања специфичном 2Д електрофорезом на полиакриламидном гелу и изолацију неуређених протеина из

хомогената целих гусеница које су подвргаване различитим температурним третманима за развој отпорности на ниске температуре. Изоловани протеини су идентификовани помоћу течне хроматографије куповане са тандемском масеном спектрометријом. Такође, *in silico* анализама су утврђени степен структурне неуређености идентификованих протеина, процењени квантитативни и квалитативни параметри њихове неуређености и урађена је функционална карактеризација протеина. Испитана је и релативна експресија одабраних гена за протеине са утврђеним значајним степеном структурне неуређености у гусеницама 5. инстара држаним у условима недијапаузе, као и дијапаузе уз развијање отпорности на хладноћу. Резултати су показали да у овој врсти постоји значајан удео IDP-ова и протеина са неуређеним регионима, као и да је удео протеина са IDR-овима корелисан са стицањем отпорности на хладноћу. Експресија већине анализираних гена је такође корелирана са том адаптацијом. Резултати ове дисертације дају нове увиде у молекуларне основе процеса који омогућавају врсти кукурузног пламенца да се прилагоди неповољним условима живота које карактеришу смањена стопа метаболизма, ограничени извори енергије за потребе базалног метаболизма, као негативне ефекте излагања ниским температурама на структурну и функционалну хомеостазу протеина и процеса у којима су они укључени.

кључне речи: структурно-неуређени протеини; структурно-неуређени региони протеина; *Ostrinia nubilalis*; отпорност на хладноћу; дијапауза; експресија гена

1. Introduction

For more than a century it has been a firmly held belief that proteins must exist in a stable, 3D conformation in order to be functional. This belief was established by the work of Emil Fischer (1894), and later strengthened by Hsien Wu's research (1929), as well as that of Mirsky and Pauling (1936). Through their work it was postulated that an enzyme's shape or structure is determined by the substrate that binds to it. This reasoning was extended to include all proteins in general, forming the *structure-function* paradigm. However, the paradigm was challenged by the growing number of proteins with multiple different binding partners of various shapes and sizes. This led to the formulation of the induced fit theory which explained that a protein's active site adapts to the shape of a substrate until it is firmly bound (Koshland, 1958). Then, the end of the 20th and the beginning of the 21st century saw several research groups, independently from one another, recognize that higher, ordered tiers of protein structure are not a requirement for protein function. Several terms were coined to describe these proteins, such as *natively unfolded* (Uversky, 2002), *intrinsically unstructured* (Wright and Dyson, 1999) or *intrinsically disordered* (Dunker et al., 2001), with the consensus settling on the lattermost term (Hatos et al., 2020).

1.1. Intrinsically disordered proteins – IDPs

Intrinsically disordered proteins are present in all organisms, from prokaryotes to eukaryotes, and it is estimated that they constitute a significant portion of the proteome in an individual organism (Burra et al., 2010; Pancsa and Tompa, 2012; Xue et al., 2012). Considering the fact that these proteins are in an unstructured, disordered state when placed under physiological conditions, they possess considerable resistance to those adverse conditions that would lead to the denaturation of globular, ordered proteins (Uversky et al., 2001; Uversky, 2009). Intrinsic disorder (ID) is a structural property that is also present in limited amounts in proteins that are considered typically ordered. In those proteins, some parts of their structures contain such a composition of amino acids that they exist as structurally disordered regions of varying lengths (Necci et al., 2016). Intrinsically disordered proteins and proteins containing long intrinsically disordered regions (long IDRs) are involved in numerous biological processes in the cell. To name a few, they can act as molecular chaperones, scavenge toxins and heavy metals, function as regulatory factors, as well as be involved in a plethora of cell signaling pathways (Deiana et al., 2019; Vinterhalter et al., 2019). From a research standpoint, intrinsically disordered proteins are of great importance as their

dysregulation and altered functionality are at the core of several severe disorders, such as Parkinson's and Alzheimer's diseases (Uversky et al., 2008; Babu, 2016).

1.1.1. The code for protein intrinsic disorder

Whether a protein will adopt a stable and rigid tertiary conformation or remain unfolded in an extended and flexible structural state under physiological conditions is written down in its primary, amino acid sequence. Compositional analyses have revealed that there are significant and quite distinct differences in the amino acid compositions between globular and disordered proteins, as well as proteins containing long IDRs. In ID-containing proteins, the primary sequence is particularly enriched in amino acids that have thus been termed disorder-promoting – Lys, Ser, Arg, Pro, Glu, Gly, Gln and Ser (Table 1). Conversely, they are depleted in Cys, Asn, Phe, Ile, Leu, Val, Trp and Tyr, which have therefore been termed order-promoting amino acids (Table 1) (Uversky et al., 2000; Williams et al., 2000; Campen et al., 2008).

Apart from this compositional bias, intrinsic disorder is also characterized by reduced complexity of sequences. Unlike in structured proteins, the primary sequence of IDPs is less variable when it comes to the utilization of different amino acids. As a result, their sequences often contain stretches of multiple sequence and single amino acid repeats, such as poly-Glu, and are informationally simpler in comparison to the primary sequences of globular proteins. Even in ordered proteins multiple sequence repeats are often enriched in disorder-promoting residues (Simon and Hancock, 2009; Jorda et al., 2010; Darling and Uversky, 2017).

Owing to their peculiar amino acid makeup, intrinsically disordered proteins differ from structured proteins also when it comes to several physicochemical properties. Namely, IDPs lack the hydropathy and relatively neutral net charge at physiological conditions, both of which facilitate the folding of proteins into stable, ordered structures. Instead, due to many uncompensated charged residues, intrinsically disordered proteins have an overall high net charge. Depletion of hydrophobic amino acids also increases the solvability of these proteins. Taken together, these properties are at the core of why IDPs do not fold into ordered conformations unprompted. The high net charge leads to electrostatic repulsions between residues along the polypeptide chain, while the low hydropathy eliminates a major driving force of protein compaction into an ordered structure (Uversky et al., 2000; Uversky, 2011, 2019).

Table 1. Proteinogenic disorder- and order-promoting amino acids, their three- and one-letter codes and properties conferred by their side chains. Amino acids are ranked from top to bottom according to their disorder- and order-promoting properties, respectively (Campen et al., 2008).

Disorder-promoting		Order-promoting	
Amino acid	**Side chain property**	**Amino acid**	**Side chain property**
Proline (Pro, P)	pyrrolidine loop	Tryptophan (Trp, W)	hydrophobic, aromatic
Glutamic acid (Glu, E)	polar, negatively charged	Phenylalanine (Phe, F)	hydrophobic, aromatic
Serine (Ser, S)	polar, uncharged, hydroxymethyl group	Tyrosine (Tyr, Y)	hydrophobic, aromatic, hydroxymethyl group
Lysine (Lys, K)	polar, positively charged	Isoleucine (Ile, I)	hydrophobic, aliphatic
Glutamine (Gln, Q)	polar, uncharged	Methionine (Met, M)	hydrophobic, contains sulphur
Histidine (His, H)	polar, positively charged	Leucine (Leu, L)	hydrophobic, aliphatic
Aspartic acid (Asp, D)	polar, negatively charged	Valine (Val, V)	hydrophobic, aliphatic
Arginine (Arg, R)	polar, positively charged	Asparagine (Asn, N)	polar, uncharged
Glycine (Gly, G)	single hydrogen atom	Cysteine (Cys, C)	thiol group
Alanine (Ala, A)	hydrophobic	Threonine (Thr, T)	polar, uncharged, hydroxymethyl group

1.1.2. Localized intrinsic disorder – intrinsically disordered regions

While proteins can be completely unstructured and unable to fold into a stable conformation along the entire length of their polypeptide chains, otherwise ordered proteins can also contain segments in their sequences that are fundamentally disordered. Computational analyses of available proteomic data have revealed that such intrinsically disordered regions (IDRs) are abundant in eukaryotic proteins, and even enriched in the functional group of signaling proteins (Iakoucheva et al., 2002; Xue et al., 2012). Intrinsically disordered regions are flexible, extended and highly dynamic, transitioning through different conformational states depending on their environment or interaction with other molecules (van der Lee et al., 2014). A typical mode of operation for an IDR

is to remain in a disordered state until it interacts with a binding partner. At such a moment, it will undergo a disorder-to-order transition upon binding with the physiological partner in question. Such a transition is temporary, and the IDR will return to a disordered state once the interaction has ended (Eliezer, 2009; Uversky, 2011, van der Lee et al., 2014). Ordered proteins can also contain segments that are essentially folded, however they possess an inherent tendency to unfold due to their amino acid composition. In contrast to the previously mentioned disorder-to-order transition, these segments can undergo an opposite transition – order-to-disorder. This type of transition, where an ordered protein needs to unfold some of its segments in order to be functional, is usually induced by changes in the protein's environment or when its structure is modified in some way (Jakob et al., 2014; Uversky, 2015). Intrinsically disordered regions are typically located at the termini of proteins and their position is correlated with protein function. Proteins that are involved in the activation or repression of transcriptional processes have C-termini that are enriched in disordered residues, while disorder-rich N-termini are commonly found in proteins that exhibit DNA-binding activity (Lobley et al., 2007; Pentony and Jones, 2010; Vuzman et al., 2010).

1.1.3. Environmental hardiness of IDPs

Globular proteins are beholden to stable environmental conditions in order to preserve their structure and, ultimately, function. When environmental conditions deteriorate, it can lead to the denaturation of proteins and subsequent loss of function. It is well known that globular proteins are susceptible to adverse factors such as elevated temperatures and drastic changes in pH of their immediate environment. Intrinsically disordered proteins, on the other hand, are already in a denatured state when conditions are at the physiological level. Exposing these proteins to environments that are detrimental to globular proteins elicits a drastically different response from IDPs (Tantos et al., 2009; Uversky, 2009, 2013a; Smith and Jelokhani-Niaraki, 2012).

In ordered proteins, high temperatures disrupt the conformational forces, such as hydrogen bonds, that hold their 2-D and 3-D structure together. As these forces dissipate, the ordered protein begins to denature and its functionality is irreversibly lost (Tanford, 1968). When it comes to IDPs, however, elevated temperatures cause these proteins to fold in part and obtain transient secondary structures. It is likely that high temperatures strengthen hydrophobic interactions, which are otherwise weak in IDPs, and force the unstructured proteins to compact and fold. Unlike the

irreversible denaturation of globular proteins, this induced partial folding of IDPs is reversible and the proteins return to a disordered state once the temperature has been lowered (Uversky, 2009, 2013a). Cold temperatures can also lead to reduction or loss of function in globular proteins under certain conditions (Marqués et al., 2003; Chandrayan and Guptasarma, 2008). Contrary to heat denaturation, native proteins can retain or restore their functionality after being frozen and returned to room temperature. Nonetheless, repeated freezing and thawing will disturb their structure and function. In the case of IDPs, they remain largely unaffected by the events of temperature decrease and water freezing, mostly due to their lack of any stable conformation (Tantos et al., 2009).

Apart from extreme temperatures, drastic changes in environmental pH can cause structure loss in ordered proteins (Tanford, 1968). Intrinsically disordered proteins, on the other hand, contain a significant amount of charged residues due to their biased amino acid composition. Electrostatic repulsion events between those residues prevent IDPs from folding. In alkaline or acidic environments, these charged residues are neutralized and the electrostatic repulsions are removed, opening the path for IDPs to fold into transient secondary structures (Uversky, 2009, 2013a; Smith and Jelokhani-Niaraki, 2012).

1.1.4. Natural abundance of intrinsic disorder

After being discovered, intrinsically disordered proteins were first believed to be exceptions to the long established rule of structure equaling function (Peng et al., 2015). However, different computational methods were developed that allowed the prediction of intrinsic disorder in available protein sequences, leading to a change in perspective (Xue et al., 2012). Once considered to be rare occurrences, intrinsically disordered proteins and protein regions have been shown to be actually abundant in nature (Uversky, 2011). It is now an established fact that IDPs and proteins containing IDRs are present in all living organisms, as well as in viruses, in varying degrees (Kulkarni and Uversky, 2018). A common feature of intrinsic disorder is that it is consistently higher in proteomes of eukaryotes compared to prokaryotic proteomes (Pancsa and Tompa, 2012; Peng et al., 2015). Depending on species, the fraction of disordered residues in eukaryotic proteomes can vary between 35 and 50% (Xue et al., 2012). The content of structural disorder seems to be strongly correlated with the lifestyle an organism leads, as the highest amounts of disorder were predicted in parasites that change several hosts (Pancsa and Tompa, 2012).

1.1.5. Intrinsic disorder and protein function

One of the most important features, if not the most important feature, of intrinsically disordered proteins, as well as intrinsically disordered protein regions, is that they are functional despite lacking higher tiers of ordered structure. Rather, it is their intrinsically disordered nature that gives rise to the diverse array of functions these proteins can perform and the processes they are involved in. There are several modes how intrinsically disordered proteins can exhibit their function. One such mode is the transition from a disordered state into an ordered one upon interaction with a binding partner. Once a partner is recognized, the intrinsically disordered protein, protein region, undergoes induced folding and binds with the partner with high specificity and low affinity (Uversky, 2011). Protein-protein interactions are often enabled by specific sequences such as short linear motifs (SLiMs), molecular recognition features (MoRFs) and eukaryotic linear motifs (ELMs). Most of these motif types have been located in protein regions that are intrinsically disordered (Fuxreiter et al., 2007; Tompa, 2012). Due to their flexibility and plasticity, as well as the ability to recognize and bind to different partners, IDPs are often involved in processes pertaining to regulation of signaling pathways, transcription, translation and cellular differentiation (Tompa, 2012; Wright and Dyson, 2015; Bondos et al., 2021). Transcriptional factors in particular have been found to be disordered in a significant degree (Minezaki et al., 2006), while intrinsic disorder has also been correlated with chaperone function (Reichmann et al., 2012).

Apart from this operational mode of disorder-to-order transitions, IDP functions can also stem from proteins remaining persistently disordered (Tompa, 2012; Uversky, 2015). As has been mentioned, the terminal parts of proteins, their tails, are often disordered (Lobley et al., 2007; Pentony and Jones, 2010; Vuzman et al., 2010). These intrinsically disordered protein tails, such as those found in neurofilaments H and M, function as entropic bristles, using their flexibility to ensure that neighboring neural filaments remain separate, as well to maintain the shape of axons (Brown and Hoh, 1997, Uversky, 2013b). Additionally, these entropic bristles can sweep away molecules that would interfere with the proper folding of target protein, increase the solubility of proteins they are fused with and prevent protein aggregation (Santner et al., 2012).

1.2. The European corn borer, *Ostrinia nubilalis* (Hbn., 1796)

The European corn borer (ECB), *Ostrinia nubilalis* (Hbn., 1796) is an insect species belonging to the order Lepidoptera which includes moths and butterflies. It is a member of the Pyraloidea superfamily which is comprised of two families – Crambidae and Pyralidae. The *Ostrinia* genus in general belongs to the subfamily Pyraustinae, which has a contentious classification. For a long time the subfamily was placed under the Pyralidae family, while more recently phylogenomic studies have firmly placed the Pyraustinae subfamily as members of the Crambidae family (Table 2) (Mutuura and Munroe, 1970; Solis, 2007; Léger et al., 2020; Yang et al., 2021). Even the species name was not without contention, and had undergone several changes (e.g. *Pyralis nubilalis*, *P. silacealis*, *Pyrausta nubilalis*) before being settled on the present *Ostrinia nubilalis* (Caffrey and Worthley, 1927; Mutuura and Munroe, 1970).

Table 2. Scientific classification of the European corn borer, *O. nubilalis*.

Taxonomic category	
Kingdom	Animalia
Phylum	Arthropoda
Class	Insecta
Order	Lepidoptera
Superfamily	Pyraloidea
Family	Crambidae
Subfamily	Pyraustinae
Genus	*Ostrinia*
Species	*Ostrinia nubilalis*

1.2.1. History, distribution and ecology

Ostrinia nubilalis is a polyphagous pest moth species that is widespread in temperate regions of Europe, as well as those of northern Africa and western Asia. It is believed that the moth originated in Europe, as the first records of the insect’s presence on this continent are dated to around the 1500s (Bethenod et al., 2005), while its earliest economic impact on European crops was recorded in 1835 (Caffrey and Worthley, 1927). Apart from the Old World, the ECB was introduced to the North American continent most likely by way of broom corn shipments (*Sorghum bicolor*) from

Austria-Hungary and Italy between 1909 and 1914, with the first specimens being discovered near Boston, Massachussets in 1917 (Smith, 1920; Caffrey and Worthley, 1927). From there, the insect spread westward to the Rocky Mountains in both the United States and Canada (Willet and Harrison, 1999).

Corn (*Zea mays*) is the primary host plant that larvae of the ECB attack and feed on. Young larvae largely feed on the corn tassel, moving into the stalk as they grow and undergo several molting phases. Once inside, the larvae cause significant structural damage to the corn stalk while feeding, which can often lead to the plant snapping and lodging if it is infested with multiple larvae. *O. nubalis* infestation of corn stalks can be recognized by the characteristic tunnels that they bore throughout the plant. Apart from corn, the ECB has been confirmed as a pest of over 200 different crops, such as hop (*Humulus lupulus*), mugwort (*Artemisia vulgaris*), broom corn (*Sorghum vulgare*), potato (*Solanum tuberosum*), sweet pepper (*Capsicum anuum*) and tomato (*Solanum licopersicum*) (Capinera, 2000; Kuhar et al., 2004; Sole et al., 2010). Common for all these plants is that their stems provide enough space for ECB larvae to bore and settle into.

1.2.2. Insect morphology

The European corn borer is a holometabolous insect and undergoes a complete metamorphosis during its development, passing through four distinct stages: embryo or egg, larva, pupa and imago or adult. ECB eggs are oval shaped and flattened, and deposited in clusters of 15 to 20 at a time. Freshly laid eggs are creamy or greenish white in color and opaque in the middle (Fig. 1A). As they mature, the eggs take on a yellowish hue and the black head capsule of the enclosed larva can be seen due to rapid chitinization (Caffrey and Worthley, 1927; Capinera, 2000). Larvae hatch around 24 hours after the head capsule has visibly formed. Newly hatched larvae are less than 2 mm in length, their body pale yellow in color with a pinkish dorsum, and with a dark brown head capsule. The body is divided into 14 segments – 3 thoracal with a pair of true legs on each, 10 abdominal with pairs of prolegs on the 3rd, 4th and 5th abdominal segments, and one terminal segment, also with a pair of prolegs. A dark streak is present on the dorsal side, as well as small dark marks on every segment. As larvae feed and grow, they molt and in field conditions usually go through a total of five developmental stages or instars (Fig. 1B). However, in some cases they can go through six or even seven (Caffrey and Worthley, 1927; Capinera, 2000). Final instar larvae pupate inside the bored tunnels. Early pupae are light brown in color and darken with age

(Fig. 1C). Sexual dimorphism is pronounced at this stage, as male pupae are smaller than the female ones, and the genital openings are differently positioned between the sexes (Caffrey and Worthley, 1927; Capinera, 2000). Adult moths that emerge from pupae are small, again with pronounced sexual dimorphism. Males have a wingspan of 20 to 26 mm and grayish brown in color with dark zigzag lines on their wings (Fig. 1D). Females have a larger wingspan of 25 to 34 mm and are of lighter color than the males. Their wings also contain dark zigzag lines (Fig. 1E). Despite the difference in wingspan, male and female moths are of similar body length – 13 to 14 mm (Caffrey and Worthley, 1927; Capinera, 2000).

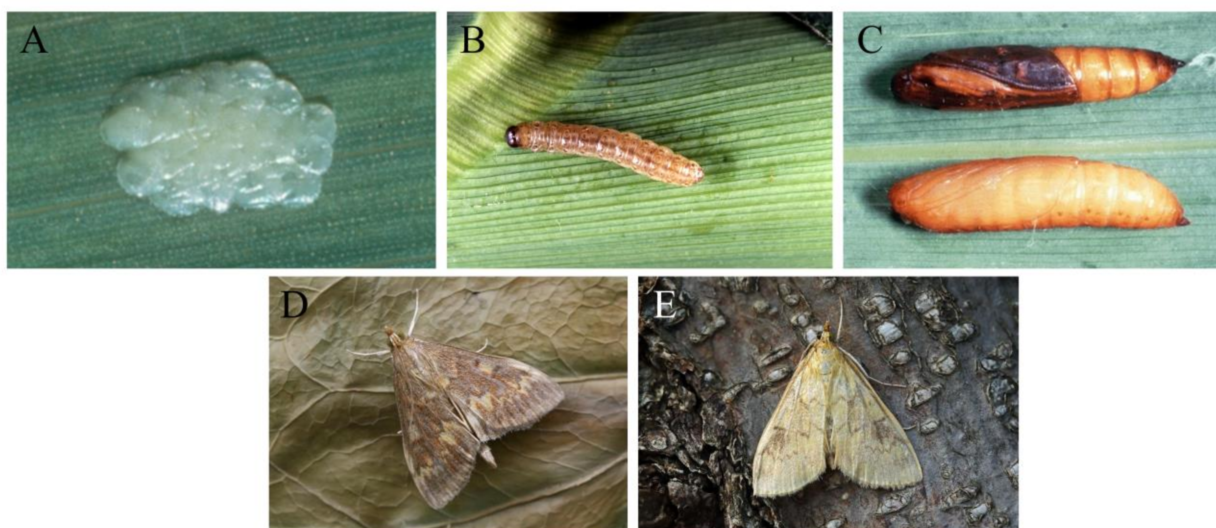

Figure 1. Life stages of *O. nubilalis*: (A) egg cluster; (B) 5th instar larva; (C) pupae – male (above), female (below); (D) male adult; (E) female adult. Image credits: (A) and (B) United States Department of Agriculture; (C) Fabrizio Santi (CC BY-NC-ND 4.0); (D) and (E) ©entomart (entomart.be).

1.2.3. Life cycle of *Ostrinia nubilalis*

The four aforementioned life stages constitute the life cycle of one ECB generation. In field conditions, one life cycle takes between 4 and 6 weeks to complete. Up until the 1980s, *O. nubilalis* was predominantly a univoltine species in Europe, producing one generation per year (Bača et al., 2007). However, in the last 30 years, voltinism of the ECB has been affected by climate changes, primarily temperature fluctuations and amounts of precipitation. The ECB has progressed to being a multivoltine species, most often producing two generations per year – one summer and one winter. That said, there have been years when three generations were detected, with an additional summer generation being produced due to adult moth flights happening in late spring and early

summer (Vajgand, 2010), and the number of yearly generations is likely to increase following the rise of global temperatures (Kocmánková et al., 2008). In the upper parts of North America the ECB usually produces one or two generations per year, while in southern locations three generations are more common with some regions experiencing four per year. In such locations adult moth flights were recorded from as early as April and as late as September (Capinera, 2000).

The first generation of the year spawns from 5th instar larvae that had overwintered in a state of arrested development called diapause (explained in detail in *1.2.4.1. Diapause as an overwintering strategy*). In late spring, overwintering larvae terminate their diapause program and pupate. This developmental stage lasts around 12 days in field conditions, with pupae requiring an ambient temperature of at least 13°C to develop correctly. First generation adult moths emerge from the cocoons and are mostly active during the nighttime. Imagos feed on plant nectar and live between 18 and 24 days. Female moths begin to lay eggs from 3 to 4 days old. Eggs, 15 to 20 at a time, are usually deposited on the underside of leaves, close to the midrib. The period of oviposition lasts around two weeks. With this, the life cycle of the winter generation is completed (Capinera, 2000; Boyd and Bailey, 2001; Sekulić et al., 2008). Depending on environmental conditions, egg development can last from 4 to 9 days, and generally requires an ambient temperature of at least 15°C. Larvae that hatch from these eggs represent the first generation of the year, which is also considered the summer generation (Fig. 2).

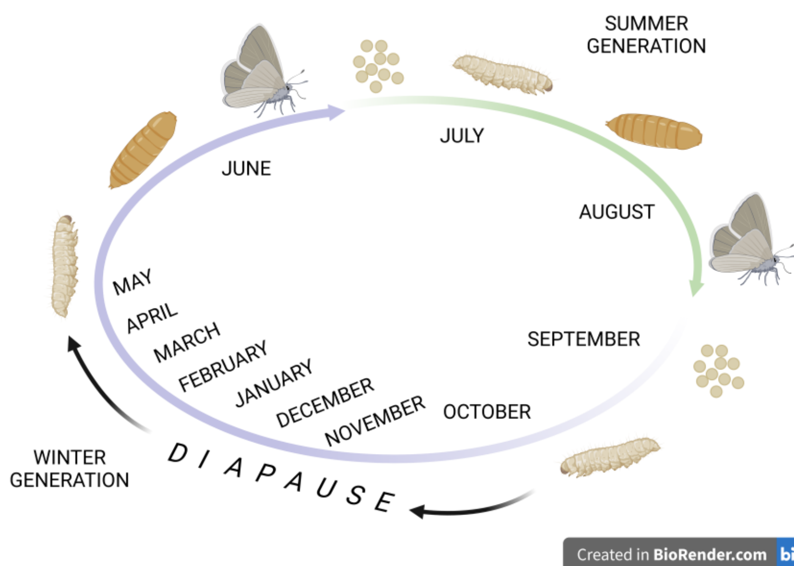

Figure 2. Life cycle of bivoltine *O. nubilalis* specimens. Image by Vanja Tatić (created with BioRender.com).

As has been said, during larval development individuals go through at least 5 instar stages. On average it takes around 50 days for larvae to fully develop, depending on weather conditions, after which they pupate and metamorphose into adult moths (Capinera, 2000; Boy and Bailey, 2001). Regardless of whether there is one or more summer generations of moths during a year, the latest one will lay eggs from which the winter generation will hatch (Capinera, 2000). As summer comes to an end and autumn begins, the days shorten and ambient temperatures drop, all of which serves as an environmental signal that will trigger the larvae prepare for the coming winter.

1.2.4. *Ostrinia nubilalis* as a model system for diapause and cold adaptation

Apart from being an economically important and widespread pest insect species, the ECB is also known for being able to seasonally arrest its development by entering diapause and waiting out the cold winter months, after which it resumes its regular ontogeny. The life cycle of this species has been extensively described, allowing for experiments to be set up with great consistency. Because of this, the ECB has for several decades been used as a reliable model system for studying the mechanisms that govern diapause and adaptations for enduring cold environmental temperatures. In the ECB, these two processes are tightly linked, as exposure to cold temperatures during diapause is one of the necessary triggers for adequate metabolic changes to occur which ensure the insect's survival during the winter months (Popović et al., 2021).

1.2.4.1. Diapause as an overwintering strategy

One of the most significant seasonal changes in the environment that affects insects inhabiting temperate and polar zones is the gradual onset of cold temperatures which can have lethal consequences unless countered. This period is also accompanied by ever increasing food scarcity, which forces insects to rely on energy reserves that were accumulated in the preceding months. An additional challenge that is posed to insects during this time is how to stretch out the limited reserves and endure until conditions improve and food becomes available again. To that end, insects have, over the long course of evolution, acquired adaptations that help them face and survive such harsh living conditions.

In order to conserve their limited energy stores, insects enter a hypometabolic resting state called diapause. During diapause growth and development are arrested, the insect's metabolic rate is significantly reduced (Danks, 1987; Denlinger, 2009) and tolerance to abiotic stressors is increased

(Rinehart et al., 2000). Despite the reduction in the intensity of metabolic processes, diapause is complex and dynamic, and not just a passive stage in an insect's life cycle. Diapause is also important for insect survival as it is induced ahead of the seasonal changes to which the insect needs to adapt. The most common environmental cue for insects that is time to begin preparations for diapause is the shortening of the photoperiod (Nylin, 2013). Insects can enter diapause at any life stage, and the exact life stage when that occurs is species-specific (Nation, 2008). Diapause is divided into three distinct ecophases, some with their own *subphases* (Koštál, 2006; Koštál et al., 2017):

- Pre-diapause – *induction* and *preparation*;
- Diapause – *initiation* and *maintenance*;
- Post-diapause or quiescence.

Once diapause is induced, an insect will undergo numerous changes on behavioral, morphological, physiological, biochemical and molecular level in order to ensure its survival during the harsh period. These include cessation of feeding, reduction of mobility, alterations in the structure of cuticle, suppression of less important physiological processes, the aforementioned lowering of general metabolic rate, expression of specific stress-related genes and others (Denlinger, 2002; Storey and Storey, 2004; MacRae, 2010). Concurrently with these metabolic changes diapausing organisms also develop increased resistance to abiotic stressors such as dehydration, UV radiation and low/high temperatures. In insects of polar and temperate regions, such as *O. nubilalis*, increased hardiness to cold weather conditions is the most common adaptation that occurs during diapause.

1.2.4.2. Effects of low and subzero temperatures on biological systems

Environmental temperatures greatly affect different components of biological systems, which rely on specific temperatures ranges in order to remain functional and ensure the survival of an individual. Cold and sub-zero ambient temperatures present living beings with unique and dangerous challenges which are fatal if they cannot be overcome by adaptations on behavioral, physiological, biochemical and molecular levels. In cold environments, organisms need to cope with issues such as freezing of intra- and extracellular water, denaturation of proteins induced by cold temperatures, decreased fluidity of membranes, desiccation, lowered transcription and translation rates, just to name a few.

In living organisms water can constitute 90% or more of their total body weight, depending on environmental and developmental factors (Hadley, 1994). Water is also the most abundant molecule in living systems, accounting for 95–99% of all molecules in a particular organism (Edney, 1977). Because of this overwhelming presence, water is the biomolecule that is most susceptible to be adversely affected by low temperatures. In turn, freezing can cause the most devastating consequences for living beings. As temperatures drop below the freezing point of water, ice crystals begin to form in the extracellular spaces. The growing ice crystals can disrupt and pierce membranes, leading to cell death. As more water turns to ice, less of it is readily available to dissolve biomolecules, minerals and other compounds. Moreover, this establishes a concentration gradient between the extra- and intracellular environments and water begins to flow outside of the cell in order to reestablish the concentration equilibrium across the cell membrane. However, this process will ultimately end with the cell dying, as the extracellular freezing will cause the cell to lose a critical amount of water (Mazur, 1970, 1984; Zacchariassen and Kristiansen, 2000; Chian and Quinn, 2010).

Less dramatic, but no less dangerous are the effects cold temperatures can have on biological processes in living systems. These effects are mediated by water, as it is the major component of body fluids where biological processes take place. As the temperature of fluids decreases, so does the available kinetic energy needed for enabling biochemical reactions to occur. This slows down biological processes and leads to a general depression of metabolism. Additionally, cold temperatures increase the stability of DNA and RNA secondary structures, making it difficult for transcriptional and translational machinery to access them and carry out protein synthesis (D'Amico et al., 2006). Enzymatic activity is also reduced, or completely absent, at low ambient temperatures. Enzymes undergo reversible denaturation, as the cold temperatures reduce the interactions keeping protein subunits together and they subsequently dissociate from each other. Similar effects occur in non-enzymatic proteins as well, with the reversible disassembly of supramolecular structures of microtubules and collagen fibrils under cold conditions (Gaskin et al., 1974; Gelman et al., 1979; Privalov, 1990).

To overcome these obstacles and ensure their survival in harsh conditions, organisms have adopted different strategies. These strategies differ based on the environment they inhabit, whether the

changes in temperature are seasonal or year-round and are depended on the life cycle phase the organism is currently in.

1.2.4.3. Molecular adaptations in O. nubilalis during diapause and cold hardening

In *O. nubilalis*, diapause and acclimation to cold ambient temperatures occur during the 5th instar of the larval stage. Diapause itself is photoperiodically induced by the shortening of days in autumn (Beck, 1962). This signals the larvae to begin preparing for the coming winter by accumulating energy reserves. During the course of diapause, larvae will not feed and will need to rely on these stores as sources of fuel for their metabolism (Hahn and Denlinger, 2007, 2011). At the same time ambient temperatures begin to gradually drop, which induces a cascade of changes on physiological and biochemical levels. One of the major changes is in how glycogen, the major reserve polysaccharide, is utilized. As glycogen is being broken down, the resulting glucose is directed towards the synthesis of glycerol. Glycerol is an important cryoprotective molecule that is synthesized in the larval fat body and exported into the haemolymph. As diapause progresses, the level of glycerol in the haemolymph increases, which lowers the supercooling point of the larva to below –20°C and enables it to become cold hardy (Nordin et al., 1984; Grubor-Lajšić et al., 1991; Andreadis et al., 2008). Apart from glycerol, other cryoprotective compounds are produced and accumulated, all of them making the insect increasingly cold hardy. These include sorbitol, trehalose and free amino acids such as alanine, proline and arginine (Koštál et al., 2007; Clark and Worland, 2008; Kojić, 2009; Popović, 2014; Purać et al., 2015).

Due to the redirection of glycogen and lack of feeding during diapause, the general metabolic rate needs to decrease in order to conserve energy and ensure survival. Diapausing larvae experience a depression of metabolism by cutting down on energy intensive processes, such as tissue differentiation, cell division and general organism growth and development (Popović et al., 2021). In addition, the cold temperatures slow down metabolic processes, further lowering an organism's metabolic rate (Storey and Storey, 2004). The energy profile of cold-acclimated diapausing larvae is also distinct from non-acclimated diapausing ones, indicating that exposure to low temperatures during diapause is likely an important component for triggering the necessary metabolic changes for survival (Storey and Storey, 2012). The levels of energy metabolism-related molecules, such as adenine, uridine and guanosine nucleotides, as well as NAD^+/$NADP^+$ coenzymes, are depleted in non-acclimated larvae compared to cold-acclimated larvae. In addition to this, the relative

expression of genes encoding different components of the electron transport chain in the non-acclimated diapausing larvae is indicative of increased metabolic rate. Taken together, these findings show that diapausing larvae which have not been exposed to cold temperatures are metabolically very active, which leads to the untimely depletion of their already restricted energy reserves and high mortality (Popović et al., 2021).

Other changes that have been reported to occur in regards to diapause and cold hardiness include shifting from aerobic to anaerobic metabolism (Uzelac et al., 2020), increase in the number of aquaporins in cell membranes (Izumi et al., 2007), alterations in the lipid composition of membranes by increasing the content of unsaturated fatty acids (Vukašinović et al., 2013, 2015, 2018) and expression of specific sets of stress protection genes and genes regulating the cell cycle and apoptosis (Denlinger, 2002; Rinehart et al., 2007; MacRae, 2010; Popović et al., 2015; Koštál et al., 2017).

Despite this extensive research, the molecular mechanisms governing insect diapause and development of stress-tolerance, in this case to cold ambient temperatures, have not been completely elucidated. While the biochemical and physiological aspects of these processes in the model system *O. nubilalis* have been broadly covered, one aspect that remains insufficiently studied is the effect of these metabolic and physiological changes on the proteomic level.

2. Thesis hypothesis and aims

Studies have shown that IDPs exhibit higher thermal stability at high, as well as low temperatures. Additionally, some IDPs better retain their functionality at cold temperatures, than ordered proteins. Therefore, considering that *O. nubilalis* is a species whose 5th instar larvae become cold hardy during diapause, it is an ideal model system for *in vivo* studies on the roles of IDPs in the molecular ecophysiology of cold hardiness. The hypothesis of this dissertation is that the content of IDPs will be increased during the resting phase of *O. nubilalis* larvae at low temperatures, as a result of gradual development of cold hardiness.

Taking into consideration the unique primary structure of IDPs and proteins with long IDRs, its effects on the thermal stability of these proteins, as well as the scarcity of information on their presence and importance in species that have developed adaptations to surviving cold temperatures in their environment, the aims in order to the hypothesis were as follows:

- to determine the presence of IDPs and proteins with long IDRs in *O. nubilalis* larvae by optimizing the existing methods for their detection;
- to analyze the content of IDPs and proteins with long IDRs in different experimental groups where specimens were placed in diapause-inducing conditions and gradually acclimated to cold temperatures;
- to bioinformatically determine the qualitative and quantitative properties of intrinsic disorder in the identified IDPs and proteins with long IDRs, as well as evaluate the level of intrinsic disorder;
- to bioinformatically determine the (potential) molecular functions of identified IDPs and proteins with long IDRs;
- to explore if there are significant changes in the qualitative and quantitative composition of IDPs and proteins with long IDRs in different experimental conditions that were more or less conducive for the development of cold hardiness;
- to analyze the expression of genes encoding IDPs and long IDR-containing proteins with either a high level of intrinsic disorder, and/or a significant biological role in thermal stress conditions;
- to ascertain, according to the results of the above mentioned analyses, the role of IDPs in the ecophysiological basis of diapause and development of cold hardiness in larvae of *O. nubilalis*.

3. Materials and methods

3.1. Experimental design of the study

3.1.1. Pilot sampling and experimental design

Diapausing (winter) 5th instar larvae of the ECB were collected from the fields of the Maize Research Institute in Zemun Polje (44°87′ N, 20°33′ E), Serbia, during the winter season of 2018/2019. In total, two experimental groups, consisting of 3 biological replicates with 5 larvae each, were formed for the pilot stage:

- Dnca – non-cold-acclimated diapausing group;
- Dca – cold-acclimated diapausing group.

The setup for the Pilot experiment is outlined in Figure 3. The collected larvae were first acclimated at 15°C for two weeks, after which one group – Dnca, was frozen in liquid nitrogen and stored at −80°C until protein analysis. The remaining larvae were placed in insect homes built from waxed cardboard and gradually chilled by lowering the temperature by 1°C each day, with additional acclimation for two weeks when specific checkpoint temperatures were reached (5, −3, and −16°C). During the entire acclimation process, larvae were kept under diapause-promoting light:dark conditions (12h L : 12h D). After acclimation to –16°C, larvae were frozen in liquid nitrogen and stored at −80°C until protein analysis (Dca experimental group).

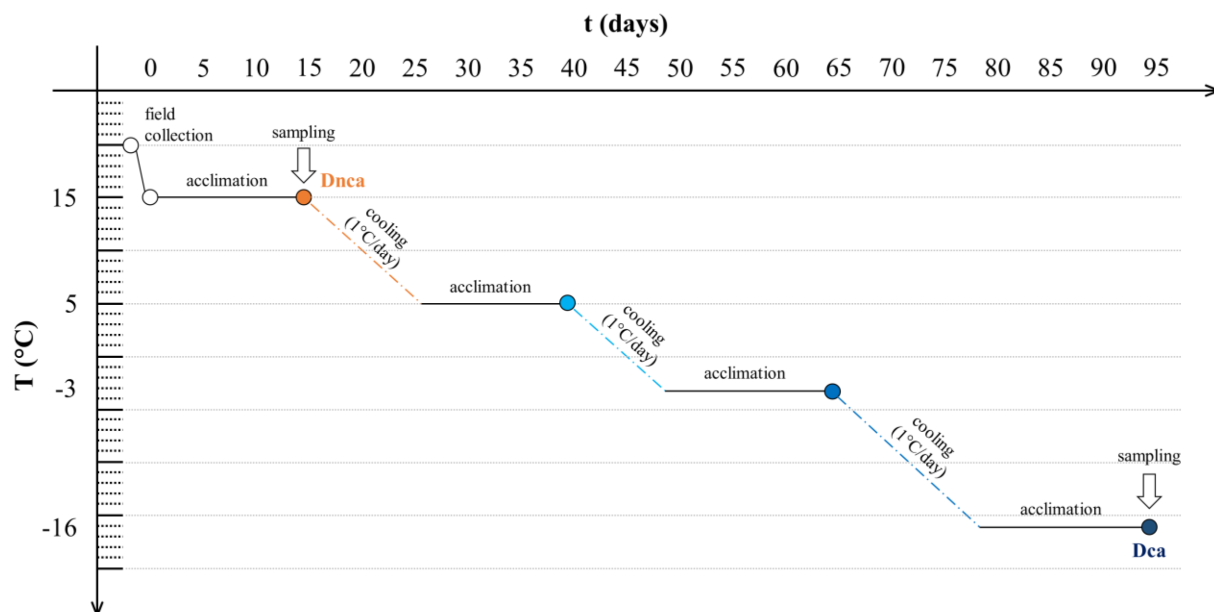

Figure 3. Experimental setup and cold acclimation regiment in the Pilot. Details in text.

3.1.2. Main experiment sampling and design

For the Main experiment, corn stalks infested with ECB larvae were provided by the Institute of Field and Vegetable Crops in Novi Sad, Serbia, during the winter season of 2019/2020. The infesting larvae were at the 5th instar stage and in early phases of diapause. Enough larvae were collected for cold acclimation experiments, while the rest were left to complete their life cycle and produce a generation of non-diapausing larvae.

The collected diapausing larvae were placed in sterilized glass jars with insect feed prepared according to the following recipe (ingredients and measurements give in Table 3):

- Added 1 L of cold water to wheat and barley kernels, mixed thoroughly and left to rest for 30 minutes or overnight;
- Dissolved agar in 120 mL of cold water with a magnetic stirrer and combined with the wheat and barley mixture;
- Added brewer's yeast, mixed thoroughly, brought to a boil while stirring occasionally, and then proceeded with constant stirring while the mixture cooked;
- Cooled the mixture to 70°C after sufficient thickness had been achieved and gradually added methylparaben, formaldehyde, citric and acetic acids, vitamin C and vitamin B complex;
- Cooled the mixture to 40–50°C after the preservatives and vitamins were added and poured into glass jars;
- Sterilized glass jars containing insect feed under UV light for 1 hour.

Table 3. Composition of insect feed used for maintaining diapausing and rearing non-diapausing larvae.

Component	Amount (g)	Component	Amount (g)
Wheat kernels	160.0	Formaldehyde 37%	6.0
Barley kernels	25.0	Citric acid	5.0
Agar	20.0	Acetic acid	1.4
Brewer's yeast	25.0	Vitamin C	5.0
Methylparaben	1.7	Vitamin B complex	0.07

The glass jars with feed and larvae were placed into incubators for gradual cold acclimation of the samples and forming of experimental groups. The same acclimation regiment and light:dark conditions from the Pilot setup were applied, with samples being frozen in liquid nitrogen and stored at –80°C after every acclimation checkpoint (Fig. 4):

- D(15) – diapausing larvae acclimated to 15°C for two weeks;
- D(5) – diapausing larvae acclimated to 5°C for two weeks;
- D(–3) – diapausing larvae acclimated to –3°C for two weeks;
- D(–16) – diapausing larvae acclimated to –16°C for two weeks.

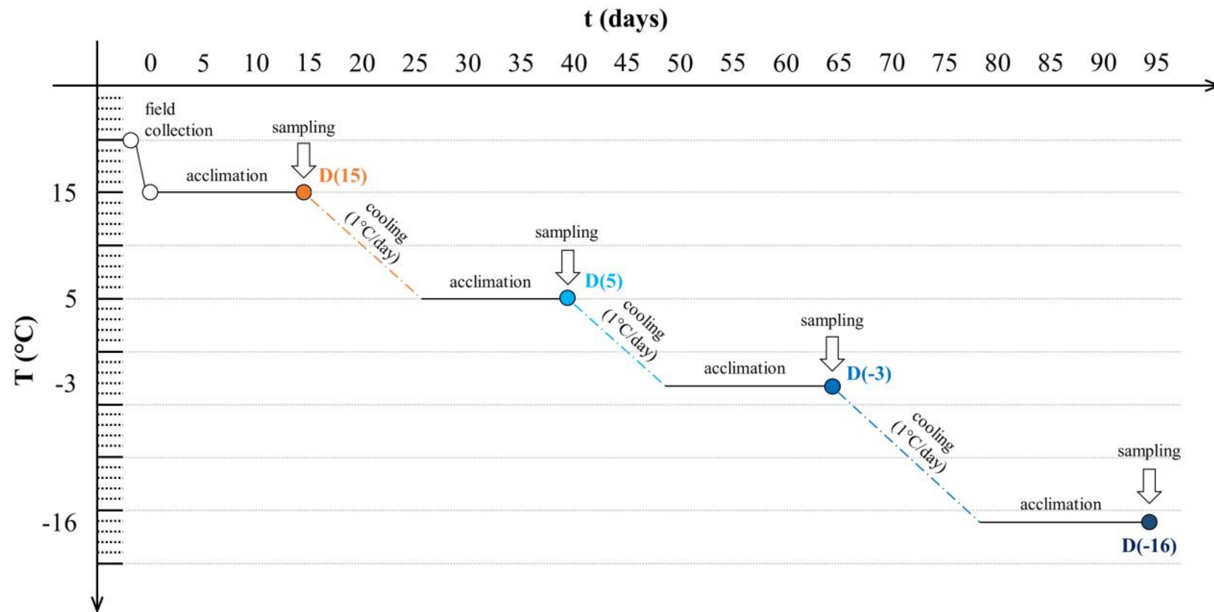

Figure 4. Setup and cold acclimation regiment in the Main Experiment. Details in text.

The larvae that remained in the corn stalks were allowed to terminate diapause and transform into pupae. The pupae were collected from the stalks during the summer season of 2020 and placed in cages covered with paper. Adult moths emerged from the pupae after metamorphosis, mated and laid eggs in batches on the paper covers, which were placed in glass jars with insect feed. Larvae that emerged from the eggs were reared under non-diapausing light conditions (18H L : 12H D) until they reached the 5th instar stage, when they were collected for the final experimental group – non-diapausing (ND) group. The larvae were frozen in liquid nitrogen and stored at –80°C until analysis.

Every experimental group consisted of 4 biological replicates with 3 larvae each:

- 1 biological replicate for protein analyses;
- 3 biological replicates for relative gene expression analysis.

3.2. Proteomic and bioinformatics analyses

3.2.1. Sample preparation for proteomic analyses

Whole-body larvae (5 per sample for the Pilot experiment and 3 per sample for the Main experiment) were homogenized in ice-cold 50 mM K-phosphate buffer pH 7.5 with 1 mM DTE using mortar and pestle to make a 20% w/v homogenate. The homogenates were then additionally lysed with sonication for 2 min (24 sonic pulses lasting 5 s each, with 10 s pauses in-between) using the Qsonica Q500 (Qsonica, Newtown, CT, USA). Sonication was followed by centrifugation for 10 min at 7000 rpm, 4°C to remove insoluble debris and lipids. Supernatants were divided into two microtubes per sample. One microtube from each sample was placed in a water bath at 100°C for 5 min in order to remove globular proteins and enrich the content of IDPs (heated sample type). The other microtube was left untreated for comparison (non-heated sample type). After the heat treatment, all samples were centrifuged for 10 min at 12 000 rpm, 4°C to further purify them and the supernatants were transferred to clean microtubes.

Total proteins were assayed on a microplate using the commercial Quick Start™ Bradford Protein Assay kit (Bio-Rad, Hercules, CA, USA, cat. no. 5000203). Samples and protein standards from the kit were pipetted onto the microplate (5 µL) and mixed with 250 µL of the staining reagent. Absorbances were measured on a microplate reader at 595 nm and used to construct a protein standard curve in Microsoft Excel and determine protein concentrations in the samples.

After the assay, a 5× concentrated protease inhibitor cocktail (Roche cOmplete™ ULTRA tablets, Merck KGaA, Darmstadt, Germany, cat. no. 5892970001) was added to the samples to a final concentration of 1×.

3.2.2. 2D PAGE enrichment validation

To validate the results of IDP enrichment, proteins from both untreated and heat-treated samples were separated using a modified in-house 2D PAGE method. Gels were cast and electrophoresis was run using the Mini-PROTEAN 3 system (Bio-Rad, Hercules, CA, USA). For the first dimension, a 10-well, 1 mm thick discontinuous native PAGE consisting of a 12.5% separating and 4% stacking gel was prepared according to the following recipe:

Table 4. Composition of gels prepared for first dimension PAGE separation.

Component	12.5% separating gel	4% stacking gel
dH_2O	2.0 mL	2.5 mL
1.5M Tris-HCl pH 8.8	1.5 mL	/
0.5M Tris-HCl pH 6.8	/	1.0 mL
30% acrylamide-bisacrylamide	2.5 mL	0.54 mL
10% APS	100 µL	75 µL
TEMED	10 µL	10 µL
Total (mL)	~6.1 mL	~4.9 mL

First, the separating gel was cast by pipetting ~4.5 mL of the separating gel mixture between the assembled glass plates and overlaying it with dH_2O until it polymerized after ~20 minutes. Next, the water was discarded and the assembly was filled to the top with the stacking gel mixture. A 10-well comb was placed inside the stacking gel and removed after polymerization, which took ~20 minutes. The finished discontinuous gel was assembled together with the electrophoresis module and placed in a buffer tank filled with Tris-glycine pH 8.3 running buffer.

Samples were prepared by taking aliquots from the heated and non-heated sample types and pooling them in two mixtures, respectively. The mixtures were loaded onto the gel with 2X native sample buffer (0.5M Tris-HCl pH 6.8, glycerol, 0.5% bromophenol blue). In total, wells were loaded with 20 µL containing 20 µg of protein. All remaining empty wells were filled with a mix of dH_2O and sample buffer to ensure even distribution of the current throughout the gel. Electrophoresis was run for 50 min at 180 V. After the separation, individual lanes were cut out as strips and placed in 10 mL of 1.5 M Tris-HCl pH 8.8 containing 8M urea for 45 min to solubilize the proteins. The separating gel (12.5%) for the second dimension was prepared according to the recipe outlined in Table 5 below:

Table 5. Composition of second dimension PAGE separating gel.

Component	12.5% separating gel
dH_2O	0.75 mL
1.5M Tris-HCl pH 8.8	1.5 mL
30% acrylamide-bisacrylamide	2.5 mL
Urea (final conc. 8M)	2.88 g
10% APS	50 µL
TEMED	5 µL
Total (mL)	~6.0 mL

The total volume of the mixture ended up being ~6 mL after the urea dissolved completely. Less APS and TEMED were added compared to the previous separating gel as the urea also makes the gel polymerize much quicker. Additionally, a spacer was placed between the assembled glass plates in order to reduce the width of the gel, compared to its default dimensions. The separating gel containing 8M urea was cast by pipetting ~4.5 mL of the separating gel mixture between the assembled glass plates and overlaying it with dH_2O until it polymerized after ~20 minutes. After casting the gel, the water was discarded and a strip with solubilized proteins was placed on top of it instead of a stacking gel, making sure not to introduce any bubbles between the separating gel and the strip. The spacer that was placed beforehand ensured that the width of the separating gel corresponded to the length of the strip. A small volume of 4% separating gel mixture was added to fix the strip onto the separating gel. The second dimension was run for 30 min at 400 V with Tris-glycine pH 8.3 as the running buffer. Following the separation, protein spots were visualized by staining the gels with the Pierce™ Silver Stain Kit (Thermo Scientific, Waltham, MA, USA, cat. no. 24612), according to the manufacturer's instructions.

3.2.3. LC-MS/MS protein identification

Protein identification was done using shotgun LC–MS/MS and the Mascot search engine. Up to 20 µg of total protein from whole-body homogenates were double-digested in-solution using Trypsin/Lys-C mixture, followed by Trypsin digestion.

Samples were first prepared for digestion in Microcon-10 kDa centrifugal filters according to the following steps:

1. Added 200 µL of LC-MS grade H_2O to the filter and centrifuged at 13 500 g, 4°C for 10 min to rinse it, with ~30 µL of water remaining in the filter after the rinse; discarded the eluate from the outer vial;
2. Added a solution containing up to 20 µg of protein to the filter and topped up to 200 µL with 200 mM NH_5CO_3, centrifuged at 13 500 g, 4°C for 10 min; discarded the eluate from the outer vial;
3. Added 200 µL of 200 mM NH_4CO_3, centrifuged at 13 500 g, 4°C for 10 min; discarded the eluate from the outer vial;
4. Added 200 µL of 50 mM NH_4CO_3, then centrifuged at 13 500 g, 4°C for 10 min; discarded the eluate from the outer vial;

5. Placed the filter upside-down in a new outer vial and centrifuged at 1000 g for 2 min to transfer the protein solution from the filter to the outer vial, then pipetted the solution into a 0.5 mL Lo-Bind Eppendorf tube.

After the preparation steps, protein samples were digested according to the following protocol:

1. Added 1.5 µL of LC-MS MeOH to the protein sample for a final MeOH concentration of 5%;
2. Added 5 µL of 0.5% Rapigest and 2 µL of 200 mM DTT to the protein sample and incubated at 60°C for 30 min;
3. Cooled the sample to room temperature, then added 5 µL of 200 mM NH_5CO_3 and 2.5 µL of 200 mM iodoacetamide;
4. Incubated the sample in the dark for 30 min at room temperature;
5. Added 1 µL of stock Trypsin/Lys-C Mix (20 µg of the mixture in 80 µL of LC–MS grade H_2O) to the sample and incubated at 37°C for 1 h;
6. Added trypsin in a 1:25 trypsin:protein ratio at 37°C for 1 h;
7. Terminated the digestion by adding 1.5 µL of formic acid for a final concentration of at least 2% *v/v*;
8. Completely dried the sample in a vacuum dryer at 50°C.

After the digestion, samples were desalted and cleaned up using Pierce C 18 Spin Columns placed in 2 mL Lo-Bind Eppendorf tubes according to the following steps:

1. Added 200 µL of 50% MeOH to the column and centrifuged at 1500 rpm for 1 min. This step was repeated once more and the eluate discarded after;
2. Added 200 µL of 0.5% TFA, 5% ACN solution to the column and centrifuged at 1500 rpm for 1 min. This step was repeated once more and the eluate discarded after;
3. Added 200 µL of 0.1% TFA to the column and centrifuged at 1500 rpm for 1 min. This step was repeated once more and the eluate discarded after;
4. Dissolved the dried sample in 50 µL of 0.1% TFA and applied to the column, then centrifuged at 1500 rpm for 1 min;
5. Collected the eluate and reapplied to the column, then centrifuged at 1500 rpm for 1 min;
6. Added 100 µL of 0.1% TFA to the column and centrifuged at 1500 rpm for 1 min. This step was repeated once more;

7. Placed the column was in a new 2 mL Lo-Bind Eppendorf tube;
8. Added 50 µL of 0.1% TFA, 70% ACN solution to the column and centrifuged at 1500 rpm for 1 min to elute the sample. This step was repeated once more;
9. Completely dried the sample in a vacuum dryer at 50°C and stored in at –20°C until analysis.

Tryptic digests were subjected to nano-LC–MS/MS analysis using a Dionex Ultimate 3000 RSLC nanoLC system (Dionex, Sunnyvale, CA, USA) coupled to Bruker Maxis II ETD Q-TOF instrument (Bremen, Germany) via a CaptiveSpray nanoBooster ionization source. Peptides were separated online using Acquity M-Class BEH130 C18 analytical column (1.7 µm, 130 Å, 75 µm × 250 mm Waters, Milford, MA, USA) following trapping on an Acclaim PepMap 100 C-18 trap column (5 µm, 100 Å, 100 µm × 20 mm, Thermo Fisher Scientific, Waltham, MA, USA). The temperature was set at 48°C, and a flow rate of 300 nl/min was applied. The gradient method was from 4% B to 50% B in 90 min; solvent A was 0.1% formic acid in water, whilst solvent B was 0.1% formic acid in acetonitrile. Sample ionization was achieved in the positive electrospray ionization mode. Data-dependent analysis was performed using a fixed cycle time of 2.5 s. MS spectra were acquired over a mass range of 150–2200 *m/z* at 3 Hz, while CID was performed at 16 Hz for abundant precursors and at 4 Hz for ones of low abundance.

Data were evaluated with ProteinScape 3.0 software (Bruker Daltonic GmbH, Bremen, Germany) using the Mascot search engine version 2.5.1 (Matrix Science, London, UK). Considering the limited availability of *O. nubilalis* protein sequences in published databases, MS/MS spectra were also searched against *O. nubilalis*, *O. furnacalis* (a species closely related to the ECB), as well as all lepidopteran protein sequences available in the NCBI database. The following parameters were applied: trypsin as enzyme, 7 ppm peptide mass tolerance, 0.05 Da fragment mass tolerance, and 2 missed cleavages. Carbamidomethylation on cysteines was set as a fixed modification, while deamidation (NQ) and oxidation (M) were set as variable modifications.

3.2.4. Bioinformatic analyses

After protein identification, FASTA sequences for all identified proteins were downloaded from the NCBI database and used for the prediction of structural disorder. The structural disorder of proteins was determined with the IUPred long disorder predictor (https://iupred3.elte.hu/). The predictor is based on estimating the total pair-wise inter-residue interaction energy gained when a

polypeptide chain folds. An amino acid is considered to be disordered if its IUPred score is at least 0.5. Mean disorder was computed as the average of residue scores, which range from 0.0 to 1.0. Overall disorder rate (percental disorder, ranging from 0 to 100%) represents the fraction of disordered amino acids in a polypeptide chain. Proteins are considered globular if their overall disorder rate is below 10%; nearly ordered if the rate is between 10% and 30%; partially disordered if the rate is between 30% and 70%; (mostly) disordered if the rate is above 70%. All proteins were further analyzed for the presence of long intrinsically disordered regions (long IDRs) — sequences of at least 20 consecutive disordered amino acids. Additionally, the amino acid composition of the proteins was analyzed to determine the absolute number of individual amino acids that make up each polypeptide, as well as their ratios.

Lastly, functional characterization was performed on the identified sequences. Functional information on the identified proteins was collected from various databases such as UniProt (www.uniprot.org/, Consortium, 2021), Pfam (http://pfam.xfam.org/, Mistry et al., 2021), Interpro (https://www.ebi.ac.uk/interpro/, Blum et al., 2021), and GeneOntology (http://geneontology.org/, Ashburner et al., 2000). Data on their molecular functions, cellular localization, the biological processes they are involved in, and the do-mains they contained was collected. All of the analyses were performed using homemade PERL scripts run locally.

3.3. Gene expression analysis of selected IDPs

3.3.1. Isolation of total RNA

Total RNA was isolated from whole bodies of larvae from the Main experiment. Each of the five experimental groups was comprised of three biological pools and each pool contained three larvae. RNA isolation was performed using the TRIzol Reagent (Invitrogen, Waltham, MA, USA, cat. no. 15596026) and mortar and pestle. Microtubes (2 mL) containing 1.5 mL of TRIzol Reagent were prepared in advance. The mortar and pestle were placed at -20°C for chilling before sample homogenization. Larvae were placed into the mortar and liquid nitrogen was poured over to freeze them and turn them brittle. After the samples were pulverized into fine dust, they were transferred into the prepared microtubes for the next RNA isolation steps. The mortar and pestle were wiped with 70% EtOH between each sample. Next, the steps outlined in the manufacturer's protocol for TRIzol Reagent were followed:

1. Centrifuged for 5 minutes at 12 000 g and 4°C, transfered the supernatants to new 2 mL microtubes due to the high fat content of the larvae;
2. Incubated for 5 minutes and added 0.5 mL of chloroform;
3. Gently inverted the microtubes for 3 minutes in order to mix the contents;
4. Centrifuged for 15 minutes at 12 000 g and 4°C to separate the phases;
5. Transferred the top, aqueous phase containing total RNA to new 1,5 mL microtubes;
6. Added 1 mL of ice cold isopropanol to precipitate the RNA;
7. Incubated for 10 minutes, then centrifuged for another 10 minutes at 16 000 g and 4°C;
8. Discarded the supernatant and resuspended the RNA pellet in 1.5 mL of 75% ethanol for salt cleanup;
9. Centrifuged for 10 minutes at 16 000 g and 4°C to precipitate the RNA again;
10. Discarded the supernatant and repeated steps 8 through 9 two more times;
11. Incubated the microtubes with purified RNA pellets for 10 minutes at 60°C in a TSC ThermoShaker (Biometra, Jena, Germany) to evaporate residual ethanol;
12. Added 40 µL of DEPC-treated H_2O to the RNA pellets and incubated on the thermoshaker for 10 minutes at 60°C to improve the dissolving of RNA.

3.3.2. RNA concentration and quality assessment

Total RNA concentration and purity were assessed using the BioSpec-nano spectrophotometer (Shimadzu, Kyoto, Japan). Absorbance was measured at 230, 260 and 280 nm. The 260 nm wavelength absorbance corresponded to the concentration of the RNA, and the instrument calculated the results according to the following formula:

$$\mathrm{RNA}\left[\frac{\mu g}{\mu L}\right] = A_{260}\ x\,\mathrm{R}\ x\,\mathrm{F}$$

where:

A260 – the absorbance measured at the 260 nm wavelength,

R – sample dilution and

F – 40, indicating 40 µg/µL of purified RNA when OD (optical density) equals 1.

The 260/230 nm and 260/280 nm absorbance ratios indicated how pure the samples were from salts and proteins/phenol, respectively.

To determine the integrity of the isolated RNA, samples were run on 1.5% agarose gel. The gel was prepared as follows:

- Dissolved 1.5 g of agarose (VWR Peqlab, Erlangen, Germany, cat. no. 732-2788P) in 100 mL of tris-acetate-EDTA (TAE) buffer and heated until the solution cleared up;
- Cooled the solution and added 3 µL of GelRed Nucleic Acid Stain (Biotium, Fremont, CA, USA, cat. no. #41003), swirling to mix thoroughly;
- Poured the gel in the cast, placed the well comb and let the gel harden;
- Removed the well comb and rinsed the gel with TAE buffer to remove excess stain;
- Placed the gel in the electrophoresis tank and equilibrated in running buffer for 30 minutes.

The samples were mixed with 10X Blue Juice Gel Loading Buffer (Thermo Fisher Scientific, Waltham, MA, USA, cat. no. 10816015) and loaded onto the gel (1 µL RNA, 1 µL loading buffer, 8 µL DEPC-treated H_2O). Electrophoresis was run on a Biometra Compact M system (Biometra, Jena, Germany) at 90V for 45 minutes until the dye reached the halfway point on the gel. After the run, the gel was placed in the ChemiDoc XRS+ System (Bio-Rad, Hercules, CA, USA) to image 28S and 18S RNA bands and measure their intensities. The 28S/18S intensity ratio was used to

calculate the RNA integrity number (RIN) which shows whether the isolated RNA was intact or degraded.

3.3.3. RT-PCR cDNA synthesis

Complementary DNA for qPCR analysis was synthesized using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Waltham, MA, USA, cat. no. 4368814). The standard reaction mixture of 20 µL was up-scaled to 40 µL to produce a higher amount of cDNA. Complementary DNA was synthesized from 4 µg of total RNA from each sample. The reverse transcription process was carried out on an Eppendorf Mastercycler EP Gradient S (Eppendorf, Germany). Reaction mixtures were prepared as follows:

Table 6. Composition of individual reaction mixtures for RT-PCR cDNA synthesis.

Component	Volume
10X RT Buffer	4.0 µL
25X dNTP Mix (100 mM)	1.6 µL
10X RT Random Primers	4.0 µL
Multiscribe™ Reverse Transcriptase	2.0 µL
DEPC-treated H_2O	8.4 µL
RNA Template (200 ng/µL)	20.0 µL
Total per reaction	40.0 µL

Next, the mixtures were pipetted into 200 µL PCR microtubes, vortexed briefly and centrifuged to spin down the contents. The prepared microtubes were placed in the thermal cycler and reverse transcription was carried out according to the following amplification steps:

Table 7. Thermal cycler settings for RT-PCR cDNA synthesis.

Step	Settings	
	Temperature (°C)	Time (min)
1	25	10
2	37	120
3	85	5
4	4	Hold

After reverse transcription, the synthesized cDNA was diluted 8 times with nuclease-free H_2O, bringing the concentration to 12.5 ng/µL, and stored at –20°C until further analysis.

3.3.4. Primer design

Genes to be analyzed were selected according to the intrinsic disorder content of proteins that are coded by them, as well as their functions. Actin and ribosomal protein s03 genes were selected as

reference genes due to their stable expression levels. The other selected genes of interest cover different function groups, such as regulation of muscle contraction and cytoskeleton formation, immune response, signal transduction and protein metabolism. Due to the limited availability of *O. nubilalis* nucleotide sequences in online databases, most primers were designed using sequences belonging to the closely related *O. furnacalis*. Nucleotide sequences were collected from the NCBI database (National Centre for Biotechnology Information) and primers designed using the Primer-BLAST tool (Ye et al., 2012). During the design, *in silico* analysis was also performed to check whether the primers would recognize non-specific targets.

The following sequences were available for *O. nubilalis*. Their short names and NCBI accession numbers are included:

1. *actin* (EL928709.1),
2. *ribosomal protein s03, rps3* (EL929086.1),
3. *heat shock protein 90, hsp90* (EL929806.1),
4. *small heat shock protein 20.4, hsp20.4* (AB568467.1) and
5. *small heat shock protein 20.1, hsp20.1* (AB568468.1).

The following sequences were available for *O. furnacalis*. Their short names and NCBI accession numbers are included:

1. *heat shock cognate 70 kDa protein, hsc70* (JF708084.1)
2. *heat shock protein 68-like, hsp70* (XM_028309302.1),
3. *troponin T, tnt* (XM_028316365),
4. *thymosin beta, thym* (XM_028305412),
5. *tropomyosin-2, tropmy2* (XM_028309660) and
6. *moesin/ezrin/radixin homolog 1, moe* (XM_028309570).

Primer sequences can be found in Table 27 of Supplementary Materials. The primer pairs were ordered from and synthesized by Vivogen LLC (Belgrade, Serbia).

Primers were delivered in lyophilized form and centrifuged for 30 seconds at 8000 rpm to spin down the contents. Stock solutions (100 µM) were prepared as follows:

- Dissolved lyophilisates in 10 times the synthesized amount of primers (25 nmol) with DEPC-treated H_2O;
- Vortexed the microtubes to mix the contents;

- Spun down microtubes briefly and placed on thermoshaker for 10 minutes at 60°C;
- Centrifuged for 30 seconds at 8000 rpm to spin down condensates.

Working solutions (10 µM) were prepared by diluting the stock solutions 10 times with DEPC-treated H_2O.

3.3.5. Primer specificity analysis

The specificity of primer pairs derived from *O. furnacalis* sequences was additionally confirmed by non-quantitative PCR and running the products on an agarose gel. To minimize potential amplification differences that can exist between the samples and ensure good signal strength, aliquots of synthesized cDNA were taken from every sample, mixed together and used as the template for the amplification reactions. The same instrument that was used for cDNA synthesis, the Eppendorf Mastercycler EP Gradient S, was employed here as well.

Single reaction mixtures were prepared with the 2X GoTaq Hot Start Green Master Mix (Promega, Madison, WI, USA, cat. no. M7422) as follows:

Table 8. Composition of individual reaction mixtures for primer specificity analysis.

Component	Volume
2X Master Mix	7.0 µL
F primer (10 µM)	0.7 µL
R primer (10 µM)	0.7 µL
DEPC-treated H_2O	0.6 µL
cDNA mixture (~12.5 ng/µL)	5.0 µL
Total per reaction	14.0 µL

Next, the mixtures were pipetted into 200 µL PCR microtubes, vortexed briefly and centrifuged to spin down the contents. The prepared microtubes were placed in the thermal cycler and PCR amplification was carried out according to the following steps:

Table 9. Thermal cycler settings for primer specificity analysis.

Step	Settings		Comments
	Temperature (°C)	Time (min:sec)	
1	95	10:00	Initial denaturation
2	95	00:30	Repeat steps 2–4 40 times
3	60	00:30	
4	72	00:30	
5	72	07:00	Final elongation

After amplification, the samples were loaded onto 0.8% agarose gel to check for products. The gel was prepared as follows:

- Dissolved 0.2 g of agarose (VWR Peqlab, Erlangen, Germany, cat. no. 732-2788P) in 25 mL of tris-borate-EDTA (TBE) buffer and heated the solution cleared up;
- Cooled the solution and added 1 µL of GelRed Nucleic Acid Stain (Biotium, Fremont, CA, USA, cat. no. 41003), swirling to mix thoroughly;
- Poured the gel in the cast, placed the well comb and let the gel harden;
- Removed the well comb and rinsed the gel with TAE buffer to remove excess stain;
- Placed the gel in the electrophoresis tank and equilibrated in running buffer for 30 minutes.

The entire volume of samples from the PCR microtubes (14 µL) was loaded onto the gel and electrophoresis was run on a Mini-Sub Cell GT Horizontal Electrophoresis System (Bio-Rad, Hercules, CA, USA) at 90V for 45 minutes until the dye reached the halfway point on the gel. After the run the gel was placed under a UV light to visualize the bands.

3.3.6. Primer efficiency determination

Amplification efficiency of the primer pairs was determined with quantitative PCR on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). The same cDNA mixture that was prepared for the analysis of primer specificity was used as template to create 10-fold serial dilutions for the qPCR analysis – 1:1, 1:10, 1:100, 1:1000 and 1:10000.

Single reaction mixtures were prepared with the 2X Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA, cat. no. K0221) according to the steps outlined in **Primer specificity analysis**. All samples were run as technical duplicates.

The following amplification protocol was used:

Table 10. Thermal cycler settings for primer efficiency determination.

Step	Settings		Comments
	Temperature (°C)	Time (min:sec)	
1	95	10:00	Initial denaturation
2	95	00:30	Repeat steps 2–4 40 times
3	60	00:30	
4	72	00:30	
5	65–95	00:05	Melting curve analysis

Melting curve analysis was performed to confirm the specificity of the primer pairs at this step as well. Product melting curves were recorded in the temperature range of 65 to 95°C, while the rate of temperature increase was set to 0.5°C every 5 seconds. After amplification, the recorded cycle threshold (Ct) values were plotted in Microsoft Excel against logarithmized cDNA dilutions and the slope of the graph line was used to calculate primer pair efficiencies according to the following equation:

$$\text{Efficiency (\%)} = \left(10^{\frac{-1}{slope}} - 1\right) x\ 100$$

3.3.7. Quantitative PCR of genes of interest

Relative expression of selected genes of interest was determined using *actin* and *ribosomal protein s03* as reference genes and non-diapausing (ND) group as the control.

Single reaction mixtures were prepared as outlined in **Primer efficiency determination**, and the same detection system and amplification protocol were used as well. Samples were pipetted in technical duplicates onto 96-well microplates, sealed with cover and centrifuged for 2 minutes at 4000 rpm to spin down the contents prior to placement in the detection system.

After amplification, relative expression of analyzed genes was calculated according to the methodology described in Ganger et. al. (2017), which is derived from Livak and Schmittgen (2001) and Pfaffl (2001) methods. In this methodology, the obtained Ct values for reference and genes of interest are adjusted with logarithmized primer efficiency values ($\log_{10}E$), which allows all further calculations to be performed on the log scale. Next, relative expressions of genes of interest (Goi) are normalized by subtracting their Ct values from the Ct values of reference genes (Ref1 and Ref2). The final results are expressed in ΔCt values:

$$\Delta \text{Ct} = \frac{\log_{10}\text{E}_{\text{Ref1}}\ x\ \text{Ct}_{\text{Ref1}} + \log_{10}\text{E}_{\text{Ref2}}\ x\ \text{Ct}_{\text{Ref2}}}{2} - \log_{10}\text{E}_{\text{Goi}}\ x\ \text{Ct}_{\text{Goi}}$$

3.3.8. Statistical analysis of qPCR results

Results of quantitative PCR were statistically analyzed using the Statistica version 14.0 software (StatSoft, Inc., Tulsa, OK, USA). The ΔCt values of genes from all experimental groups were compared to determine whether there were differences in gene expression between them. First, normal distribution and equal group variance of the data were confirmed with Brown-Forsythe test. Next, the statistical significance of the differences was tested with one-way analysis of variance (one-way ANOVA) and *post hoc* Tuckey analysis for significance level $p < 0.05$. The results are presented as univariate scatterplots, which is recommended for studies with small sample sizes (Weissgerber et al., 2015).

4. Results

4.1. Proteomic analyses

4.1.1. 2D PAGE confirmation of IDP enrichment procedure

A specific two-dimensional electrophoretic assay was performed on samples from Pilot experimental groups, in order to detect the presence of IDPs following their enrichment and isolation. This assay is based on the heat-stability and resistance to chemical denaturation of IDPs, resulting in a pattern where disordered proteins align in a diagonal line in the second dimension of electrophoretic separation. As seen in Figure 5A, proteins from the non-heated sample, which was rich in globular polypeptides, were mostly retained in the first dimension and did not transfer into the second gel. In fact, the proteins did not migrate far during the separation in the first dimension. This was likely due to the presence of high molecular weight arylphorins, preventing other proteins from separating. On the other hand, a large proportion of the proteins from the heat-treated sample (Fig. 5B) were aligned on the diagonal line, signifying that they are mostly or fully disordered. Heat-stable globular proteins are generally found above the diagonal line, as indicated by the arrows. Thus, heating of the samples and removal of globular proteins allowed the proteins to separate in the first dimension and transfer into the second gel.

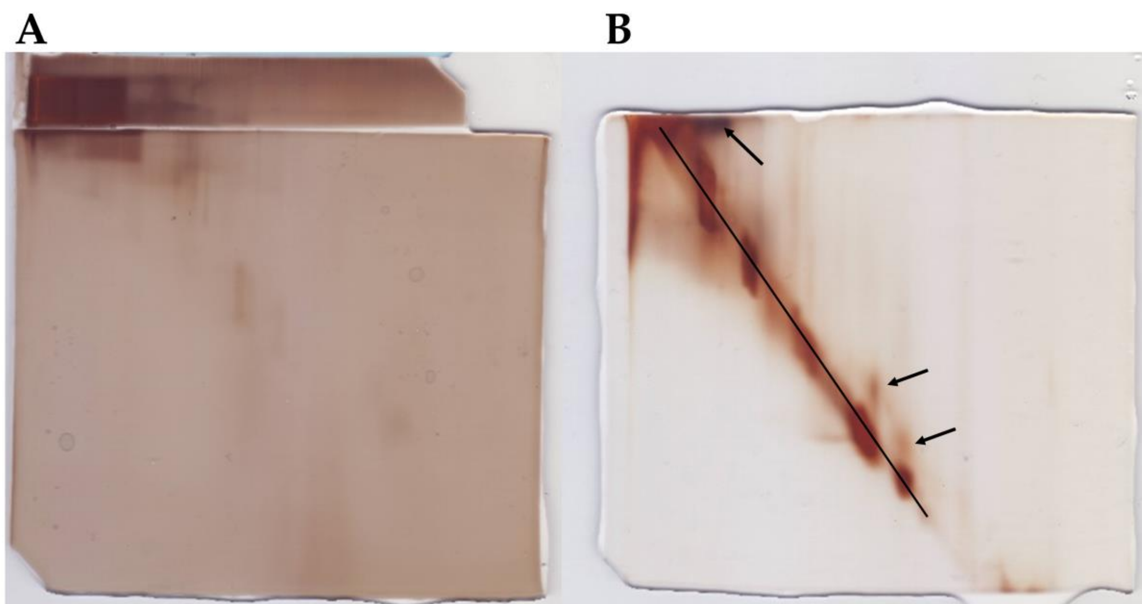

Figure 5. Results of in-house 2D PAGE for detecting intrinsically disordered proteins. (A) Proteins from non-heated samples are mostly locked in the gel from the first dimension (strip overlaying the larger gel) (B). Proteins from heat-treated samples have successfully entered the second dimension. The black line represents the diagonal along which IDPs are located. Arrows denote ordered proteins that stay above the diagonal.

4.1.2. LC-MS/MS identification of proteins

Following the enrichment validation by 2D PAGE, protein identification from individual homogenates of both experimental setups could be proceeded with. In total, 2103 proteins were identified – 820 from the Pilot setup and 1283 from the Main setup.

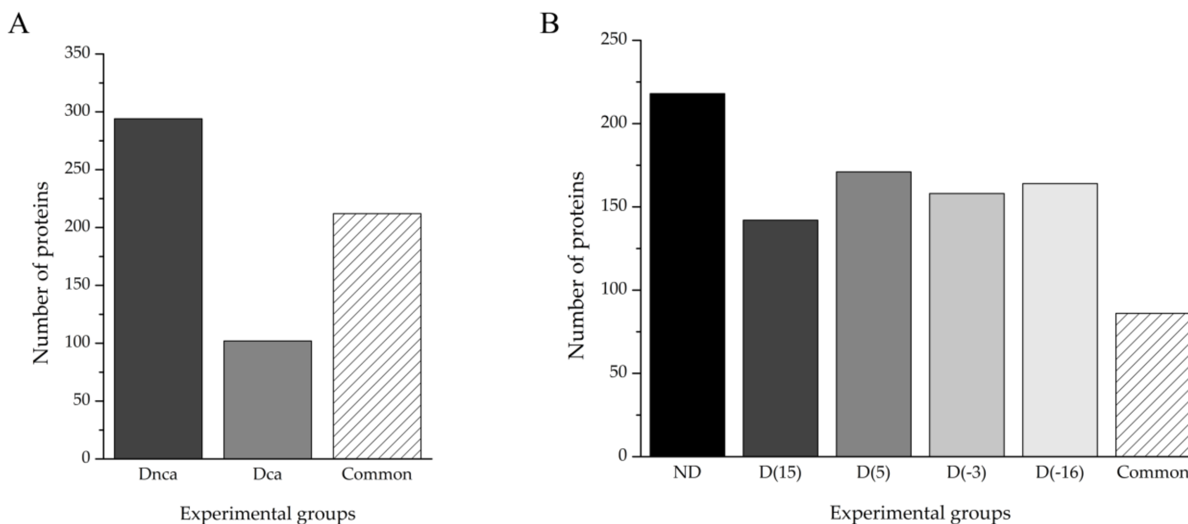

Figure 6. Total unique and common proteins isolated from the (A) Pilot and (B) Main experimental setups. Common – proteins that were found in all samples of Pilot or Main experimental setup (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

In the Pilot experimental setup, after accounting for shared entries between the two experimental groups, there were 608 unique proteins in total, with 47% of hits (290) being linked to polypeptides that have only been predicted from nucleotide sequences. Out of the total number of proteins, 294 were present only in the Dnca experimental group, with 102 only in the Dca experimental group; the remaining 212 proteins were found in both (Fig. 6A).

In the Main experimental setup, after accounting for shared entries between the five experimental groups, there were 600 unique proteins in total, with 45% of hits (268) being linked to polypeptides that have only been predicted from nucleotide sequences. Out of the total number of proteins, 218 were present only in the ND experimental group, 142 only in the D(15) experimental group, 171 only in the D(5) experimental group, 158 only in the D(–3) experimental group and 164 only in

the D(–16) experimental group; the remaining 86 proteins were found in all five experimental groups (Fig. 6B).

When accounting for duplicate entries between all groups in both experimental setups, there were a total of 892 unique proteins – 292 were identified only in groups from the Pilot setup, 284 were identified only in groups from the Main setup and 316 proteins were identified in groups from both experimental setups.

The breakdown of the number of unique proteins present in each experimental group, as well as of common proteins in the two different experimental setups, is presented in Table 11 below:

Table 11. Total, unique and shared proteins identified in groups of Pilot and Main experimental setups.

Experimental setup	Experimental group	Total number of proteins	Unique proteins	Common proteins
Pilot	Dnca	506	294	212
	Dca	314	102	
Main	ND	304	218	86
	D(15)	228	142	
	D(5)	257	171	
	D(–3)	244	158	
	D(–16)	250	164	

Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks.

4.1.3. Effect of sample heating on protein identification

A comparison of total identified proteins was made between the heat-treated and untreated samples of all experimental groups in both setups. Heating the samples resulted in the identification of additional unique proteins in all groups, when compared with the non-heated samples.

In the Pilot setup (Fig. 7A), an additional 180 unique proteins were uncovered in the Dnca group after heating the samples, while 265 heat-sensitive proteins were eliminated. The two sample types had 61 proteins in common. Within the Dca group, 96 proteins were found only in the heated sample, 184 in the non-heated samples, and 34 proteins were shared between the two sample types

When it comes to the Main setup, sample heating lead to an unexpected outcome in the ND group (Fig. 7B). Almost the same number of unique proteins were identified in the non-heated and heated ND samples (138 and 129, respectively). The two sample types also had 37 proteins in common. The highest amount of proteins were identified in the non-heated samples of D(5) and D(–3) experimental groups – 165 and 151, respectively. With the exception of the ND group, the heated samples contained significantly less unique proteins, and their amounts distribution between the groups is similar.

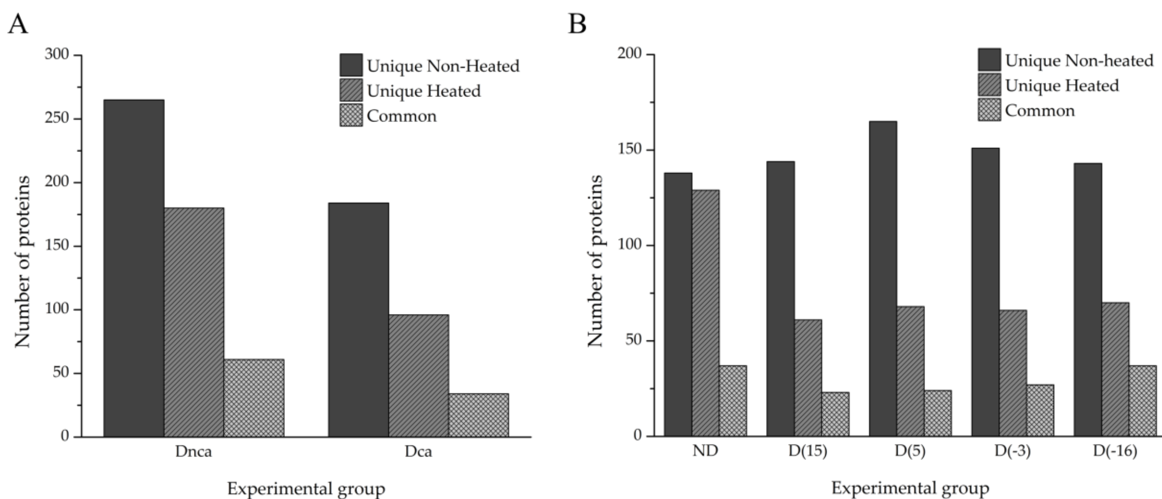

Figure 7. Effect of sample heating on the number of identified proteins in the (A) Pilot and (B) Main experimental setups. Unique Non-heated — proteins found only in the non-heated samples; Unique Heated — proteins found only in heated samples; Common — proteins that were found in both heated and non-heated samples (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

The breakdown of the number of unique and common proteins per sample type and experimental group is presented in Table 12 below:

Table 12. Total numbers of unique and shared proteins identified in different sample types of Pilot and Main experimental setup groups.

Experimental setup	Experimental group	Unique Non-heated	Unique Heated	Common
Pilot	Dnca	265	180	61
	Dca	184	96	34
Main	ND	138	129	37
	D(15)	144	61	23
	D(5)	165	68	24
	D(–3)	151	66	27
	D(–16)	143	70	37

Unique Non-heated — proteins found only in the non-heated samples; Unique Heated — proteins found only in heated samples; Common — proteins that were found in both heated and non-heated samples (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

4.2. Bioinformatic analyses

4.2.1. Determination of protein disorder

In order to determine the extent of structural disorder present in proteins identified in this study, percental disorder was calculated for the total proteins from every individual experimental group. Proteins with an average percental disorder of 70% or higher were considered as mostly disordered (MDPs), with partially disordered proteins (PDPs) if the average percental disorder was between 30% and 70%, nearly ordered (NOPs) for values between 10% and 30%, and ordered (OPs) if the percental value was no higher than 10%.

In the Dnca group of the Pilot setup, MDPs accounted for 31 of all identified proteins; 81 were PDPs, 75 were NOPs, and the remaining 319 were OPs. In the Dca group, 16 proteins were MDPs, 51 were PDPs, and 45 were NOPs; there were 198 OPs (Fig. 8A).

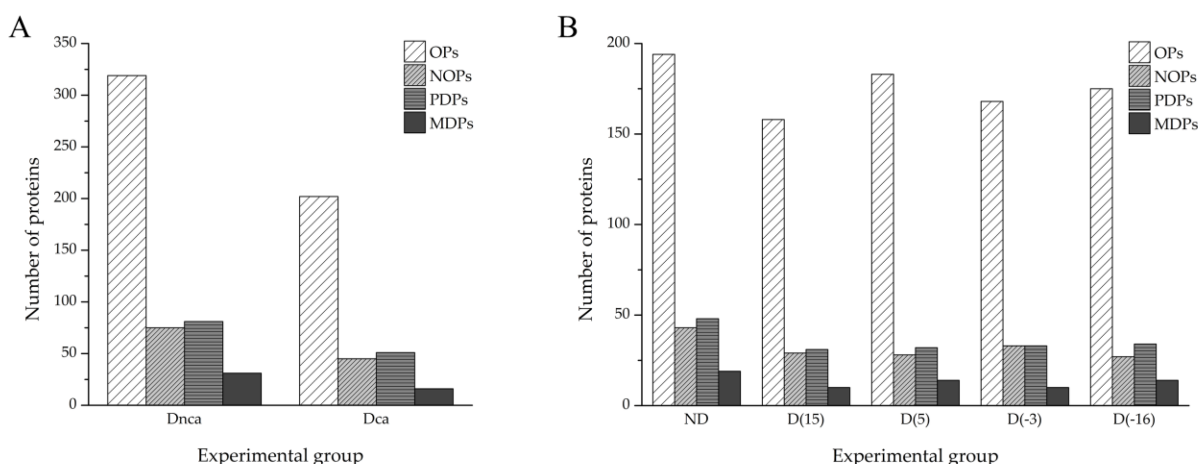

Figure 8. Total number of proteins with varying degrees of intrinsic disorder in the Pilot (A) and Main (B) experimental setups. OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

When it comes to the Main setup (Fig. 8B), the ND group had the most number of proteins identified across all four disorder content categories. In this group, OPs accounted for 194 hits, NOPs and PDPs were similar in number (43 and 48, respectively), while 19 proteins were MDPs. The individual diapausing groups generally differ in the amount of OPs that were identified in them.

Out of all those groups, the highest number of OPs were identified in the D(5) and D(–16) groups – 183 and 175, respectively. In terms of MDPs, PDPs and NOPs, similar numbers of these proteins were identified between the diapausing groups.

Percental distribution of proteins with varying degrees of intrinsic disorder was also assessed in every group of the two experimental setups. Between the two groups of the Pilot setup, the percental distributions are almost identical – 63% OPs, 15% NOPs, 16% PDPs and 6% MDPs in the Dnca group versus 64% OPs, 14% NOPs, 16% PDPs and 5% MDPs in the Dca group (Fig. 9A). In the Main setup, the main difference is in the distribution of OPs. It is slightly lower in the ND group compared to the diapausing groups on average – 64% versus 69.8%, respectively, while the other protein groups are of similar distribution (Fig. 9B).

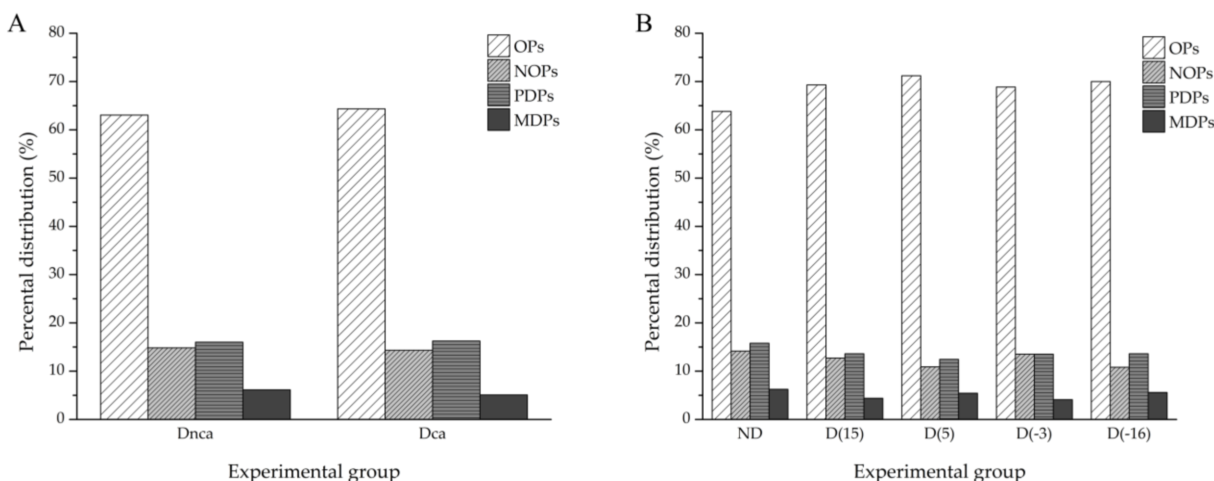

Figure 9. Percental distribution of proteins with varying degrees of intrinsic disorder in the Pilot (A) and Main (B) experimental setups. OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

The breakdown of the number of proteins with varying degrees of intrinsic disorder in different experimental groups is presented in Table 13 below:

Table 13. Total numbers of proteins with varying degrees of intrinsic disorder identified in different experimental groups of Pilot and Main setups.

Experimental setup	Experimental group	OPs	NOPs	PDPs	MDPs
Pilot	Dnca	319	75	81	31
	Dca	202	45	51	16
Main	ND	194	43	48	19
	D(15)	158	29	31	10
	D(5)	183	28	32	14
	D(–3)	168	33	33	10
	D(–16)	175	27	34	14

OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

4.2.2. Effect of sample heating on protein disorder distribution

The heat treatment had a profound effect on the distribution of proteins with various degrees of intrinsic disorder in the groups of both Pilot and Main experimental setups. The heat-treated samples contained more partially and mostly disordered proteins compared to the non-heated samples, while still retaining a significant portion of heat-resistant ordered proteins.

4.2.2.1. Pilot experimental setup

When it comes to the Pilot setup (Fig. 10), OPs accounted for ~80% of all identified proteins in the non-heated samples, while PDPs and MDPs together made up for only ~7.6%. As for NOPs, on average they made up ~12% of all identified proteins in this sample type.

After the heat treatment, OP content was reduced by around half, dropping to around 40% of all identified proteins, compared to ~80% in the non-heated samples. Proteins with higher disorder content (PDPs and MDPs), on the other hand, were significantly enriched by the heat treatment. Going by percental distribution, there were 4.4 times as many PDPs in the heated samples compared to the non-heated samples on average (from ~6.7% to ~30%).

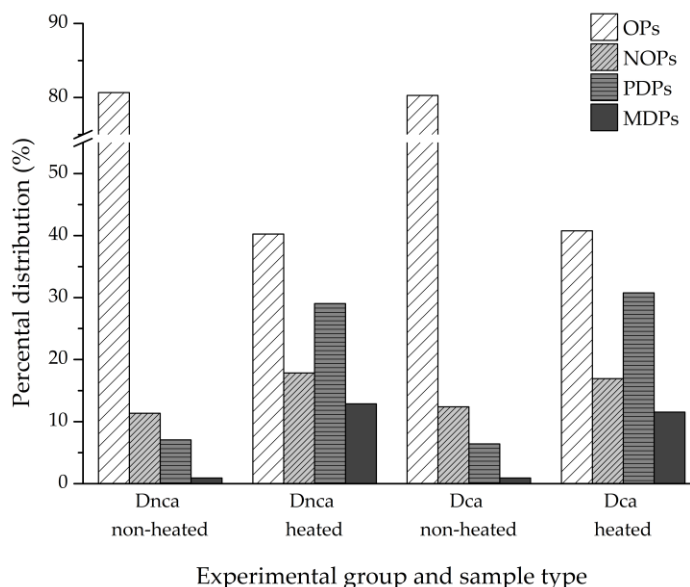

Figure 10. Effect of heat treatment on the percental distribution of proteins with varying degrees of intrinsic disorder in the two sample types of Pilot setup experimental groups. OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group).

When it comes to MDPs, they accounted for ~12.2% of all identified proteins in the heated samples on average, 13 times more than in the non-heated samples. The content of NOPs was also slightly enriched by heat treatment, going from ~12% in the non-heated samples to ~17% in the heated samples.

The breakdown of the absolute values of the data shown in Figure 10 is presented in Table 14 below:

Table 14. Total numbers of proteins with varying degrees of intrinsic disorder identified in the two sample types of Pilot setup experimental groups.

Experimental group	Sample type	OPs	NOPs	PDPs	MDPs
Dnca	Non-heated	263	37	23	3
	Heated	97	43	70	31
Dca	Non-heated	175	27	14	2
	Heated	53	22	40	15

OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group).

4.2.2.2. Main experimental setup

When it comes to the experimental groups of the Main setup, slightly more OPs were identified compared to the Pilot setup. On average, in the non-heated samples, OPs constituted ~82% of all proteins, compared to ~80% in the same sample type of the Pilot setup. On the other hand, the average content of NOPs, PDPs and MDPs in the Main setup groups is slightly lower in comparison to the Pilot setup.

Heat treating these samples (Fig. 11) has led to a similar pattern of enrichment as with the Dnca and Dca groups. Going by percental distribution, the amount of OPs in the heated samples dropped by more than a third (from ~82% to ~49%), while the amounts of MDPs and PDPs went up ~10 and 4 times, respectively. The content of NOPs was also slightly enriched by heat treatment, going from ~11% in the non-heated samples to ~14% in the heated samples.

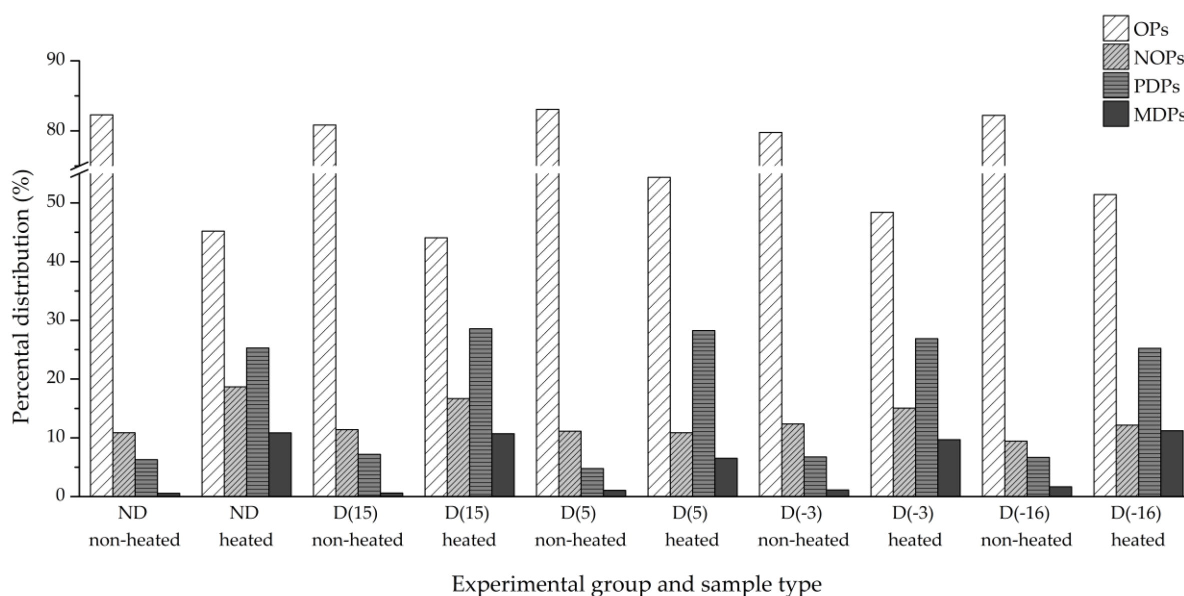

Figure 11. Effect of heat treatment on the percental distribution of proteins with varying degrees of intrinsic disorder in the two sample types of Main setup experimental groups. OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

The breakdown of the absolute values of the data shown in Figure 11 is presented in Table 15 below:

Table 15. Total numbers of proteins with varying degrees of intrinsic disorder identified in the two sample types of Main setup experimental groups.

Experimental group	Sample type	OPs	NOPs	PDPs	MDPs
ND	Non-heated	144	19	11	1
	Heated	75	31	42	18
D(15)	Non-heated	135	19	12	1
	Heated	37	14	24	9
D(5)	Non-heated	157	21	9	2
	Heated	50	10	26	6
D(–3)	Non-heated	142	22	12	2
	Heated	45	14	25	9
D(–16)	Non-heated	148	17	12	3
	Heated	55	13	27	12

OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

4.2.3. Determination of long intrinsically disordered regions (long IDRs)

Long intrinsically disordered regions (long IDRs) have a higher potential to possess biological relevance. A protein region is considered as such if it contains at least 20 consecutive disorder-promoting amino acids. Proteins identified in both experimental setups were further analyzed to determine whether they contain long IDRs.

4.2.3.1. Pilot experimental setup

The results from the Pilot setup show that the proteins from both the Dnca and Dca (Fig. 12A and B, respectively) experimental groups show mostly similar patterns when it comes to the distribution of long IDRs in their sequences. Ordered proteins are devoid of IDRs, as are most of the NOPs. The majority of PDPs contain either 1 or 2 such regions. When it comes to MDPs, most of them possess one long disordered segment, followed by proteins containing two, four, and three long IDRs, respectively. It can also be seen that the ratio of NOPs that contain at least 1 long IDR is higher in the cold-acclimated Dca group compared to the Dnca group.

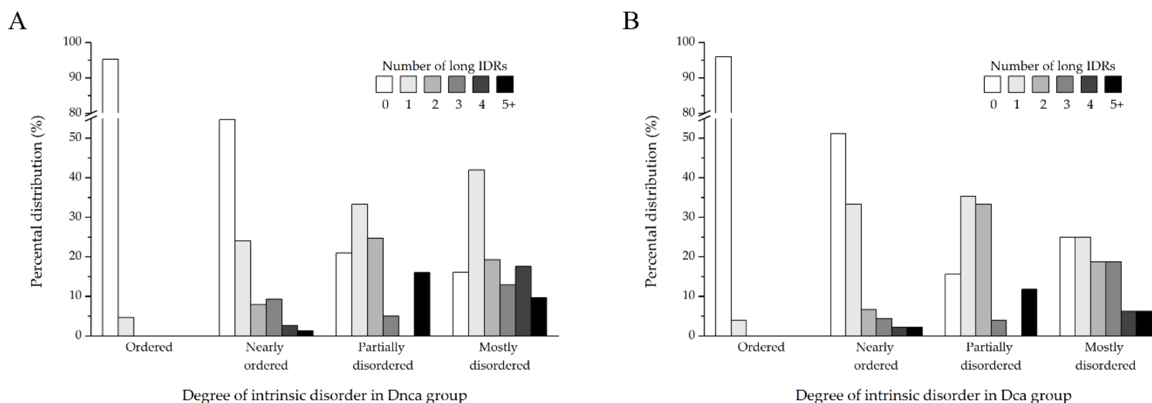

Figure 12. Distribution of long intrinsically disordered regions in proteins with various degrees of intrinsic disorder in the (A) Dnca – non-cold-acclimated diapausing and (B) Dca – cold-acclimated diapausing experimental groups. The different numbers of long IDRs in proteins are color-coded.

The breakdown of the absolute values of the data shown in Figure 12 is presented in Table 16 below:

Table 16. Long IDR content of proteins with various degrees of intrinsic disorder in the experimental groups of Pilot setup.

Experi-mental group	Disorder category	Number of long IDRs					
		0	**1**	**2**	**3**	**4**	**5+**
Dnca	OPs	304	15	0	0	0	0
	NOPs	41	18	6	7	2	1
	PDPs	17	27	20	4	0	13
	MDPs	5	13	6	4	0	3
Dca	OPs	194	8	0	0	0	0
	NOPs	23	15	3	2	1	1
	PDPs	8	18	17	2	0	6
	MDPs	4	4	3	3	1	1

IDR – intrinsically disordered region; OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins ,30–70%; MDPs – mostly disordered proteins, at least 70% (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group).

4.2.3.2. Main experimental setup

When it comes to the Main setup, there is a difference in the long IDR content between the ND control and the different diapausing experimental groups (Fig. 13).

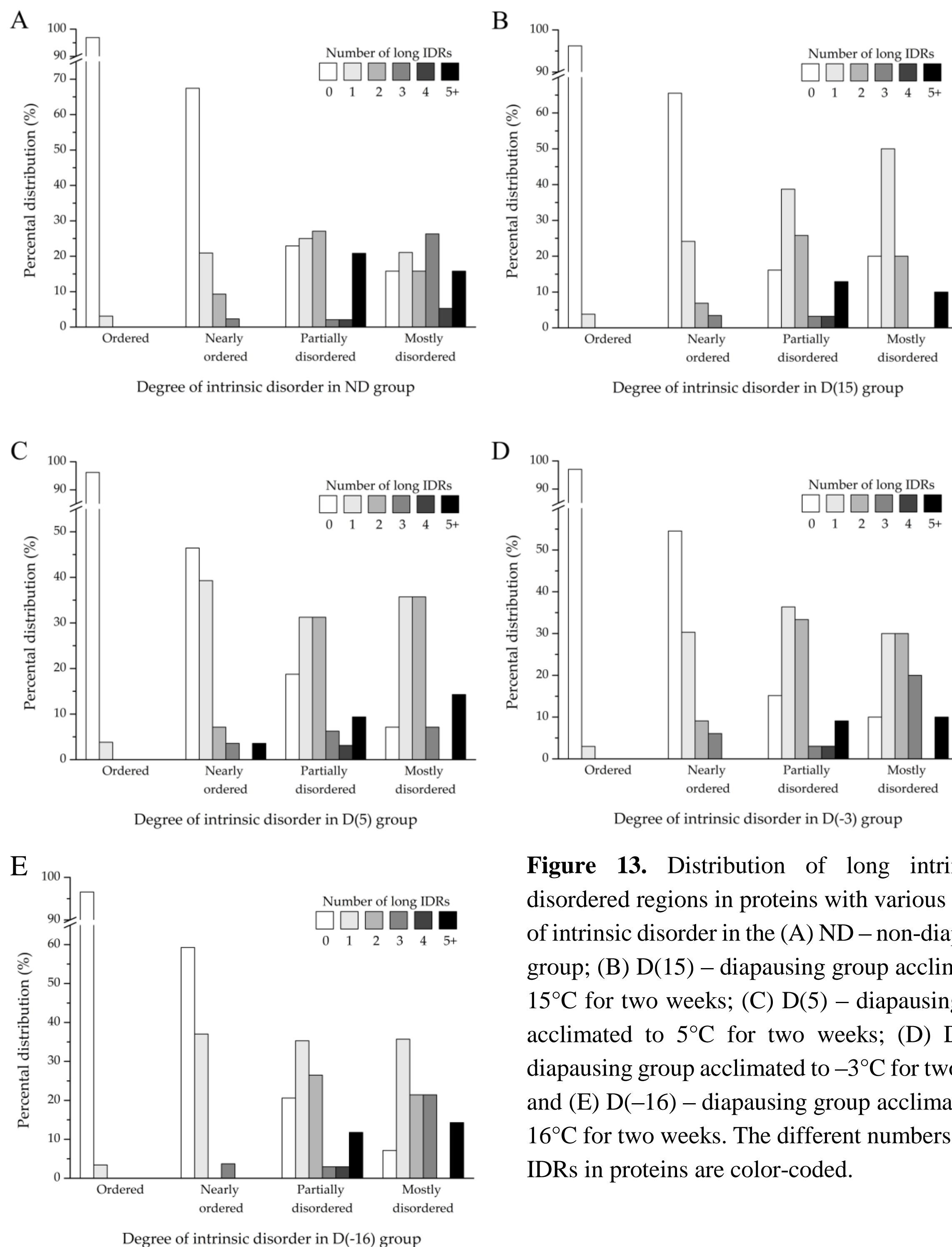

Figure 13. Distribution of long intrinsically disordered regions in proteins with various degrees of intrinsic disorder in the (A) ND – non-diapausing group; (B) D(15) – diapausing group acclimated to 15°C for two weeks; (C) D(5) – diapausing group acclimated to 5°C for two weeks; (D) D(–3) – diapausing group acclimated to –3°C for two weeks and (E) D(–16) – diapausing group acclimated to –16°C for two weeks. The different numbers of long IDRs in proteins are color-coded.

Most notably, the amount of NOPs that contain at least 1 long IDR is on average higher in the diapausing groups in comparison to the ND control. Additionally, diapausing groups have a significantly higher content of proteins with 1 or 2 long IDRs than the ones in the ND control. The breakdown of the absolute values of the data shown in Figure 13 is presented in Table 17 below:

Table 17. Long IDR content of proteins with various degrees of intrinsic disorder in the experimental groups of Main setup.

Experimental group	Disorder category	Number of long IDRs					
		0	**1**	**2**	**3**	**4**	**5+**
ND	OPs	188	6	0	0	0	0
	NOPs	29	9	4	1	0	0
	PDPs	11	12	13	1	1	10
	MDPs	3	4	3	5	1	3
D(15)	OPs	152	6	0	0	0	0
	NOPs	19	7	2	1	0	0
	PDPs	5	12	8	1	1	4
	MDPs	2	5	2	0	0	1
D(5)	OPs	176	7	0	0	0	0
	NOPs	13	11	2	1	0	1
	PDPs	6	10	10	2	1	3
	MDPs	1	5	5	1	0	2
D(–3)	OPs	163	5	0	0	0	0
	NOPs	18	10	3	2	0	0
	PDPs	5	12	11	1	1	3
	MDPs	1	3	3	2	0	1
D(–16)	OPs	169	6	0	0	0	0
	NOPs	16	10	0	1	0	0
	PDPs	7	12	9	1	1	4
	MDPs	1	5	3	3	0	2

IDR – intrinsically disordered region; OPs – ordered proteins, 10% at most; NOPs – nearly ordered proteins, 10–30%; PDPs – partially disordered proteins, 30–70%; MDPs – mostly disordered proteins, at least 70% (Dnca – non-cold-acclimated diapausing group; Dca – cold-acclimated diapausing group; ND – non-diapausing group; D(15) – diapausing group acclimated to 15°C for two weeks; D(5) – diapausing group acclimated to 5°C for two weeks; D(–3) – diapausing group acclimated to –3°C for two weeks; D(–16) – diapausing group acclimated to –16°C for two weeks).

4.2.4. Amino acid compositional analyses

4.2.4.1. Total amino acid content

The amino acid composition of every identified protein was analyzed, and the ratios of individual amino acids that make up the polypeptides were determined (Fig. 14). Depending on the degree of structural disorder, the analyzed proteins were comprised of varying amounts of disorder- and order-promoting amino acids. On average, mostly disordered proteins are composed of 68.9% disorder-promoting and 31.1% order-promoting amino acids. The composition of partially disordered proteins is similar to that of MDPs (64.9%/35.1%), while nearly ordered and ordered proteins are more balanced in terms of disorder- and order-promoting amino acid distribution — 60.4%/39.6% and 56.1%/43.9%, respectively.

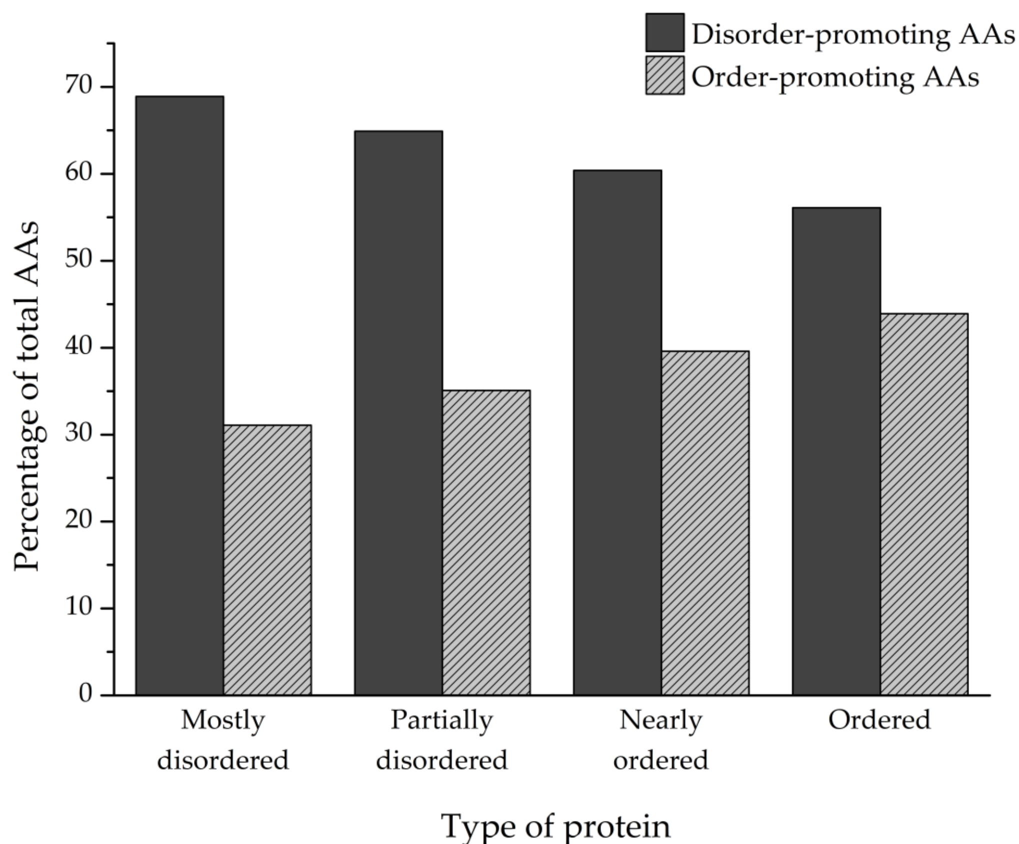

Figure 14. Ratios of disorder-promoting (P, E, S, K, Q, H, D, R, G, A) and order-promoting amino acids (T, C, N, V, L, M, I, Y, F, W) in mostly disordered (MDPs), partially disordered (PDPs), nearly ordered (NOPs), and ordered proteins (OPs).

4.2.4.2. Individual amino acid content

The identified MDPs and PDPs are particularly enriched in disorder-promoting amino acids (Fig. 15A), such as glutamate (11.91% and 11.64% of total amino acids in a sequence on average, respectively), lysine (9.34% and 9.26%), and glutamine (6.51% and 5.26%), compared to nearly ordered (7.85% E, 8.15% K, and 4.42% Q) and ordered proteins (6.56% E, 7.17% K, and 3.49% Q). The only exception is glycine, which is more prevalent in nearly ordered (7.27%) and ordered proteins (7.47%) than in MDPs and PDPs (5.95% and 5.85%, respectively). Next, MDPs contain almost double the amount of proline as the other three groups of proteins. When it comes to order-promoting amino acids (Fig. 15B), MDPs are comprised of low amounts of cysteine (0.4%), phenyl-alanine (1.97%), tyrosine (2.05%), isoleucine (3.57%), and valine (5.3%) in comparison to nearly ordered (1.1% C, 2.5% Y, 3.5% F, 5.13% I, and 7.26% V) and ordered proteins (2.05% C, 3.28% Y, 3.95% F, 5.8% I, and 7.38% V). Leucine is the standout order-promoting amino acid that partially disordered proteins contain in amounts similar to NOPs and OPs (8.08% compared to 7.7% and 8.45%, respectively), unlike MDPs (5.37%). The distribution of the remaining amino acids is more or less similar between the four groups of proteins.

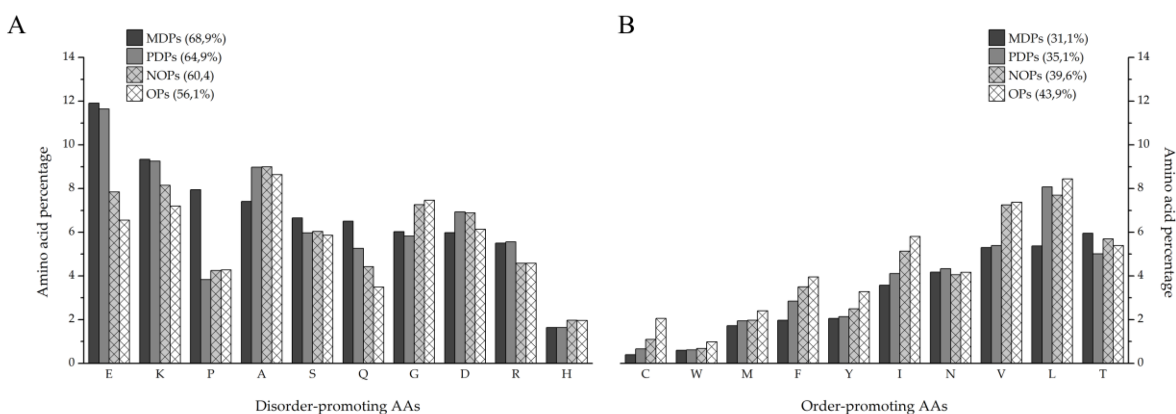

Figure 15. Ratios of individual disorder-(A) and order-promoting amino acids (B) in mostly dis-ordered (MDPs), partially disordered (PDPs), nearly ordered (NOPs), and ordered proteins (OPs), ordered by abundance in MDPs.

Due to the volume of data points a summary of individual amino acid content in the four different groups of proteins is presented in a table that is available upon request from the author. The table includes protein names, accession numbers, calculated percental disorder and ratios of individual amino acids.

4.2.5. Functional analysis of disordered proteins

To gain an insight into the biological importance of IDPs in the cold adaptation process of the ECB, we performed a bioinformatic functional analysis of the identified disordered proteins using the data from online knowledgebases Uniprot, Pfam, Interpro, and Gene Ontology. Our results have revealed that only 458 of the proteins have a Uniprot entry and at least one data point from the other listed knowledgebases. Out of that number, 143 proteins are either mostly or partially disordered or contain at least 1 long IDR (Fig. 16, Total unique).

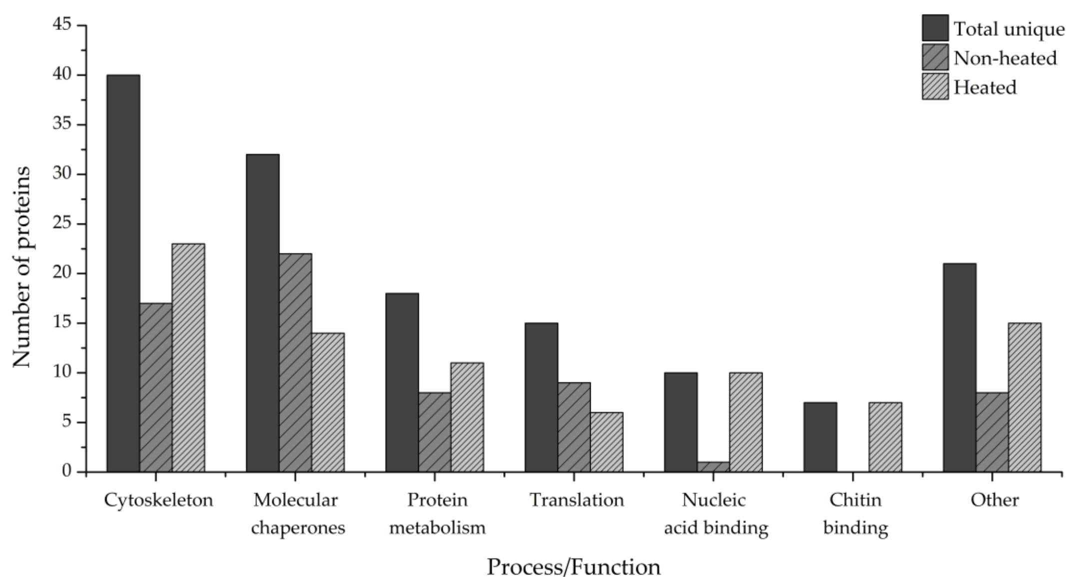

Figure 16. Biological processes and molecular functions of intrinsically disordered proteins and proteins containing long IDRs. The category Other encompasses processes and functions that make up less than 4% of total hits each. Total unique – all uniquely identified proteins; Non-heated – all proteins identified in the non-heated sample types; Heated – all proteins identified in the heated sample types.

The largest functional group (40 unique hits) is comprised of cytoskeleton structural components or proteins associated with them, for example regulators of muscle contraction or actin filament organization proteins. The second-largest group encompasses proteins functioning as molecular chaperones (32 unique hits), followed by proteins involved in protein and amino acid metabolism (18 unique hits). The rest of the proteins cover a wide range of biological processes and molecular functions, including translation (15 unique hits), nucleic acid binding (10 unique hits), chitin binding and formation of cuticle (7 unique hits), and others (Fig. 16, Total unique). Additionally, heat treatment of the samples increased the number of proteins that could be identified and functionally analyzed. Proteins that are involved in chitin binding and cuticle formation were found only in the

heated samples, as were the majority of nucleic acid binding proteins (10 total hits compared to 1 total hit). More proteins belonging to the Cytoskeleton category were also present in the heated samples (23 total hits) compared to the non-heated ones (17 total hits). More proteins that act as molecular chaperones, on the other hand, were present in the non-heated samples (22 total hits) than in the heated samples (14 total hits) (Fig. 16, Non-heated, Heated).

When it comes to processes and functions grouped in the Other category (21 unique hit), electron transport chain (ETC) components and proteins involved in lipid metabolism are the most represented with 4 unique hits each (Fig. 17, Total unique). Additionally, ETC components were only identified in the heated samples, as were proteins involved in enzyme regulation, transcription, signaling regulation and TCA cycle. In fact, the majority of proteins from the Other category were identified after sample heating, in comparison to non-heated samples (15 and 8 total hits, respectively).

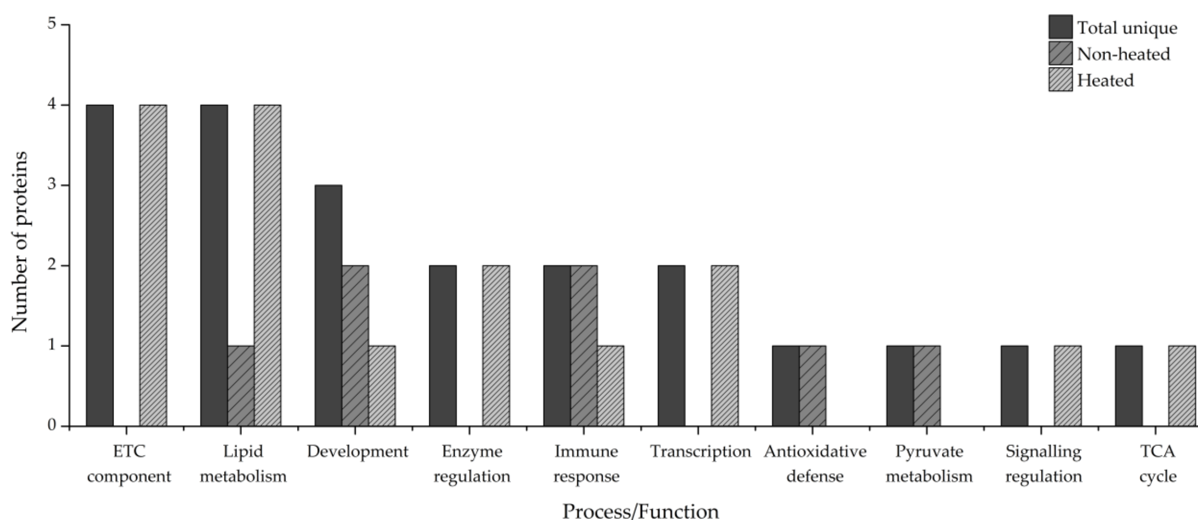

Figure 17. Biological processes and molecular functions of intrinsically disordered proteins and proteins containing long IDRs in the Other category. Total unique – all uniquely identified proteins; Non-heated – all proteins identified in the non-heated sample types; Heated – all proteins identified in the heated sample types; ETC – electron transport chain; TCA – tricarboxylic acid.

Due to the volume of data points, a summary of all proteins with intrinsic disorder that have been annotated is presented in a table that is available upon request from the author. The table includes protein names, accession numbers, calculated percental disorder, number of long IDRs and functions that were assigned to the proteins.

4.3. Relative gene expression analyses

All relative gene expression analyses were performed on samples prepared as part of the Main experimental setup, with ND (non-diapause) considered as the control group.

4.3.1. Ct values of reference genes

There were no statistically significant differences in the variances for the measured Ct values of the reference genes (Fig. 18). The Ct values of *actin* were higher than the Ct values of *rps3*, except in the ND experimental group where the values for *rps3* were slightly higher than the values for *actin*.

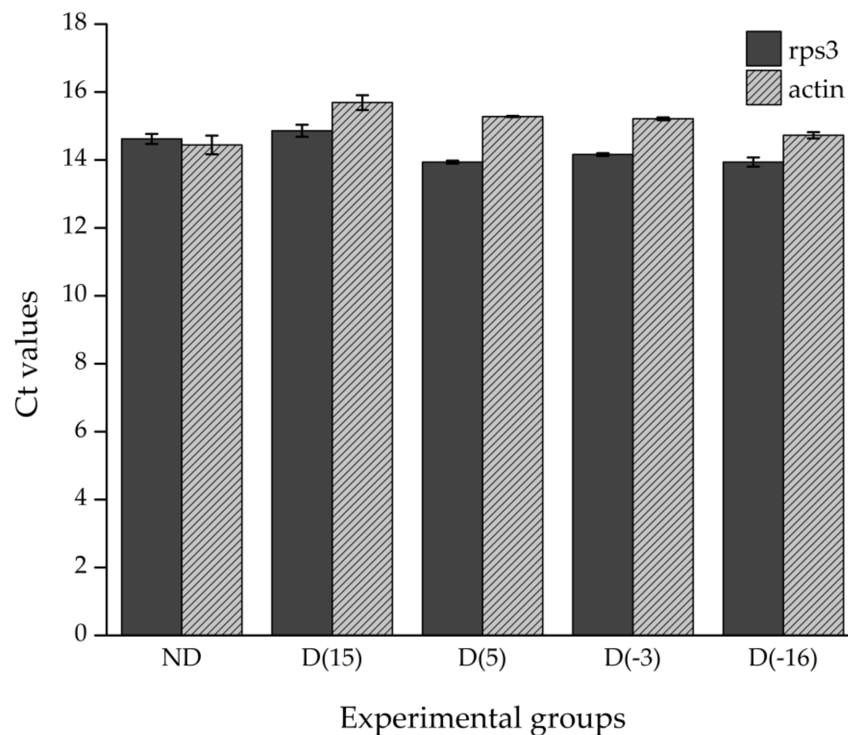

Figure 18. Ct values of reference genes – *rps3* and *actin*. The results are presented as mean Ct values ± standard error of mean of 6 qPCR reactions per experimental group – 3 biological replicates per group, performed in 2 technical replicates each. One-way ANOVA and *post hoc* Tuckey test (significance level $p<0.05$) were used to statistically analyze the results.

4.3.2. Relative expression of *hsp90* gene

The results of relative gene expression for *hsp90*, a major molecular chaperone, not only differ between the experimental groups, but the differences in the diapausing groups are also conspicuously contrasted (Fig. 19).

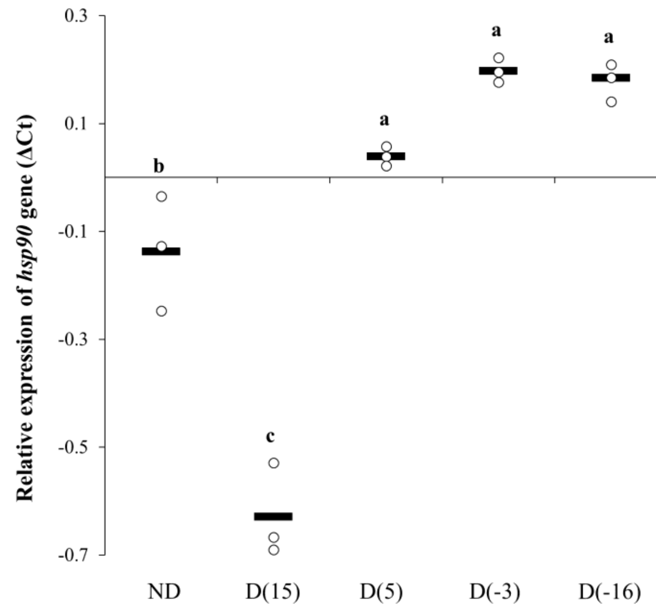

Figure 19. Relative expression of *hsp90* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

Compared to the non-diapause control and other diapausing groups which have been cold acclimated, the gene expression of *hsp90* is severely down-regulated in the D(15) group which was in early diapause. As diapause and cold acclimation progressed in the D(5), D(–3) and D(–16) groups, the expression of the *hsp90* gene increased to values higher than in the ND control.

Statistically significant differences in relative expression are also presented in Table 18, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 18. Relative expression of *hsp90* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level $p<0.05$ (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	1	1
D(15)			1	1	1
D(5)				0	0
D(–3)					0
D(–16)					

4.3.3. Relative expression of *hsc70* gene

The results of relative gene expression for *hsc70*, which is considered a constitutively expressed molecular chaperone, are contrasted between the non-diapause control and the diapausing, cold acclimated groups. (Fig. 20).

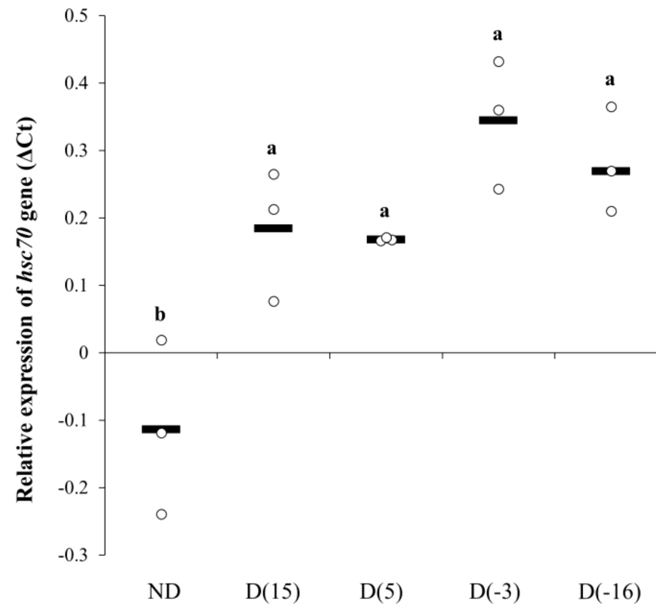

Figure 20. Relative expression of *hsc70* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

As with the previous two heat shock protein genes, the relative gene expression of *hsc70* is also upregulated in the diapausing groups when compared to the ND control. While the differences in expression between the individual diapausing groups were not statistically significant, the absolute calculated values of ΔCt were higher in the two diapausing groups that were acclimated to temperatures below 0°C (D(–3) and D(–16)), in comparison to the other two diapausing groups.

Statistically significant differences in relative expression are also presented in Table 19, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 19. Relative expression of *hsc70* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level $p<0.05$ (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	1	1
D(15)			0	0	0
D(5)				0	0
D(–3)					0
D(–16)					

4.3.4. Relative expression of *hsp70* gene

The results of relative gene expression for the inducible molecular chaperone *hsp70* differ between the experimental groups (Fig. 21). When compared with the ND group, *hsp70* expression steadily increases with the cold acclimation treatment and progression of diapause.

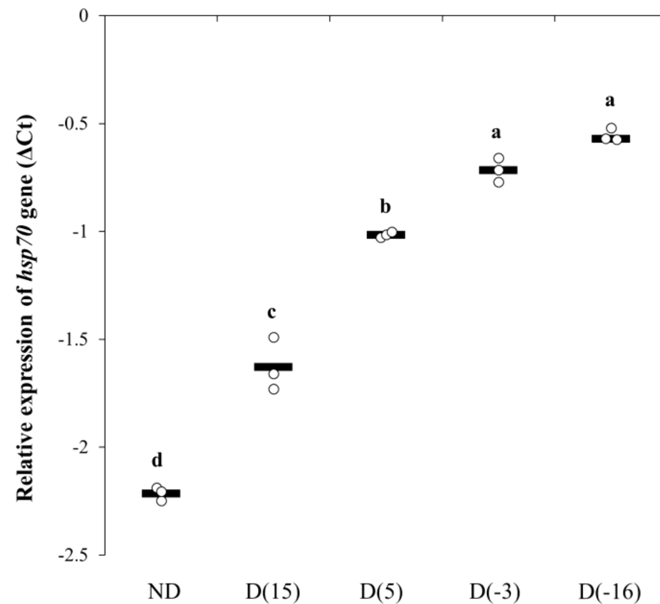

Figure 21. Relative expression of *hsp70* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

Relative expression of the *hsp70* gene begins significantly increasing already in the D(15) group, when compared with the ND control, and reaches the highest recorded values in the D(–16) group. Although differences in expression between groups D(–3) and D(–16) were not statistically significant, the absolute values were higher in the latter group.

Statistically significant differences in relative expression are also presented in Table 20, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 20. Relative expression of *hsp70* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 (1 – statistically significant; 0 – not statistically significant)

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	1	1
D(15)			1	1	1
D(5)				1	1
D(–3)					0
D(–16)					

4.3.5. Relative expression of *hsp20.4* gene

The results of relative gene expression for the small, ATP-independent chaperone *hsp20.4* are similar to the pattern of *hsp70* expression, and they differ between the experimental groups (Fig. 22).

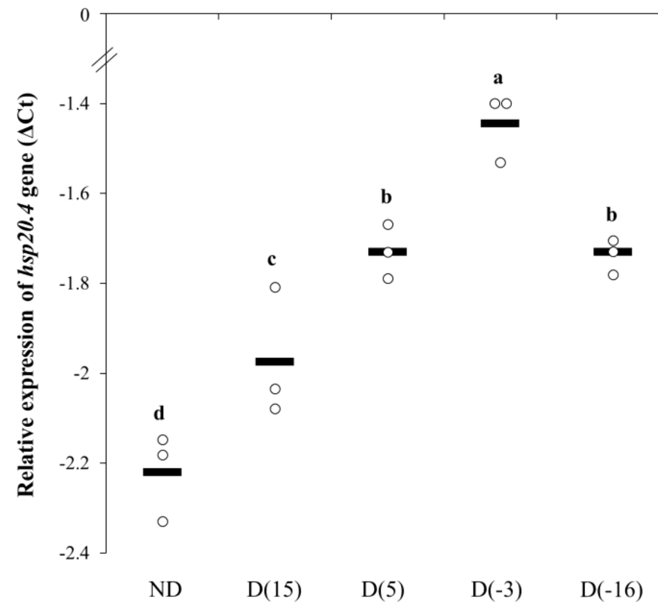

Figure 22. Relative expression of *hsp20.4* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

The lowest relative expression was recorded in the ND control, and it steadily increased as diapause and cold acclimation progressed. Expression levels peaked in the D(–3) group, before dropping in the D(–16) group which was acclimated at the lowest temperature and spent the longest time in diapause of all experimental groups.

Statistically significant differences in relative expression are also presented in Table 21, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 21. Relative expression of *hsp20.4* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	1	1
D(15)			1	1	0
D(5)				1	0
D(–3)					1
D(–16)					

4.3.6. Relative expression of *hsp20.1* gene

The results of relative gene expression for *hsp20.1*, also an ATP-independent small chaperone protein, differ between not only the ND control and diapausing groups in general, but also between the diapausing experimental groups individually (Fig. 23).

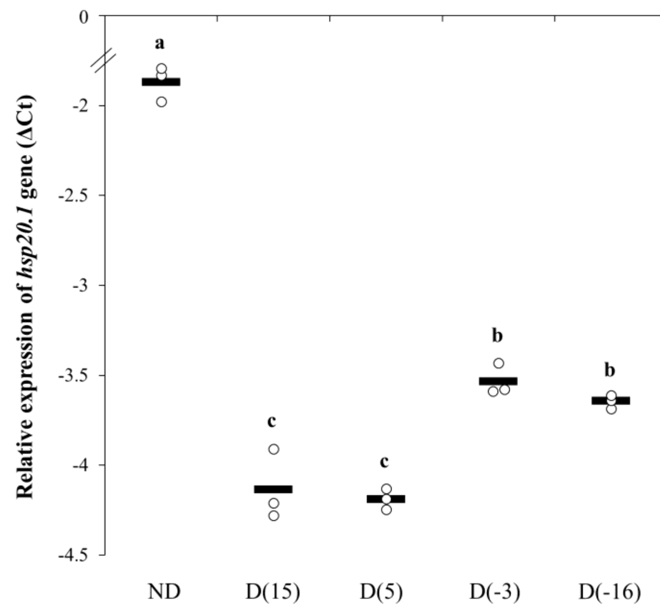

Figure 23. Relative expression of *hsp20.1* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

Unlike the previous genes coding different heat shock proteins, the relative expression of *hsp20.1* is down-regulated during diapause and exposure to low temperatures in comparison to the control group. When it comes to the individual diapausing groups, progress of the diapause program and acclimation to temperatures below 0°C have led to an increase of *hsp20.1* expression when compared with groups that were in earlier stages of diapause and acclimated to temperatures above 0°C.

Statistically significant differences in relative expression are also presented in Table 22, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 22. Relative expression of *hsp20.1* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level $p<0.05$ (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	1	1
D(15)			0	1	1
D(5)				1	1
D(–3)					0
D(–16)					

4.3.7. Relative expression of *tropmy2* gene

Tropomyosin-2 is involved in controlling muscle contraction. The results of relative gene expression for *tropmy2* in the ND control are different from those in the diapausing and cold acclimated experimental groups (Fig. 24).

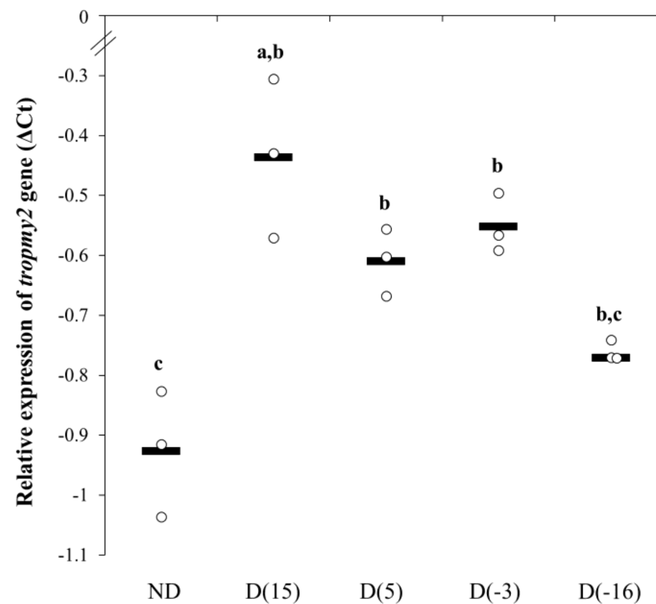

Figure 24. Relative expression of *tropmy2* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

The highest expression was recorded in the D(15) group that corresponds to early diapause and where samples were not yet acclimated to temperatures around and below 0°C. As diapause progressed and temperatures were lowered, the expression of *tropmy2* decreased as well. Of all the diapausing groups, D(–16) samples had the lowest recorded expression levels, which on par with the expression recorded in the ND control.

Statistically significant differences in relative expression are also presented in Table 23, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 23. Relative expression of *tropmy2* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	1	0
D(15)			0	0	1
D(5)				0	0
D(–3)					0
D(–16)					

4.3.8. Relative expression of *tnt* gene

Troponin T is a member of the troponin complex regulating muscle activity. The results of relative gene expression for *tnt* differ between the ND control and D(15) groups on one side and the remaining diapausing experimental groups on the other side (Fig. 25).

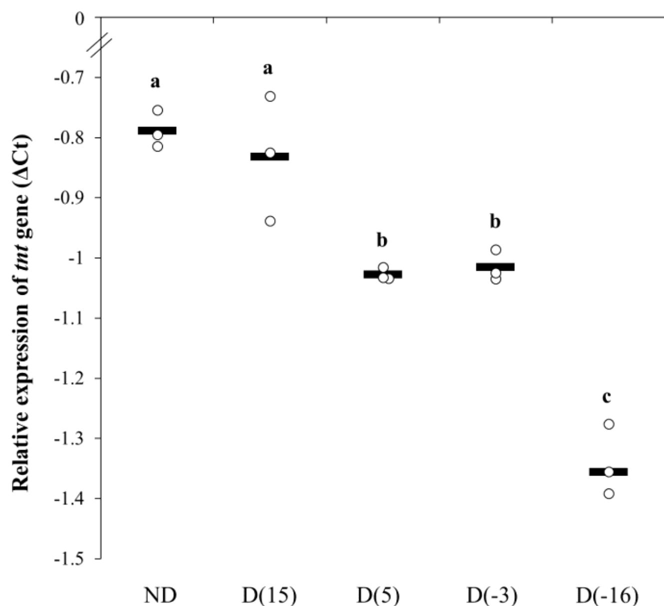

Figure 25. Relative expression of *tnt* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

The non-diapausing control, as well as the D(15) samples that were in early diapause had the highest expression of *tnt* gene. In groups that were further along in diapause and acclimated to temperatures around and below 0°C gene expression levels were decreased. In particular, out of all diapausing groups, gene expression was significantly decreased in the D(–16) group which was acclimated at the lowest temperature and in late diapause.

Statistically significant differences in relative expression are also presented in Table 24, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 24. Relative expression of *tnt* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level $p<0.05$ (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		0	1	1	1
D(15)			1	1	1
D(5)				0	1
D(–3)					1
D(–16)					

4.3.9. Relative expression of *thym* gene

Thymosin beta is a polypeptide involved insect humoral response to microbial infections. The results of relative gene expression for *thym* differ between the ND control on one side and the diapausing experimental groups on the other (Fig. 26).

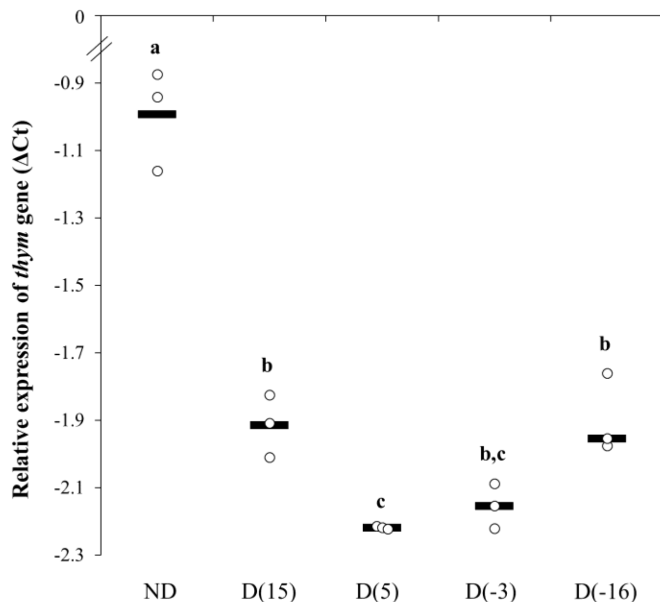

Figure 26. Relative expression of *thym* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

Expression of *thym* was significantly down-regulated in all diapausing and cold acclimated groups in comparison to the non-diapausing control. When it comes to the individual diapausing groups, expression levels were mostly similar with the exception of the D(5) group, where the lowest expression of *thym* gene was recorded.

Statistically significant differences in relative expression are also presented in Table 25, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 25. Relative expression of *thym* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	1	1
D(15)			1	0	0
D(5)				0	1
D(–3)					0
D(–16)					

4.3.10. Relative expression of *moe* gene

Moesin is a cytoskeletal protein involved in many regulatory processes. The results of relative gene expression for *moe* differ between the experimental groups (Fig. 27). When compared with the ND control, expression is down-regulated during most of diapause when temperatures are above or around 0°C.

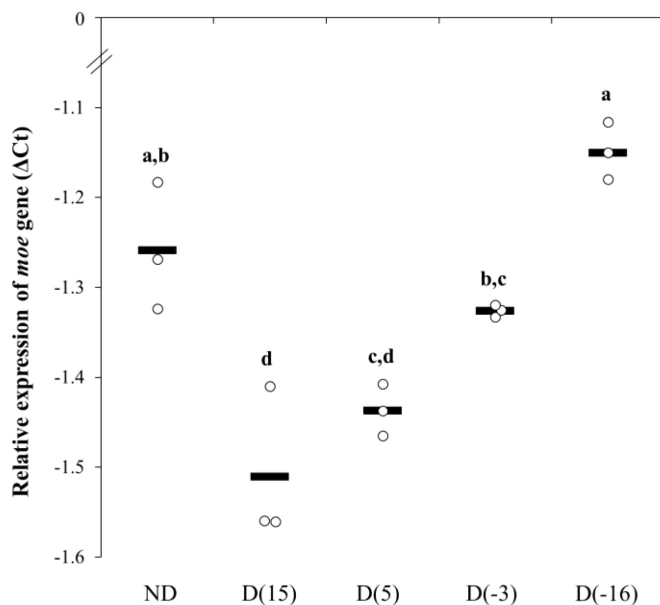

Figure 27. Relative expression of *moe* gene. The results are expressed in ΔCt and presented as univariate scatterplots. Each dot represents one biological pool comprised of 3 larvae. The black bars represent the mean of ΔCt values for each group. Statistically significant differences in relative expression between groups are labeled with letters above the scatterplots.

Looking at the diapausing groups, expression of *moe* is lowest in early diapause (D(15) group) and steadily increases as diapause and acclimation to temperatures around and below 0°C progress. By late diapause and at lowest temperatures (D(–16) group) expression returns to levels recorded in the non-diapause control.

Statistically significant differences in relative expression are also presented in Table 26, while the means of calculated ΔCt values and standard errors of means are presented in Table 28 in the Supplementary Material.

Table 26. Relative expression of *moe* gene in whole-body *O. nubilalis* larvae – one-way ANOVA and *post hoc* Tuckey test for significance level p<0.05 (1 – statistically significant; 0 – not statistically significant).

	ND	D(15)	D(5)	D(–3)	D(–16)
ND		1	1	0	0
D(15)			0	1	1
D(5)				0	1
D(–3)					1
D(–16)					

5. Discussion

5.1. Proteomic analyses

Proteomic analyses of intrinsically disordered proteins/protein regions and ID-containing proteins, based on *in vitro* methodologies and techniques, are fraught with obstacles. The difficulties stem from the fundamental structural properties of these biomolecules. Being intrinsically disordered, these proteins are usually not able to crystallize. Consequently, they are not amenable for structural analyses using the typical methods for resolving protein structure, such as X-ray crystallography (Piovesan et al., 2017). High content of protein disorder is also prohibitive when it comes to structure determination employing cryo-EM. In addition, cryo-EM is quite restrictive when it comes to the size of analytes, which limits its application in determining the structure of individual proteins (Bari and Prakashchand, 2021; Avramov et al., 2022). As such, proteins containing intrinsic disorder are far less studied than proteins that are globular and possess a well-ordered structure (Siemer, 2020).

5.1.1. Enrichment of IDP content and enrichment validation

However, the same fundamental physicochemical properties, which present obstacles in the study of IDP structures, can also provide unique advantages in other aspects of proteomic research. As has been stipulated, IDPs can behave vastly different than ordered, globular proteins in environments where conditions are not at physiological levels. Instead of denaturing or, at the very least, losing some level of biological activity, IDPs obtain transient ordered structures under these conditions. Additionally, under such circumstances, that is what allows them to perform their various functions (Uversky, 2009).These unique properties of IDPs can be employed in several ways in order to enrich samples to be analyzed with proteins containing various degrees of intrinsic disorder. One such approach entails treating samples with acidic precipitating agents, e.g. perchloric or trichloracetic acid (PCA and TCA, respectively). Applying either of these chemicals leads to the denaturing and precipitation of globular proteins in the samples, which are then subsequently removed by centrifugation (Cortese et al., 2005; Romero-Pérez et al., 2023). The remaining content is in this way enriched in intrinsically disordered proteins, although some globular proteins are quite stable even at critically low pH and therefore are still present in the sample after acid treatment. Next, the enriching effects of this treatment can be explored and validated by 2D PAGE methods, which offer considerable resolving power.

Apart from treating the samples with precipitating agents, the content of intrinsically disordered proteins can be enriched by exposing the samples to boiling temperatures (Galea et al., 2006, 2009; Zhang et al., 2018). Again owing to their unique physicochemical properties, in particular low mean hydrophobicity and high net charge, IDPs exhibit resistance to heat denaturation and low thermal aggregation. Instead of denaturing, proteins containing intrinsic disorder can adopt transient, higher orders of structure while the high temperature conditions persist (Uversky, 2009). As with the acid treatment, globular proteins will denature and can be removed from the samples with subsequent centrifugation. What will remain in the samples are heat-stable intrinsically disordered proteins.

For the purposes of the studies performed in this dissertation, heat sampling was chosen as the preferred method for enriching the samples with intrinsically disordered proteins. The method is simple, straightforward and provides adequate reduction of the content of globular proteins, which in turn uncovers proteins that would have been masked by their ordered counterparts (Galea et al., 2006; Zhang et al., 2018; Avramov et al., 2022). Acid treatment, on the other hand, requires taking into consideration that different precipitating agents and their concentrations lead to different proteomic profiles that can be detected after the treatment (Cortese et al., 2005). Following enrichment, its effects were validated using a particular, modified 2D-PAGE method. This in-house developed method (Csizmók et al., 2006; Tantos and Tompa, 2012) relies on the use of 8M urea to further filter out globular proteins from the samples that are being analyzed. Urea is used to solubilize the proteins after separation in the first dimension, and also as a component of the resolving gel for the second dimension separation. Intrinsically disordered proteins are unaffected by the urea during electrophoresis and align in a distinct diagonal pattern in the second dimension. The movement of globular proteins is impeded by the urea, and they typically end up located above the aforementioned diagonal. Looking at Figures 5A and 5B, it becomes sufficiently apparent why enrichment is the necessary first step in proteomic analyses of intrinsically disordered proteins. Non-heated sample types are abundant in globular, ordered proteins which impeded the separation in the first dimension, as well as transfer into the second dimension gel (Fig. 5A). The likely culprits are high molecular weight arylphorins – storage proteins that are present in large amounts in the haemolymph of insects such as the ECB (Taški et al., 2004). Once the offending proteins were removed by heating the samples up, separation in both the first and second dimension was more successful (Fig. 5B). In the second dimension, it can be seen that

several groups of proteins are much more visible and are aligned along the diagonal. Their positioning, as well as presence after the enrichment procedure, indicate that they are either partially or mostly disordered (Csizmók et al., 2006; Tantos and Tompa, 2012). Some globular proteins can still be noticed, as they appear in several spots above the diagonal. As it can be the case with acid treatment, where some globular proteins exhibit structural stability even at acidic pH levels, so can some globular proteins show they are thermostable and remain in the samples after heat treatment.

5.1.2. Identification of proteins by LC-MS/MS

Protein identification in non-heated and heated samples from both experimental setups was performed by matching the MS/MS spectra of detected peptides against the protein sequences that are available in the NCBI database not only for the species in question, *O. nubilalis*, but also for the Asian corn borer (ACB, *Ostrinia. furnacalis*) and all other lepidopteran species. The decision to expand the search paramateres to also include these other insects was made due to the limited number of ECB protein sequences available in the aforementioned database (~1600 entries) in comparison to, for example, the ACB (around 28 500 entries). The European corn borer has not been a subject of proteomics-level studies (Purać et al., 2016), and as such its protein entries are relatively few in number when compared with other species. Furthermore, the ACB was selected for protein identification as it is a close relative of the ECB. There is recorded gene flow between the two moth species, and they produce viable hybrid offspring after interspecies mating (Domingue et al., 2008; Li and Yang, 2022).

A total of 2103 proteins were identified in the entire study. The majority of the proteins, as was to be expected due to the limited availability of ECB sequences, were identified because the search parameters were expanded to include other lepidopteran species. Despite this, sequence similarity was at such a level that the peptides acquired from protein digestion of ECB homogenates could be matched with existing sequences from these other species with high confidence. The total takes into account that identical proteins could be identified across the different experimental groups and sample types. When the duplicates are addressed, the number of unique proteins that have been identified in this study comes to 892. Of that number, 361 proteins were common for the two experimental setups, while 292 and 284 were unique to the Pilot and Main setups, respectively.

When looking at the breakdown of the proteins identified in every group, it can be noticed that the numbers vary conspicuously. In the Pilot setup, nearly three times as many proteins were uniquely identified in the Dnca group compared to the Dca group – 294 and 102 proteins, respectively. The remaining proteins, 212 of them, were found in both of these groups (Fig. 6A).

In the Main setup, on the other hand, more unique proteins were identified in the ND control group (218) than in any of the diapausing, cold-acclimated groups (142, 171, 158 and 164, respectively). These five experimental groups all shared 86 proteins between them (Fig. 6B). These discrepancies in both setups highlight the effects and merits of the experimental designs – cold acclimation and diapause/resting state, on the proteome of the sampled insects (Grubor-Lajšić et al., 1991; Hahn and Denlinger, 2011; Kojić et al., 2018; Uzelac et al., 2020; Popović et al., 2021; Avramov et al., 2022) in increasing the number of unique proteins that can be identified in the study. Lastly, an unintentional finding was made during this process, in regards to the sequences that were recognized in the NCBI database by peptide matching. In both the Pilot and Main setups, nearly half of the identified proteins (47% and 45%, respectively) come from entries that have only been predicted from available nucleotide sequences. It is important to highlight these results, as the findings in this study lend experimental validation for the protein sequence prediction endeavors.

5.1.3. Effect of sample heating on protein identification

Protein identification was also largely affected by whether the samples were heat-treated or not. As can be expected, fewer proteins were identified in non-heated samples. On average, non-heated samples contained at least double the amount of unique proteins than their heated counterparts (Fig. 7, Table 12). The only exception to this pattern is the ND control group of the Main experimental setup (Fig. 7B). Surprisingly, the two sample types of this experimental group contained a nearly identical number of unique proteins – 165 in the non-heated samples and 151 in the heated samples. Apart from this exception, the non-heated samples of all other experimental groups had, on average, 30–50% more unique proteins than their heated counterparts. Another way to look at these results is that heating the samples allowed for additional unique proteins to be identified. There are several reasons why these proteins would be detectable only after sample heating. In the non-heated samples, their presence could potentially be masked by the abundant and overrepresented proteins, preventing their signal from being detected by LC-MS/MS identification. Some proteins are embedded as components of multi-subunit complexes, which also

impedes their detection and identification. By heating the samples overabundant proteins are removed and the large complexes are broken up, driving the identification of additional unique proteins (Avramov et al., 2022).

5.2. Bioinformatical analyses

Due to the specific compositional characteristics and biases of intrinsically disordered proteins and protein regions, different *in silico* methods based on these compositional specifics have been developed over the years to analyze different aspects of IDPs and proteins with IDRs (He et al., 2009; Deng et al., 2012; Eickholt and Cheng, 2013; Jones and Cozzetto, 2015; Necci et al., 2021). In this study, the IUPred platform (Dosztányi et al., 2005), as well as locally run PERL scripts, were employed to reliably analyze the identified proteins in regards to their effective and localized intrinsic disorder, amino acid composition and functional characterization.

5.2.1. Determination of intrinsic disorder

The disorder predictor IUPred uses an energy estimation approach for determining intrinsic disorder content in proteins directly from their amino acid sequences (Dosztányi et al., 2005; Dosztányi, 2018). The analyzed proteins were divided into four distinct categories based on their content of intrinsic disorder, expressed as percental disorder. Proteins with at least 70% of percental disorder were designated as mostly disordered (MDPs), between 70% and 30% as partially disordered (PDPs), between 30% and 10% as nearly ordered (NOPs) and below 10% as ordered proteins (OPs). In both the Pilot and the Main experimental setups more than 30% of the identified proteins were determined to be either partially or mostly disordered, or contain localized intrinsic disorder (Figs. 8, 12, and 13). These findings are in accordance with previous studies that were centered on exploring the prevalence of protein intrinsic disorder across the living world. When it comes to eukaryotes, computational analyses have predicted that, on average, at least one third of eukaryotic proteomes contain significant intrinsic disorder (Pancsa and Tompa, 2012; Habchi et al., 2014; Xue et al., 2014; Peng et al., 2015; Uversky, 2015; Bondos et al., 2021). The remaining proteins from both experimental setups were determined to contain less than 10% percental disorder, making them ordered proteins (OPs). However, despite being labeled as ordered, a significant proportion of these proteins (~75% of OPs in either setup) were calculated to contain non-zero percent of intrinsic disorder. Percental disorder value of 0% would indicate a

complete absence of intrinsic disorder and the protein would be fully ordered. Such proteins are rare, evidenced by the findings that only 7% of proteins present in the reference Protein Data Bank (PDB) have been determined to contain no disordered residues in their sequences (Gall et al., 2007).

5.2.2. Effect of sample heating on protein disorder distribution

The majority of identified proteins that were determined to be either mostly or partially disordered were only revealed after sample heating and enrichment to remove globular proteins. Heating the samples had a particularly strong effect on increasing the amounts of MDPs. In the heated samples of the Pilot experimental setup, the content of MDPs went up on average 13 times compared to the non-heated samples (Fig. 10). A similar pattern was observed in the heated samples of the Main experimental setup – an average of 10-fold increase compared to the non-heated samples (Fig. 11). Partially disordered proteins were similarly enriched by this process, increasing around 4-fold in both the Pilot and Main experimental setups.

What is interesting to note is how the content of ordered proteins was affected by the heat treatment. A non-insignificant amount of OPs remained in the samples even after heating. Their numbers dropped by at most 50% in some cases, with a considerable amount of potentially heat-stable ordered proteins still present in the samples (Figs. 10 and 11). Taken altogether, these results again underscore the importance of sample enrichment, in this case by heating, in order not to miss out on additional proteins with intrinsic disorder. Moreover, this type of comparison between non-treated and treated sample types is something that is usually left out in studies that are based on IDP enrichment methods (Cortese et al., 2005; Galea et al., 2006; Galea et al., 2009; Zhang et al., 2018; Romero-Pérez et al., 2023). Here, on the other hand, it can be seen that such comparisons can provide valuable information regarding the enrichment process, and ensure that as many proteins as possible are considered for further analyses.

5.2.3. Long intrinsically disordered regions (long IDRs)

Considering that even ordered proteins can contain segments that lack structure, all proteins from the data set in this study were additionally analyzed to determine their content of long intrinsically disorder regions (long IDRs). These regions are stretches of 20 or more disorder-promoting residues in a row (Necci et al., 2016). Long IDRs can be more informative of function, compared

to purely determining the percental disorder of a protein, as this flavor of intrinsic disorder has a higher potential for possessing biological relevance. Amongst other things, these regions often contain sites of post-translational modifications, assist in molecular recognition and binding, function as flexible linkers which promote domain movement (Oldfield and Dunker, 2014; Davey, 2019; Deiana et al., 2019). In particular, post-translational modifications of sites within long IDRs, in otherwise ordered proteins, facilitate conformational changes which enable interactions with many different binding partners (Darling and Uversky, 2018; Deiana et al., 2019).

Proteins that scored under 70% percental disorder – OPs, NOPs and PDPs, were the first to be analyzed for long IDR content, considering that these regions can be significant for the function of the more ordered proteins (Figs. 12 and 13). Additionally, these proteins make up a significant portion of the proteins identified in both experimental setups – 75% of unique proteins in the Pilot and 78% of unique proteins in the Main setup belong to these three groups. According to the analysis, the majority of these proteins (over two thirds) contain no long IDRs, with most of these proteins (>80%) belonging to the OP category. When it comes to NOPs, more than 50% of them also do not contain long IDRs, along with, surprisingly, around 20% of identified PDPs. While it might seem like an inconsistency between the observed percental disorder of analyzed proteins and whether they contain or lack long IDRs, these results might rather be indicative of how intrinsic disorder is localized in the structures of these groups of proteins. It is likely that those PDPs, which have been shown to contain no long IDRs, probably contain segments with contiguous disordered residues that are shy of the 20 amino acid threshold to be considered long IDRs. On the other hand, these segments are still long enough to influence the measured IUPred score of these proteins, which classifies them as at least partially disordered. When it comes to the two groups of more ordered proteins, OPs and NOPs, their small amounts of measured intrinsic disorder are likely located within the single long IDR that some of them possess. Looking at all these three protein groups together, regardless of experimental setup, the majority of them contain exactly one long IDR (55%), while just more than a fifth contain two long IDRs. The rest of these proteins contain between 3 and up to 36 long IDRs, such as in some titin-like giant proteins.

Long IDR content analysis was performed for MDPs as well, in order to further characterize their intrinsic disorder. Similarly to the previous three groups of proteins, the majority of MDPs contained either one or two long IDRs (around 55%, regardless of experimental setup), followed

by proteins with three, four or more long intrinsically disordered regions. However, a non-insignificant portion of these proteins (20% in the Pilot setup and 14% in the Main setup) contained no long IDRs (Figs. 12 and 13). Considering their high percental disorder, it is likely they share the same situation as PDPs lacking long IDRs – they possess segments of contiguous disordered residues that are below the 20 amino acid threshold.

5.2.4. Amino acid compositional analyses

As has been already stated, intrinsic order and disorder of proteins are written down and encoded in their primary, amino acid sequences (Uversky et al., 2000). Taking into account the inherent properties of different amino acids, they can be categorized as either order- (W, F, Y, I, M, L, V, N, C and T) or disorder-promoting (A, G, R, D, H, Q, K, S, E and P). The most order-promoting amino acid is considered to be tryptophan, while proline is the most disorder-promoting one (Campen et al., 2008). Therefore, compositional analysis of identified proteins can provide valuable information on the distribution of individual residues in the data set and help their further characterization. When it comes to MDPs and PDPs identified in this study, over two-thirds of amino acids in their primary sequences are disorder-promoting (68.9% and 64.9%, respectively, Fig. 14). Going into more detail, it can be seen that these proteins are particularly enriched in polar amino acids such as glutamic acid (E, 11.91% and 11.64%, respectively), lysine (K, 9.34% and 9.26%) and glutamine (Q, 6.51% and 5.26%), when compared with the other two groups of more ordered proteins. Additionally, MDPs contain at least double the amount of proline in comparison to other protein groups – 7.95% versus 3.83%, 4.25% and 4.27% in PDPs, NOPs and OPs, respectively (Fig. 15A).

On the side of order-promoting amino acids, an interesting observation can be made regarding leucine. On the disorder-order scale, leucine is situated in the middle of order-promoting amino acids (Campen et al., 2008). As such, it can be surprising to see that PDPs are conspicuously enriched in this amino acid, with leucine content being almost on par with OPs (Fig. 15B). However, the compositional analysis has shown that certain disorder-containing muscle proteins contain significant amounts of this amino acid. In these proteins, such as tropomyosins, leucine accounts for at least 10% of all amino acids. This can be indicative of the importance of leucine for the function of muscle proteins. Indeed, tropomyosins regulate the contraction of muscles and

contain leucine zippers in their C-termini (Brown et al., 2005), in a similar vein to transcription factors.

5.2.4.1. Glutamic acid, lysine, glutamine and proline

Glutamic acid being the most present amino acid in the data set of this study is in line with its overall importance in protein intrinsic disorder. It is the second most disorder-promoting amino acid (Campen et al., 2008) and, due to its polar and charged properties, an important building block of proteins and polypeptides. In ordered proteins, glutamic acid is involved in the formation of salt bridges and is one of the amino acids that favor formation of α-helical secondary structures (Uversky, 2013a). In IDPs/IDRs, on the other hand, the comparatively increased content of this negatively charged and hydrophilic amino acid, combined with the depletion of hydrophobic and aromatic residues, contributes to electrostatic repulsion events and prevents the usual folding of proteins and/or protein regions under physiological conditions (Uversky et al., 2000; Radivojac et al., 2007). Apart from its structure-defining roles, glutamic acid is important for the function of different molecular chaperones. C-termini of these proteins are particularly enriched in acidic residues such as glutamic acid, which facilitates electrostatic interactions between the C-termini of molecular chaperones and basic acid-rich regions of their respective target proteins (Sarkar et al., 2001; Narberhaus, 2002; Park et al., 2002). Additionally, proteins can undergo polyglutamylation, where polyglutamate chains are post-translationally added to their C-termini to regulate their function. This type of post-translational modifications is frequent in intrinsic disorder-containing microtubular proteins that, through their C-termini, bind different structural and motor proteins associated with the cytoskeletal system (Young and Ajami, 2000; Janke et al., 2008; Cao et al., 2020).

Lysine is the second most abundant amino acid present in proteins of this study, and also the fourth most disorder-promoting (Campen et al., 2008). This high content of lysine accentuates the importance of the amino acid when it comes to protein function. Interfacing of proteins with membranes is facilitated by electrostatic interactions between negatively charged lipids and positively charged/basic lysine residues (McMahon and Gallop, 2005; Okada et al., 2021). When it comes to IDPs specifically, enrichment in lysine has been linked to retained protein functionality at low temperatures, as well as protection from molecular damage caused by desiccation. In the *Pseudoalteromonas haloplanktis* α-amylase, a protein model for studying adaption to cold

temperatures, lysine residues in one of the active site domains provide the enzyme with enough flexibility to retain molecular motion and enzymatic activity in a cold environment (Siddiqui et al., 2006). In plant dehydrins, intrinsically disordered proteins that shield molecules and membranes from desiccation-induced damage, lysine-rich segments are involved in binding dehydrins with target membranes (Petersen et al., 2012; Clarke et al., 2015). Additionally, lysine is one of the amino acids that are most often targeted by post-translational modifications. In particular, lysine residues are frequently enzymatically modified by acetylation and methylation, while non-enzymatic polyphosphorylation has also been reported. These modifications regulate the activity and localization of proteins that are involved in diverse processes, such as gene expression, protein-protein interactions, signal transduction and regulation of development (Choudhary et al., 2009; Winter et al., 2015; Lazar et al., 2016; Ukmar-Godec et al., 2019; Semenyuk, 2021).

One of the hallmarks of intrinsic disorder in proteins is low sequence complexity, as these proteins and their regions often contain sequences which are repeats of singular amino acids or motifs. The low complexity regions are unable to form ordered structures as they lack the necessary amounts of hydrophobic residues (Romero et al., 2001). Glutamine, which was determined here to be abundant in disorder-containing proteins in contrast to more ordered proteins (Fig. 15A), is one amino acid that is frequently found in these low complexity regions. Notably, such glutamine-rich sequences are often a characteristic of transcription factors (Dyson and Wright, 2005), as well as other proteins that are involved in processes of cell signaling, regulation of gene expression and intermolecular interactions. These low complexity regions enriched in glutamine residues allow proteins to interact with their intended targets, while their disordered structure and subsequently function can be modified and regulated by flanking sequences (Hibino et al., 2016; Cooper and Fassler, 2019; Urbanek et al., 2020; Dyson and Wright, 2021; Kapur et al., 2022).

One amino acid that stood out the most in PDPs compared to all other protein groups was the highly disorder-promoting proline. Due to its physical properties, it is known to break up secondary structures in proteins (Imai and Mitaku, 2005; Morgan and Rubenstein, 2013). Apart from its structure disrupting role, proline can in certain conditions introduce rigidity to protein conformations and enable binding functionality of proteins. Proline-rich motifs can often form polyproline type II (PPII) helixes which spatially separate different functional regions of a protein

(Theillet et al., 2013). In many scaffold proteins for example, which contain a high amount of intrinsic disorder, PPII helixes are abundant and through them scaffold proteins can recognize and interact with different binding partners (Dosztányi et al., 2006; Cortese et al., 2008). Thus, proline is important in protein intrinsic disorder not only to ensure that proteins remain unstructured and primed for interaction events, but to enable and facilitate those interactions as well.

5.2.4.2. Compositional bias

Compositional analyses such as these are usually performed in the context of specific proteins of interest. However, there are indications that the amino acid composition of intrinsically disordered proteins could have species-specific bias. In their study of IDPs in the green algae *Chlamydomonas reinhardtii*, Zhang and coauthors (2018) have reported that IDPs that were predicted for this species have a biased amino acid composition when compared to "classical" IDPs available in the DisProt database. In particular, the green algae IDPs were depleted in disorder-promoting residues such as glutamic acid, lysine, proline, glutamine and aspartic acid. They have also compared the amino acid composition of experimentally validated IDPs with available green algae proteome data and have found that the IDPs were depleted in serine, threonine, asparagine and leucine residues. Conversely, the content of these amino acids in PDPs and MDPs from our study is generally similar to that of NOPs and OPs (Fig. 15). Taken altogether, these findings could point to the existence of a species-specific compositional bias, however strengthening that claim would require for more studies to take this approach in analyzing intrinsically disordered proteins.

5.2.5. Functional analysis of disordered proteins

What immediately stands out, after the functional analysis of proteins identified in this study, is how significant the gap in functional annotation still is, for ordered and disordered proteins alike. Out of the 892 unique proteins identified here, barely 50% of them are listed in Uniprot and annotated with data from at least one other major annotation databases, such as e.g. Pfam. When the focus is narrowed down to just the disorder-containing proteins, the gap further increases as only 143 such proteins could be covered by this analysis. Low annotation efforts are likely further exacerbated by the fact that lepidopteran insects do not have many representative model species, apart from the silkworm *Bombyx mori* (International Silkworm Genome Consortium, 2008). As such, these findings could serve to motivate and direct future efforts to lessen the gap in functional knowledge of both ordered and disordered proteins in insects.

Here it was shown that the identified proteins perform a wide array of molecular functions and are involved in very diverse biological processes. The vast majority of them (40) are linked with the cytoskeletal network, either as its structural constituents or regulators of muscle fiber contraction. Following are proteins acting as molecular chaperones (32) that either assist with the proper folding of nascent proteins or handle misfolded ones. The last two major groups of proteins are involved in protein and amino acid metabolism (18), regulation of translational processes such as elongation, or they are structural components of ribosomes (15). The rest of the proteins have a plethora of functions and cover various metabolic processes, such as the metabolism of carbohydrates and lipids, insect cuticle formation, binding of nucleic acids, as well as oxidoreductive processes in the electron transport chain (Figs. 16 and 17). Heating the samples also allowed for novel functions and processes to be identified. Proteins that are involved in the formation of insect cuticle were found exclusively in these samples. Similarly the content of proteins that are a part of the cytoskeletal network or are involved in binding nucleic acids was also enriched following sample heating (Fig. 16, Heated). Lastly, sample heating allowed for many of the proteins in the Other category to be uncovered. The number of identified proteins almost doubled after the treatment (Fig. 17). As with the previous proteomic and bioinformatics analyses, these findings again highlight the importance of heat treating the samples in order to more completely identify and functionally analyze intrinsically disordered proteins.

5.3. Ecophysiological aspects of proteins with intrinsic disorder in cold hardiness

As has been mentioned in the Introduction, proteins containing intrinsic disorder possess some unique physicochemical properties that are distinct from those found in ordered, globular proteins. Due to these properties, which stem from the protein's amino acid composition, IDPs and ID-containing proteins are more resilient to adverse environmental factors than ordered proteins. The intrinsic disorder that is present in their structures allows them to either retain their activity and functionality under such conditions, or to undergo disorder-to-order, or vice-versa, transitions which, again, have the same end result of keeping the proteins functional. Because of that, ID-containing proteins are likely majorly involved in cold adaptation processes which allow organisms to survive when outside temperatures fall even well below the freezing point of water. The European corn borer, *O. nubilalis*, was selected as a model system to explore this aspect of intrinsic disorder in proteins, as this species has been frequently used in studies pertaining to adaptations that allow organisms to survive under hypometabolic and harsh environmental conditions, such as those caused by temperatures close to or below 0°C. This approach aims to further both the functional knowledge on intrinsically disordered proteins, as well as help elucidate the molecular mechanisms that govern stress adaptation processes in a cold-adapted insect species.

5.3.1. IDP content in *O. nubilalis* cold hardiness

Organisms undergo various biochemical, physiological and molecular changes as part of their adaptive processes to perturbations in the environment. These adaptations can often also lead to changes in the content of an organism's proteome. Considering the diversity of functions that ID-containing proteins possess, it was prudent to determine whether the content of these biomolecules is affected by cold adaptation processes, and to what extent. This issue was approached from two directions, mirroring how disorder content was evaluated for the proteins that were identified in this study. Potential changes in IDP content due to cold acclimation were first assessed from the point of percental intrinsic disorder, i.e. the content of MDPs, PDPs, NOPs and OPs in the different experimental groups. In both the Pilot and Main experimental setups it can be seen that diapause and cold acclimation had a negative effect on the number of proteins that could be detected and identified (Fig. 6). As such, the amount of proteins with considerable

percental disorder (MDPs and PDPs) was also lower in the cold-acclimated groups (Fig. 8). That said, the percental distribution of the four different protein categories – MDPs, PDPs, NOPs and OPs, relative to the total number of proteins, was actually fairly similar between the experimental groups (Fig. 9). Therefore, the content of IDPs, at least according to the percental disorder criteria, did not seem to be affected by cold acclimation. However, protein intrinsic disorder in this study was also assessed by determining how many proteins (regardless of percental disorder) contain long intrinsically disordered regions, which have a higher potential for biological relevance. When looking at the distribution of proteins containing long IDRs in the different experimental groups, the effects of diapause and cold acclimation on IDP content are more evident and pronounced (Figs. 12 and 13). Most notably, the percentage of NOPs that contain at least one long IDR was increased significantly in the diapausing, cold-acclimated groups, at the expense of NOPs without long IDRs. With the exception of OPs, the percentage of proteins containing long IDRs was higher in the aforementioned experimental groups, especially when comparing the groups from the Main setup to the non-diapausing control (Fig. 13A-E). This could be particularly significant for NOPs, which are close to being considered ordered. Intrinsic disorder generally increases the flexibility of polypeptide chains, which is a favorable trait for protein functionality in cold environments (Siddiqui et al., 2006). Possessing at least one long IDR would make NOPs less rigid, allowing them to retain more of their molecular mobility and preserve their functionality at low temperatures. As such, the cold adaptation processes in the ECB seem to affect the IDP content more in a qualitative manner, rather than quantitative.

These findings are in line with the notion that environmental pressure can affect proteome composition in regards to intrinsically disordered proteins. In a large study covering 46 fully sequenced prokaryotes that inhabit different mesophilic and extreme habitats, it was shown that there is a correlation between the extremity of the environment and protein intrinsic disorder measured in the percentage of proteins containing long IDRs (Vicedo et al., 2015). Additionally, an interesting observation was made that non-related species sharing the same type of habitat have similar protein disorder content, as opposed to species that are related but inhabit drastically different environments (mesophiles versus extremophiles). Higher than average disorder content was predicted for non-related species of archaea that thrive in habitats saturated with salts (*Halobacterium sp. NRC-1* and *Haloarcula marismortui ATCC 43049*). On the other hand, higher disorder content was predicted for the halophilic *Marinobacter aquaeolei VT8* bacterium

compared to its mesophilic taxonomic relative *Pseudoalteromonas atlantica T6c*. When it comes to hot and cold environments, it was found that thermophiles and psychrophiles actually have reduced protein disorder content compared to mesophilic species, which seems to be an adaptive trait of prokaryotes (Burra et al., 2010; Vicedo et al., 2015). While those findings seem in contrast to the results obtained for IDP content in *O. nubilalis* in this study, analysis of homologous proteins from two opposing extremophiles – the psychrophilic *Colwellia psychrerythraea 34H* and the hyperthermophilic *Pyrococcus horikoshii OT3*, has shown that more homologs from *Colwellia* contain long IDRs compared to the ones from *Pyrococcus* (Vicedo et al., 2015). This reinforces the hypothesis that increased protein flexibility, by means of higher content of long IDRs, is an adaptation necessary for them to function in cold environments. Similar studies on the relationship between protein intrinsic disorder and environmental temperatures in eukaryotes, however, are lacking. In the plant model species *Arabidopsis thaliana*, heat stress lead to the expression of heat-induced proteins that were enriched in electrostatically charged amino acids, while being depleted in hydrophobic and polar residues. Additionally, they were also enriched in long IDRs (Alvarez-Ponce, 2018). Such amino acid and long IDR compositions are in contrast to proteins found in thermophilic bacteria and archaea, which can point to different strategies for temperature adaptations on the protein level between eukaryotes and prokaryotes. When it comes to animal studies, the influence of different habitat temperature regimes on cold adaptation was explored in two closely related zoarcid fish species. The eurythermal *Zoarces viviparus* thrives at 15°C, but can experience and survive temperatures as low as 0°C due to seasonal fluctuations (Pörtner and Knust, 2007). The stenothermal *Pachycara brachycephalum*, on the other hand, is adapted to living at a constant 0°C (Brodte et al., 2006). Amino acid compositional analysis of orthologous sequences from these two species has shown that there is a pattern of distinct amino acid substitutions in the cold-adapted *Pachycara brachycephalum*. In particular, the frequency of acidic residues, such as glutamic acid, was reduced, while the frequencies of basic amino acids was elevated. A net result of these substitutions is a reduction in the number of salt bridges that can be formed, which would lead to increased flexibility of the polypeptide chain, and better interactions with solvents on the protein surface (Windisch et al., 2012). While this study did not explore the mechanisms of cold adaptation on the level of intrinsically disordered proteins, it has demonstrated more evidence for the protein flexibility hypothesis under cold environmental conditions.

5.3.2. Gene expression analyses of disorder-containing proteins

In order to determine if and how the results of proteomic analyses correlate with changes on the transcriptional level, relative expression of genes encoding different ID-containing proteins that are implicated in diapause and cold adaptation processes was explored. In total, 9 genes were analyzed – five encoding HSPs (*hsp90, hsc70, hsp70, hsp20.4* and *hsp20.1*), two encoding structural proteins (*tropmy2* and *tnt*), one encoding a protein involved in insect immune response (*thym*) and one encoding a member of the ERM (ezrin, radixin and moesin) family of proteins (*moe*). All of the proteins encoded by the genes analyzed here were determined to contain at least 10% intrinsic disorder and/or at least one long intrinsically disordered protein region. Structural proteins, small HSPs and thymosin beta were over 30% disordered and considered PDPs (disorder percentage between 30% and 70%). However, HSP20.1 was the only one without long IDRs. It should be noted that, due to its high ID content, it is likely that this small heat shock protein possesses a segment of consecutive disordered residues that is just shy of the 20 amino acid threshold to be regarded as a long IDR. Additionally, HSP20.4 was determined to be 87% disordered, however only a partial protein sequence was available for this analysis. It is likely that this disorder percentage is inflated because of the missing residues and should be closer to that of HSP20.1 (~46%). Lastly, the three larger HSPs (HSP90, HSC70 and HSP70) were determined to be NOPs (disorder percentage between 10% and 30%) that contain at least one long IDR. HSP70 contained the most disorder out of the three, with 21% disorder percentage and two long IDRs.

5.2.4.1. Heat shock protein genes

Molecular chaperones known as heat shock proteins play important roles in the homeostasis of biological systems. They are responsible for ensuring that newly synthesized proteins properly fold into their mature conformations, as well as handling abnormal proteins either by correcting their misfolded structures or directing them towards protein degradation and recycling pathways (MacRae, 2010; King and MacRae, 2015; Storey and Storey, 2022). HSPs also transport nascent proteins across organelle membranes to their subcellular destinations, where they spontaneously fold into their correct conformations (Saibil, 2013). Under physiological conditions, HSPs are moderately expressed because of the said role in protein maturation. However, their expression and accumulation increase when an organism is faced with adverse environmental conditions, such as high or low temperatures, exposure to UV radiation, pressure changes, lack of oxygen or water,

and presence of pollutants. Heat shock proteins alleviate the damaging effects of these abiotic stressors by preserving the structure and function of all manner of affected proteins (Hochachka and Somero, 2002; Kregel, 2002; King and MacRae, 2015).

All analyzed HSP genes, excluding *hsp20.1*, have shown a similar pattern of expression in this study. Their relative expression levels were much higher in cold-acclimated diapausing groups when compared to the non-diapausing control group, while for *hsp20.1* the opposite was the case. Additionally, it can be seen that the expression of all these genes gradually increased in the diapausing groups as diapause progressed and the larvae were exposed to increasingly colder temperatures. The only exception is the *hsc70* gene, which exhibited equally high expression in all diapausing groups regardless of acclimation temperature.

hsp90

Relative expression of the *hsp90* gene was markedly high in all experimental groups, with the exception of the D(15) group where it was heavily downregulated. Additionally, expression of *hsp90* was higher in the cold-acclimated groups in comparison to the non-diapausing control, and it rose as larvae were exposed to gradually colder temperatures and diapause progressed (Fig. 19). These results are in accordance with previous studies that have analyzed the expression of *hsp90* in insects that undergo developmental arrest and/or experience prolonged periods of cold environmental temperatures as part of their life cycle. In several leafminers of the *Liriomyza* genus, exposure to the cold induces the expression of *hsp90* (Huang and Kang, 2007; Huang et al., 2009). It can also be induced during the resting state of diapause, as was detected in pupae of the onion fly *Delia antiqua* (Chen et al., 2005), as well as larvae of the rice borer *Chilo suppresalis*. Interestingly, the high expression of *hsp90* in diapausing rice borer larvae was not further induced by exposure to cold temperatures, as opposed to the non-diapausing larvae (Sonoda et al., 2006). On the other hand, cold shock upregulated *hsp90* expression in both diapausing and non-diapausing pupae of the flesh fly *Sarcophaga crassipalpis* (Rinehart and Denlinger, 2000). Additionally, in the flesh fly pupae, as well as those of *Helicoverpa zea*, it was detected that *hsp90* expression increases towards the end of diapause and during post-diapausal development (Rinehart and Denlinger, 2000; Zhang and Denlinger, 2010).

HSP90 activity is modulated and directed by interactions with a great number of co-chaperones. These interactions enable HSP90 to be involved in many molecular processes, from protein folding

and maturation, modification of transcription factors and various kinases, to regulation of cell communication. HSP90 also differs from other HSPs as it tends to bind and interact with proteins that are already partially folded and help them overcome their intrinsic instability in order to fold into an active conformation, even taking over substrates from HSP70 to do so (MacRae, 2010; Röhl et al., 2013). Taking all of this into consideration, increased *hsp90* expression during diapause and cold acclimation in the ECB is to be expected in order to ensure that the aforementioned processes continue unimpeded in an unfavorable metabolic landscape caused by hypometabolism and exposure to cold temperatures. High expression towards the end of diapause could also be necessary to prepare the larvae for resumption of development and intense protein synthesis that will follow.

hsc70 and hsp70

Similarly to *hsp90*, expression of both the constitutive *hsc70* and the inducible *hsp70* genes was higher in the cold-acclimated diapausing larvae when compared to the non-diapausing control. The expression of *hsp70* gradually increased as diapause progressed and temperatures dropped, while *hsc70* expression remained at a similar level regardless of diapause status and acclimation temperatures (Figs. 20 and 21). These findings on the expression of the constitutive *hsc70* are in accordance with previous similar studies done on the ECB (Popović 2014; Popović et al., 2015). Considering that it does not seem to be affected by acclimation temperatures, increased *hsc70* expression could be reflective of the metabolic changes that occur in diapausing larvae that differentiate them from their non-diapausing counterparts. Looking at other insect species, there does not seem to be a conclusive pattern to *hsc70* expression. During the adult diapause of the cold-hardy northern malt fly *Drosophila montana* expression of *hsc70* is downregulated (Kankare et al., 2010). Conversely, increase in *hsc70* expression was detected over the timecourse of diapause in larvae of the corn stalk borer *Sesamia nonagrioides* (Gkouvitsas et al., 2009), as well as in the second half of larval diapause in the bamboo borer *Omphisa fuscidentalis* (Tungjitwitayakul et al., 2008). Cold stress amplified the expression of *hsc70* in the fat body of non-diapausing silkworm larvae (Fang et al., 2021), while downregulating it in non-diapausing adults of the European firebug *Pyrrhocoris apterus* (Koštál and Tollarová-Borovanska, 2009). Expression of *hsp70*, on the other hand, showed more consistent responsiveness to abiotic stressors, befitting its inducible nature. In the aforementioned firebug, cold stress amplified *hsp70*

expression (Koštál and Tollarová-Borovanska, 2009), as well as in the *Liriomyza* leafminers (Huang and Kang, 2007; Huang et al., 2009). Expression of *hsp70* is also upregulated during larval diapause of the gall fly *Eurosta solidaginis* in late fall and winter periods (Zhang et al., 2011). It is also upregulated during larval diapause in the Antarctic midge *Belgica antarctica*. HSP70 has also been determined to be critical for survival of this species in the extremely cold environment of the Antarctic (Rinehart et al., 2006, 2007).

Considering that HSP90, HSC70 and HSP70 belong to the group of ATP-dependent heat shock proteins, this can call into question what their actual role is in metabolic states when energy sources and reserves are limited, such as during the hypometabolic state of diapause. When ATP is not a limiting factor, HSPs take up abnormal proteins and assist in their folding/refolding, or interact with different receptors and kinases to modulate cellular signaling (King and MacRae, 2015). However, under hypometabolic conditions (i.e. diapause) when ATP is limited (Popović et al., 2021), HSPs are more likely to remain in an ADP-bound state which favors substrate binding and protective sequestering until energy metabolism recovers (King and MacRae, 2015). For example, when HSP70-bound ATP is hydrolyzed, HSP70 transitions into an ADP-bound state with strong substrate affinity, while substrate dissociation rates are decreased. In such a state, HSP70 has high affinity for ATP, rebinding of which causes the substrate to release from the chaperone (Mayer et al., 2000; Arakawa et al., 2011; Mayer and Gierasch, 2019; Wu et al., 2020).

Increased expression of *hsp90*, *hsp70* and *hsc70* in late diapause (D(–16) group, Figs. 19, 20 and 21) could also have positive effects on post-diapausal development of ECB larvae. Once diapause terminates and regular development resumes, the high content of different HSP transcripts would provide the now active larvae with a pool of adequate amounts of HSPs to assist in the folding of newly synthesized proteins (Popović et al., 2015). Additionally, this would also allow the HSPs to regulate their own expression and decrease it by binding to heat shock factors (HSFs). Monomer HSFs are transcription factors that are usually bound to different HSPs, such as HSP90, HSP70 and HSC70, under normometabolic conditions (Zou et al., 1998; Guo et al., 2001; Zheng et al., 2016). When stressors cause proteins to unfold, the damaged proteins compete with HSFs for binding sites on HSPs. This causes HSF monomers to detach and trimerize, after which they are translocated into the nucleus and bind to heat shock response elements (HREs) to upregulate transcription of HSP genes (Hochachka and Somero, 2002). Once stressors subside and

normometabolism is reestablished, HSPs are free to rebind HSFs and inhibit their transcription pathway, freeing up energy to be used for other metabolic processes.

hsp20.4 and hsp20.1

Small heat shock proteins (sHSPs) belong to a group of molecular chaperones that share several structural commonalities. Their C-termini contain a conserved α-crystallin domain typical for sHSPs, while the secondary structure features a conserved β-sheet structure. The α-crystallin domain is necessary for the chaperone function of sHSPs, while the β-sheet enables the oligomerization of individual sHSPs. True to their name, these chaperones are smaller than other HSPs, with their molecular weights ranging from 12 to 43 kDa (Haslbeck, et al., 2005; Stetler et al., 2010). The majority of identified sHSPs in insects do not contain introns in their sequences, which is believed to facilitate rapid stress-responses. Lack of introns facilitates the rapidity of the response, as such mRNA do not undergo post-transcriptional processing (Li et al., 2009; King and MacRae, 2015). This type of stress response is reminiscent of prokaryotic organisms, which feature simpler transcriptional and translational machinery which can and must respond quickly to changes in the environment in order to ensure their survival. Small HSPs are incredibly diverse in number, size and molecular processes they are involved in. This diversity of sHSPs allows organisms to finely tune their cellular responses to stressful factors, as can be witnessed in the great adaptability of insects to almost any type of environment (Li et al., 2009). Small HSPs are the first line of defense against protein damage induced by stress factors, preventing aggregation and subsequent irreversible denaturation of substrate proteins. They bind abnormal proteins in an ATP-independent manner until they can be taken over by one of the members of the larger HSP families for refolding (King and MacRae, 2015; Żwirowski et al., 2017).

Analyses in this study have revealed that the two small HSPs – *hsp20.4* and *hsp20.1*, have opposite expression patterns in regards to diapause and acclimation to cold temperatures. Expression of *hsp20.4* was higher in the diapausing groups when compared to the non-diapausing control. It steadily increased throughout the diapause and as temperatures dropped, with a slight decrease towards the end of the resting phase (Fig. 22). Conversely, *hsp20.1* was highly upregulated in the non-diapausing group in comparison to any of the diapausing, cold-acclimated groups (Fig. 23). Such results could indicate that *hsp20.4* and *hsp20.1* have distinct functional roles in the different development phases and responses to stress in *O. nubilalis*. Additionally, of the two sHSPs,

presence of long IDRs was detected only in the one that was upregulated during diapause and cold acclimation – *hsp20.4*, pointing to the importance of increased polypeptide chain flexibility under hypometabolic and environmentally adverse conditions.

Small HSPs are particularly diverse in number and, as such, there does not seem to be a discernable pattern of expression between different insect species. Expression depends on a plethora of different factors, including developmental stage, type of stress, sex, tissue, to name a few. For example, expression of a small HSP is elevated near the end of adult diapause in the mosquito *Culex pipiens* (Robich et al., 2007). In *S. nonagrioides* larvae expression of *hsp20.8* is variable during the timecourse of diapause, however it increases near the end of the resting phase (Gkouvitsas et al., 2008). Some small HSPs are induced by cold temperatures, such as those from the cigarette beetle *Lasioderma serricorne*. In this species, several sHSPs (*LsHsp20.2*, *LsHsp 20.3* and *LsHsp22.2*) were upregulated in cold-challenged adults, while another – *LsHsp19.4*, was unaffected (Yang et al., 2019). In the fall armyworm *Spodoptera frugiperda* expression of several sHSPs differed between male and female adults when exposed to heat and cold stresses – *SfsHsp19.3*, *SfsHsp20*, *SfsHsp21.3* and *SfsHsp29* were upregulated in both heat- and cold-treated male adults. When it comes to female adults, both temperature treatments induced the expression of *SfsHsp20.1* (Yang et al., 2021). This diversity of expression patterns is reflective of the importance that sHSPs play in many different cellular processes (removal of non-functional proteins and prevention of protein aggregation, regulation of cell division and death, transport of newly synthesized polypeptides) during various stages of development, as well as adaptations to abiotic stressors.

5.2.4.2. Structural protein genes

Transcriptomic studies have revealed that structural protein genes are often differentially expressed in insects and other arthropods that enter diapause and exhibit cold hardening as part of their life cycle (Li et al., 2009; Pavlides et al., 2011; Teets, 2013; Bryon et al., 2017). However, the significance and impact of changes in the expression of cytoskeletal genes, in regards to cold adaptation and diapause processes, is not fully understood. Structural proteins could be involved in these processes on several levels, from regulating mobility in response to the changing environment and energy conservation, ensuring cell membrane elasticity at low temperatures, to

facilitating cellular communication and signaling (Rinehart et al., 2007; Clark et al., 2012; Bryon et al., 2017).

In this study, the expression of two genes encoding muscle-related structural proteins was analyzed – tropomyosin-2 (*tropmy2*) and troponin T (*tnt*). These proteins are involved in regulating the contraction of muscles. Tropomyosin-2 belongs to a family of actin-binding proteins that regulate the interaction between actin and myosin and, in turn, muscle contraction. Apart from this regulatory role, tropomyosins are major contributors to the many functions of the actin cytoskeleton network (Gunning et al., 2008). Troponin T is one the three subunits that make up the troponin complex. It is an anchoring protein that interacts with the other two troponin components – the calcium-binding troponin C and the inhibitory troponin I, binds tropomyosin and anchors them all to thin actin filaments (Wei and Jin, 2016; Cao et al., 2019). When muscle fibers are at rest tropomyosin blocks myosin-binding sites on actin and prevents contractions. In order for the muscle to contract, the troponin complex interacts with tropomyosin to uncover the myosin-binding sites. This allows myosin and actin to bind together and initiate muscle contraction (Gunning et al., 2008; Pavadai et al., 2020).

Expressional patterns were contrasted between the two structural protein genes (Figs. 24 and 25). In the actively developing, non-diapausing larvae *tropmy2* was highly downregulated, while *tnt* was noticeably upregulated. Looking at the diapausing, cold-acclimated groups, the expressional patterns are reversed – *tropmy2* was upregulated, while *tnt* was downregulated. Expression of *tropmy2* was already upregulated at the beginning of diapause, in the D(15) group, and remained at the same level until the end of diapause when it gradually decreases. When it comes to *tnt* expression, it gradually decreased as diapause progressed and the larvae acclimated themselves to lower temperatures. These results could elucidate the molecular mechanisms behind one of the characteristics of organisms in diapause – cessation of movement. Elevated *tropmy2* and decreased *tnt* expression would ensure that muscle contractions do not occur and energy is not needlessly expended. As has been said, energy conservation is an important aspect of diapause and concurrent cold acclimation because organisms do not feed during that time and cannot replenish their energy reserves. Since tropomyosins are also associated with the actin cytoskeleton, which is responsible for a wide array of cellular processes such as intracellular transport, regulation of cell size and

morphology, upregulation of *tropmy2* during diapause and cold-hardening likely has effects other than just prevention of mobility to conserve energy.

The results of *tropmy2* expression analysis are in line with previous studies that were undertaken on other diapausing and cold-hardy species. Gene expression of a tropomyosin isoform was upregulated during the summer diapause in *D. antiqua* pupae, along with several other cytoskeletal proteins (Hao et al., 2012). Similarly, in the two-spotted spider mite *Tetranychus urticae*, *tropmy* was upregulated in overwintering adult females. Additionally, many Ca^{2+}-binding and transporter genes (including for the calcium-binding troponin C) were also upregulated in these individuals, suggesting that Ca^{2+} signaling is important for regulation of diapause in this species (Bryon et al., 2017; Zhao et al., 2017). Upregulation of the tropomyosin-encoding gene was also detected in overwintering, diapausing adults of the mosquito *C. pipiens pallens* (Zhang et al., 2019). Rapid cold hardening, a process where short-term exposure to low temperatures increases an organism's cold resistance, also induces changes in the expression of structural proteins. A tropomyosin isoform increased in abundance in the brain of *S. crassipalpis* adults after a 2h exposure at 0°C, while several other cytoskeletal proteins were less abundant (Li and Denlinger, 2008). In some species cold hardening is achieved by preemptive desiccation where water loss is encouraged in order to avoid the formation of damaging ice crystals in bodily fluids. In such cases, cytoskeletal reorganization has been reported to occur. Cryoprotective desiccation in the Antarctic midge *B. antarctica* increased the abundance of a dozen cytoskeletal protein, one of which is tropomyosin (Li et al., 2009).

According to literature, expression of *tnt* has not been explored in relation to diapause and cold hardening. However, changes in mRNA levels of this gene have been linked to the nutritional status of individuals. For example, adults of the tobacco hornworm *Manduca sexta* showed greater flight metabolic rates when their larvae were reared on inbred plants that were susceptible to infestation by the hornworm. This increase in flight metabolic rate was correlated with alternative splicing of troponin T (Portman et al., 2014). Conversely, in the fall armyworm *S. frugiperda*, poor larval diet resulted in upregulation of troponin T isoforms that are linked with reduced muscle activity and energy consumption in adults. Higher performance troponin T isoforms, similar to the one analyzed in this dissertation (*tnt*) were decreased in abundance in these experimental setups (Marden et al., 2018; Portman et al., 2020). There were cases when proper feeding had a negative

effect on the abundance of troponin transcripts. Flight activity of female blood-fed mosquitos (*Anopheles gambiae*) is inhibited as they prepare to lay eggs, which is reflected on the downregulation of troponin transcripts (Marinotti et al., 2006). Considering that organisms that enter diapause do not feed during that time, this could help to better understand the underlying molecular mechanisms behind the reduction of insect mobility during this resting phase.

5.2.4.3. Thymosin beta and moesin genes

Thymosin beta, or β-thymosin, belongs to a family of heat-stable, multifunctional polypeptides that were first discovered in calf thymus (Goldstein et al., 1966). This group of thymosins participates in cell migration, sequestering of actin fibers, anti-inflammatory processes and intracellular signaling (Bock-Marquette et al., 2009; Xue et al., 2014; Kleinman and Sosne, 2016; Marks and Kumar, 2016). In invertebrates, however, they are notable for their roles in the humoral response to microbial infection (Schillaci et al., 2012; Feng et al., 2018; Liao et al., 2018).

Expression of the thymosin beta gene, *thym*, in larvae of the ECB was much higher in the actively developing ND group compared to any of the diapausing groups (Fig. 26). These results suggest that thymus beta has a more important role in ECB immunity during the active, feeding phase of the insect’s life cycle, compared to its importance during diapause. Particularly, thymosin beta might be important for antimicrobial defense during insect metamorphosis, as it is upregulated by the hormone 20-hydroxyecdysone which controls arthropod molting and metamorphosis (Zhang et al., 2011). Thymosin beta protein levels were also highly elevated in silkworm pupae (Ma et al., 2015), further corroborating the role of specific thymosins during insect active development. When it comes to their role in diapause and cold acclimation processes, studies of these aspects are sparse. Cold response of the Colorado potato beetle *Leptinotarsa decemlineata* was investigated with a transcriptomic and proteomic approach. Elevated protein levels of thymosin beta were detected in beetles exposed to –5°C compared to non-chilled adults, while its transcript abundance was not explored (Govaere et al., 2019). While the results of that study are contrasted with the expressional analysis performed in the ECB, where *thym* was downregulated in cold-acclimated groups, it should be noted that measured protein levels do not necessarily have to correlate with detected transcript levels. Increased transcription in one phase of an organism’s life cycle could provide a pool of transcripts ready for translation in a subsequent phase of development. That said, this study on the potato beetle (Govaere et al., 2019) was undertaken in laboratory conditions that did not

mimic the natural setting an adult beetle would encounter. The Colorado potato beetle is a freeze-avoiding insect, and it overwinters during its adult stage by burrowing up to 50 cm in the ground to protect itself from the cold (Senanayake et al., 2000). As such, it does not experience the same degree of cold stress as diapausing ECB larvae that overwinter in corn stalks that are only a few centimeters across. This discrepancy can lead to different responses to cold stress on the molecular level, such as the one observed for protein and transcript levels of thymosin beta.

Moesin, shorthand for moesin/ezrin/radixin homolog 1, is one of the three highly conserved, closely related and homologous ERM family proteins. They are cytoskeletal proteins that mediate communication between the cortical actin cytoskeleton and proteins present in the plasma membrane. ERM proteins also regulate a variety of signaling pathways by assembling multiprotein complexes at the interface points between the membranes and cytoskeleton (Fehon et al., 2010). Most invertebrate species contain just one ERM-like gene, such as the gene encoding moesin in *D. melanogaster* (Shabardina et al., 2019; Bajusz et al., 2021).

Gene expressional analysis has shown that *moe* was upregulated in the non-diapausing, as well as diapausing larvae acclimated to −16°C, in comparison to the other diapausing groups. This indicates that increased transcription of *moe* is necessary both during active development and under hypometabolic conditions when larvae are exposed to temperatures well below the freezing point of water. In the diapausing groups in general, expression was the lowest at the beginning of diapause and gradually increased as diapause progressed with concurrent lowering of ambient temperatures (Fig. 27). As such, *moe* seems to be implicated in the multifactorial cold adaptive process of the ECB. In insects, moesin has been identified as an important functional component of the nucleus because it participates in the transport of mRNA (Kristó et al., 2017). Dysregulation of moesin functionality and activity leads to developmental and physiological abnormalities, as well as sterility. It is also involved in regulating the expression of hundreds of genes, in particular genes encoding heat shock proteins. In *D. melanogaster* knocking out moesin severely reduces the expression of *hsp* genes (Bajusz et al., 2021). As has been said, heat shock proteins are essential both for actively developing organisms – to ensure the proper synthesis of proteins, as well as those experiencing adverse environmental conditions – to improve the organism's resistance to those abiotic stressors. Therefore, the determined expression patterns of *moe* in this ECB study are aligned with the functional importance of moesin. Additionally, it can be seen that the expression

patterns of different *hsp* genes in the diapausing groups (Figs. 19, 21 and 22) resemble that of *moe* (Fig. 27), in that the expression gradually increased as diapause progressed along with the reduction of ambient temperatures. This further shows that there is a positive link between these proteins, at least on the transcriptional level, when it comes to the cold acclimation process and maintenance of physiological processes under hypometabolic conditions.

5.3.3. IDPs as proposed mediators of *O. nubilalis* cold adaptation

As a type of developmental arrest phase, diapause represents a period during which the organism is in a hypometabolic state and needing to redirect its energy reserves towards not only maintaining the basal metabolism, but also towards the synthesis of protective molecules required for developing resistance to cold temperatures and other stressful factors (dehydration, oxidative stress etc.). However, the thermodynamic conditions in which these processes are taking place are kinetically unfavorable, as kinetic energy drops when temperatures decrease. Therefore it is of utmost importance to shield the organism on a cellular level from the negative effects of the decrease in kinetic energy of the system in cold temperature conditions. These negative effects could manifest not only as aberrant structural changes, but also as arrests of biochemical and physiological processes. For this reason, molecules that retain their flexibility in such adverse conditions could represent one of the ways organisms adapt to cold temperatures on a molecular level. Intrinsically disordered proteins and proteins with long IDRs are just such biomolecules, specifically proteins, as they possess many important structural, regulatory and developmental functions in the cell and organism on a whole. The results obtained in this dissertation have shown that, although the content of intrinsically disordered proteins does not increase during the time course of diapause and gradual development of cold hardiness, it is rather stable in fact, the ratio of proteins containing long IDRs increases in these conditions. The increase in the ratio of IDR-containing proteins is also evidenced by the cold- and diapause program-induced upregulation of most of the genes encoding proteins that have been shown to possess structural disorder to varying degrees. In this manner the organism compensates its limited resources for general synthesis of *de novo* proteins, conserves energy, yet still acquires a part of the proteome that retains its high flexibility in adverse conditions, ensuring the continuation of protein structural and functional homeostasis and all processes they are involved in (Fig. 28)

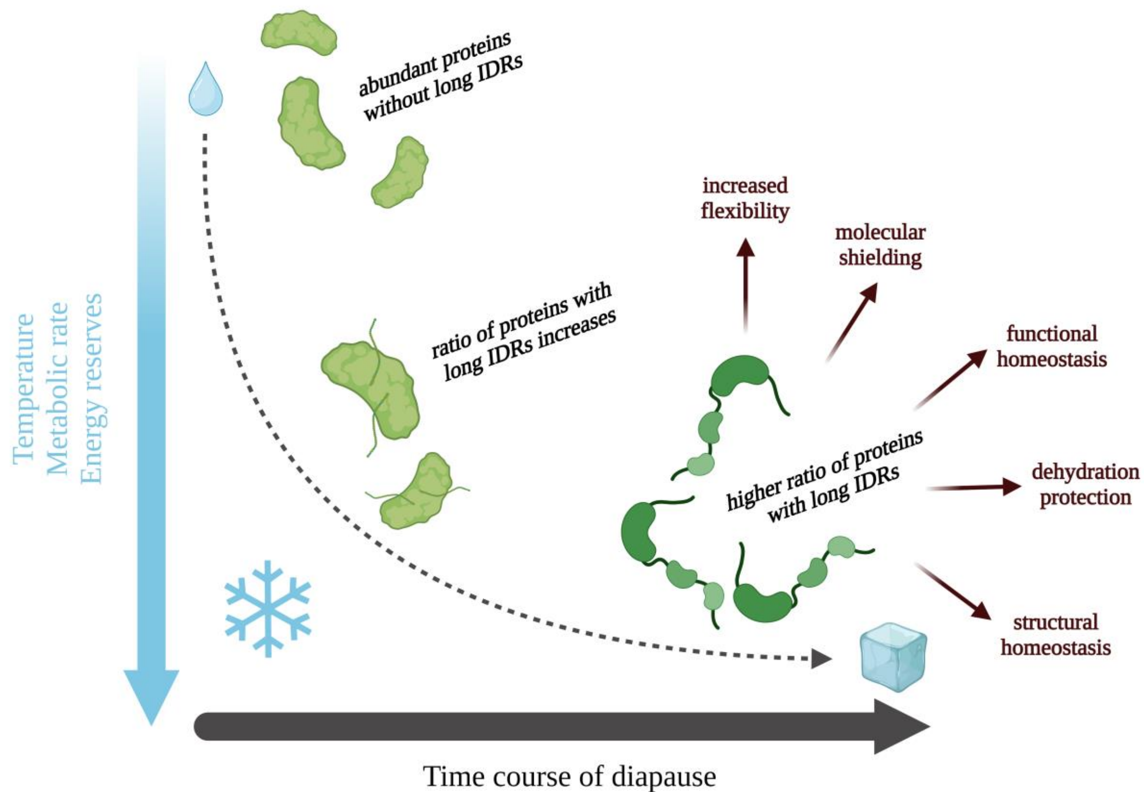

Figure 28. Proposed mechanism of proteomic changes induced by the decrease in ambient temperature and progression of a hypometabolic resting state over time. These changes would lead to an increase of proteins with more flexibility in their structure, allowing them to remain mobile and active in thermodynamically unfavorable conditions of life at low environmental temperatures. In turn, they ensure that the functional and structural homeostasis of biological processes and biomolecules is maintained during this hypometabolic and energy-restricted period. Image created with Biorender.com.

6. Conclusions

The following conclusions can be made according to the proposed hypothesis, defined aims and results of analyses performed on *O. nubilalis* 5th instar larvae that were subjected to different experimental conditions in order to trigger the development of cord hardiness:

- the methods for isolating IDPs and proteins with long IDRs were successfully implemented and optimized in *O. nubilalis* larvae, while heat treating the crude homogenates enriched the content of IDPs and proteins with long IDRs which was also validated by 2D PAGE and LC-MS/MS;
- numerous proteins, that were only predicted from nucleotide sequences available in databases, were detected by LC-MS/MS;
- the presence of IDPs and proteins with long IDRs in all experimental groups was detected by bioinformatical analysis, as were the strong effects of IDP enrichment by sample heating;
- there is a considerable fraction of heat-stable ordered proteins, as a large number of structured proteins was detected after sample heating;
- proteins with varying degrees of structural disorder were identified and confirmed by bioinformatical analyses;
- the content of specific disorder-promoting amino acids is more pronounced in proteins with higher degree of determined intrinsic disorder, which is in accordance with the results of previous studies on IDPs;
- the ratio of IDPs does not increase during the gradual cold acclimation process, which partially disproves the proposed hypothesis;
- the ratio of proteins with long IDRs increases during the gradual cold acclimation process, which partially confirms the proposed hypothesis;
- bioinformatical analysis has shown that the identified IDPs and proteins with long IDRs are involved in numerous biological processes, exhibited by the variety of their molecular functions, with most of the belonging to groups of cytoskeletal proteins, molecular chaperones, proteins involved in translation and protein metabolism, as well as regulatory proteins with nucleic acid-binding activity;

- a fraction of proteins remained unidentified and many proteins could not be assigned any functions due to information gaps in available databases;
- the expression of most genes encoding IDPs and proteins with long IDRs was upregulated during diapause and gradual development of cold hardiness;
- IDPs and proteins with long IDRs are involved in the maintenance of structural and functional homeostasis during a hypometabolic resting phase (diapause) and they contribute to the total potential of cold hardiness.

To my knowledge this is the first study on IDPs and proteins with long IDRs in an *in vivo* model that attains cold hardiness during its resting phase. These findings enrich the knowledge of the ecophysiological connection between diapause and cold hardiness, as well as the molecular mechanisms that allow not only the development but also the maintenance of cold hardiness during the long time course of the resting phase. Additionally, these results have proven, in no uncertain terms, the existence of heat-resistant ordered proteins that should be further investigated. The results of this study have potential applications in various fields – from cryopreservation in human and veterinary medicine to food and pharmaceutical industries.

7. Literature

A

Alvarez-Ponce, D., Ruiz-González, M. X., Vera-Sirera, F., Feyertag, F., Perez-Amador, M. A., & Fares, M. A. (2018). Arabidopsis heat stress-induced proteins are enriched in electrostatically charged amino acids and intrinsically disordered regions. *International Journal of Molecular Sciences*, *19*(8), 2276.

Andreadis, S. S., Vryzas, Z., Papadopoulou-Mourkidou, E., & Savopoulou-Soultani, M. (2008). Age-dependent changes in tolerance to cold and accumulation of cryoprotectants in overwintering and non-overwintering larvae of European corn borer *Ostrinia nubilalis*. *Physiological Entomology*, *33*(4), 365-371.

Arakawa, A., Handa, N., Shirouzu, M., & Yokoyama, S. (2011). Biochemical and structural studies on the high affinity of Hsp70 for ADP. *Protein Science*, *20*(8), 1367-1379.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29.

Avramov, M., Schád, É., Révész, Á., Turiák, L., Uzelac, I., Tantos, Á., Drahos, L., & Popović, Ž. D. (2022). Identification of Intrinsically Disordered Proteins and Regions in a Non-Model Insect Species *Ostrinia nubilalis* (Hbn.). *Biomolecules*, *12*(4), 592.

B

Babu, M. M. (2016). The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions*, *44*(5), 1185-1200.

Bača, F., Gošić-Dondo, S., Kaitović, Ž., & Hadzistević, D. (2007). European corn borer (*Ostrinia nubilalis* Hbn) population fluctuation at Zemun Polje between 1986 and 2005. *Maydica*, *52*(3), 325-328.

Bajusz, C., Kristó, I., Abonyi, C., Venit, T., Vedelek, V., Lukácsovich, T., Farkas, A., Borkúti, P., Kovács, Z., Bajusz, I., Marton, A., Vizler, C., Lipinszki, Z., Sinka, R., Percipalle, P. & Vilmos, P. (2021). The nuclear activity of the actin-binding Moesin protein is necessary for gene expression in *Drosophila*. *The FEBS Journal*, *288*(16), 4812-4832.

Bari, K. J., & Prakashchand, D. D. (2021). Fundamental Challenges and Outlook in Simulating Liquid–Liquid Phase Separation of Intrinsically Disordered Proteins. *The Journal of Physical Chemistry Letters*, *12*(6), 1644-1656.

Beck, S. D. (1962). Photoperiodic induction of diapause in an insect. *The Biological Bulletin*, *122*(1), 1-12.

Bethenod, M. T., Thomas, Y., Rousset, F., Frérot, B., Pélozuelo, L., Genestier, G., & Bourguet, D. (2005). Genetic isolation between two sympatric host plant races of the European corn borer, *Ostrinia nubilalis* Hübner. II: assortative mating and host-plant preferences for oviposition. *Heredity*, *94*(2), 264-270.

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I,. Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C. E. Wu, C. H., Bateman, A. & Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic acids research*, *49*(D1), D344-D354.

Bock-Marquette, I., Shrivastava, S., Pipes, G. C. T., Thatcher, J. E., Blystone, A., Shelton, J. M., Galindo, C. L., Melegh, B., Srivastava, D., Olson, E. N., & DiMaio, J. M. (2009). Thymosin beta4 mediated PKC activation is essential to initiate the embryonic coronary developmental program and epicardial progenitor cell activation in adult mice in vivo. *Journal of Molecular and Cellular Cardiology*, *46*(5), 728–738.

Bondos, S. E., Dunker, A. K., & Uversky, V. N. (2021). On the roles of intrinsically disordered proteins and regions in cell communication and signaling. *Cell Communication and Signaling*, *19*(1), 88.

Boyd, M. L., & Bailey, W. C. (2001). European corn borer: a multiple-crop pest in Missouri (2001).

Brodte, E., Knust, R., & Pörtner, H. O. (2006). Temperature-dependent energy allocation to growth in Antarctic and boreal eelpout (Zoarcidae). *Polar Biology*, *30*, 95-107.

Brown, H. G., & Hoh, J. H. (1997). Entropic exclusion by neurofilament sidearms: a mechanism for maintaining interfilament spacing. *Biochemistry*, *36*(49), 15035-15040.

Brown, J. H., Zhou, Z., Reshetnikova, L., Robinson, H., Yammani, R. D., Tobacman, L. S., & Cohen, C. (2005). Structure of the mid-region of tropomyosin: bending and binding sites for actin. *Proceedings of the National Academy of Sciences*, *102*(52), 18878-18883.

Bryon, A., Kurlovs, A. H., Van Leeuwen, T., & Clark, R. M. (2017). A molecular-genetic understanding of diapause in spider mites: current knowledge and future directions. *Physiological Entomology*, *42*(3), 211-224.

Burra, P. V., Kalmar, L., & Tompa, P. (2010). Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PloS one*, *5*(8), e12069.

C

Caffrey, D. J., & Worthley, L. H. (1927). *A progress report on the investigations of the European corn borer* (No. 1476). US Department of Agriculture.

Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N., & Dunker, A. K. (2008). TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein and peptide letters*, *15*(9), 956-963.

Cao, T., Sujkowski, A., Cobb, T., Wessells, R. J., & Jin, J. P. (2020). The glutamic acid-rich–long C-terminal extension of troponin T has a critical role in insect muscle functions. *Journal of Biological Chemistry*, *295*(12), 3794-3807.

Cao, T., Thongam, U., & Jin, J. P. (2019). Invertebrate troponin: insights into the evolution and regulation of striated muscle contraction. *Archives of biochemistry and biophysics*, *666*, 40-45.

Capinera, J. L. (2000). European corn borer, *Ostrinia nubilalis* (Hubner) (Insecta: Lepidoptera: pyralidae). University of Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, EDIS.

Chandrayan, S. K., & Guptasarma, P. (2008). Partial destabilization of native structure by a combination of heat and denaturant facilitates cold denaturation in a hyperthermophile protein. *Proteins: Structure, Function, and Bioinformatics*, *72*(2), 539-546.

Chen, B., Kayukawa, T., Monteiro, A., & Ishikawa, Y. (2005). The expression of the HSP90 gene in response to winter and summer diapauses and thermal-stress in the onion maggot, *Delia antiqua*. *Insect Molecular Biology*, *14*(6), 697-702.

Chian, R. C., & Quinn, P. (Eds.). (2010). *Fertility cryopreservation*. Cambridge University Press.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V. & Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, *325*(5942), 834-840.

Clark, M. S., Denekamp, N. Y., Thorne, M. A., Reinhardt, R., Drungowski, M., Albrecht, M. W., Klages, S. Beck, A., Kube, M. & Lubzens, E. (2012). Long-term survival of hydrated resting eggs from *Brachionus plicatilis*. *PLoS One*, *7*(1), e29365.

Clark, M. S., & Worland, M. R. (2008). How insects survive the cold: molecular mechanisms — a review. *Journal of Comparative Physiology B*, *178*, 917-933.

Clarke, M. W., Boddington, K. F., Warnica, J. M., Atkinson, J., McKenna, S., Madge, J., Barker, C. H. & Graether, S. P. (2015). Structural and functional insights into the cryoprotection of membranes by the intrinsically disordered dehydrins. *Journal of Biological Chemistry*, *290*(45), 26900-26913.

Consortium, T.U. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489.

Cooper, D. G., & Fassler, J. S. (2019). Med15: glutamine-rich mediator subunit with potential for plasticity. *Trends in biochemical sciences*, *44*(9), 737-751.

Cortese, M. S., Baird, J. P., Uversky, V. N., & Dunker, A. K. (2005). Uncovering the unfoldome: enriching cell extracts for unstructured proteins by acid treatment. *Journal of Proteome Research*, *4*(5), 1610-1618.

Cortese, M. S., Uversky, V. N., & Dunker, A. K. (2008). Intrinsic disorder in scaffold proteins: getting more from less. *Progress in biophysics and molecular biology*, *98*(1), 85-106.

Csizmók, V., Szőllősi, E., Friedrich, P., & Tompa, P. (2006). A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins. *Molecular & Cellular Proteomics*, *5*(2), 265-273.

D

D'Amico, S., Collins, T., Marx, J. C., Feller, G., Gerday, C., & Gerday, C. (2006). Psychrophilic microorganisms: challenges for life. *EMBO reports*, *7*(4), 385-389.

Danks, H. V. (1987). *Insect dormancy: an ecological perspective*. Biological Survey of Canada (Terrestrial Arthropods).

Darling, A. L., & Uversky, V. N. (2017). Intrinsic disorder in proteins with pathogenic repeat expansions. *Molecules*, *22*(12), 2027.

Darling, A. L., & Uversky, V. N. (2018). Intrinsic disorder and posttranslational modifications: the darker side of the biological dark matter. *Frontiers in genetics*, *9*, 158.

Davey, N. E. (2019). The functional importance of structure in unstructured protein regions. *Current opinion in structural biology*, *56*, 155-163.

Deiana, A., Forcelloni, S., Porrello, A., & Giansanti, A. (2019). Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One*, *14*(8), e0217889.

Deng, X., Eickholt, J., & Cheng, J. (2012). A comprehensive overview of computational protein disorder prediction methods. *Molecular BioSystems*, *8*(1), 114-121.

Denlinger, D. L. (2002). Regulation of diapause. *Annual review of entomology*, *47*(1), 93-122.

Denlinger, D. L. (2009) Encyclopedia of Insects. In: Encyclopedia of Insects (Vincent H. Resh and Ring T. Cardé, editors). Elsevier, pp. 267–271.

Domingue, M. J., Musto, C. J., Linn Jr, C. E., Roelofs, W. L., & Baker, T. C. (2008). Olfactory neuron responsiveness and pheromone blend preference in hybrids between *Ostrinia furnacalis* and *Ostrinia nubilalis* (Lepidoptera: Crambidae). *Journal of insect physiology*, *54*(8), 1261-1270.

Dosztányi, Z. (2018). Prediction of protein disorder based on IUPred. *Protein Science*, *27*(1), 331-340.

Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I., & Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. *Journal of proteome research*, *5*(11), 2985-2995.

Dosztányi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, *21*(16), 3433-3434.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., & Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of molecular graphics and modelling*, *19*(1), 26-59.

Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology*, *6*(3), 197-208.

Dyson, H. J., & Wright, P. E. (2021). NMR illuminates intrinsic disorder. *Current opinion in structural biology*, *70*, 44-52.

E

Edney, E.B. (1977) Water balance in land arthropods. Springer-Verlag, New York.

Eickholt, J., & Cheng, J. (2013). DNdisorder: predicting protein disorder using boosting and deep networks. *BMC bioinformatics*, *14*(1), 1-10.

Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. *Current opinion in structural biology*, *19*(1), 23-30.

F

Fang, S. M., Zhang, Q., Zhang, Y. L., Zhang, G. Z., Zhang, Z., & Yu, Q. Y. (2021). Heat shock protein 70 family in response to multiple abiotic stresses in the silkworm. *Insects*, *12*(10), 928.

Feng, X. W., Huo, L. J., Yang, M. C., Wang, J. X., & Shi, X. Z. (2019). Thymosins participate in antibacterial immunity of kuruma shrimp, *Marsupenaeus japonicus*. *Fish & shellfish immunology*, *84*, 244-251.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme.*Berichte der deutschen chemischen Gesellschaft*,*27*(3), 2985-2993.

Fuxreiter, M., Tompa, P., & Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, *23*(8), 950-956.

G

Galea, C. A., High, A. A., Obenauer, J. C., Mishra, A., Park, C.-G., Punta, M., Schlessinger, A., Ma, J., Rost, B., Slaughter, C. A., & Kriwacki, R. W. (2009). Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *Journal of proteome research*, *8*(1), 211-226.

Galea, C. A., Pagala, V. R., Obenauer, J. C., Park, C. G., Slaughter, C. A., & Kriwacki, R. W. (2006). Proteomic studies of the intrinsically unstructured mammalian proteome. *Journal of proteome research*, *5*(10), 2839-2848.

Gall, T. L., Romero, P. R., Cortese, M. S., Uversky, V. N., & Dunker, A. K. (2007). Intrinsic disorder in the protein data bank. *Journal of Biomolecular structure and dynamics*, *24*(4), 325-341.

Ganger, M. T., Dietz, G. D., & Ewing, S. J. (2017). A common base method for analysis of qPCR data and the application of simple blocking in qPCR experiments. *BMC Bioinformatics*, *18*(1).

Gaskin, F., Cantor, C. R., & Shelanski, M. L. (1974). Turbidimetric studies of the in vitro assembly and disassembly of porcine neurotubules. *Journal of molecular biology*, *89*(4), 737-755.

Gelman, R. A., Poppke, D. C., & Piez, K. A. (1979). Collagen fibril formation in vitro. The role of the nonhelical terminal regions. *Journal of Biological Chemistry*, *254*(22), 11741-11745.

Gkouvitsas, T., Kontogiannatos, D., & Kourti, A. (2008). Differential expression of two small Hsps during diapause in the corn stalk borer *Sesamia nonagrioides* (Lef.). *Journal of insect physiology*, *54*(12), 1503-1510.

Gkouvitsas, T., Kontogiannatos, D., & Kourti, A. (2009). Cognate Hsp70 gene is induced during deep larval diapause in the moth *Sesamia nonagrioides*. *Insect molecular biology*, *18*(2), 253-264.

Goldstein, A. L., Slater, F. D., & White, A. (1966). Preparation, assay, and partial purification of a thymic lymphocytopoietic factor (thymosin). *Proceedings of the National Academy of Sciences*, *56*(3), 1010-1017.

Govaere, L., Morin, M. D., Frigault, J. J., Boquel, S., Cohen, A., & Lamarre, S. G. (2019). Transcriptome and proteome analyses to investigate the molecular underpinnings of cold response in the Colorado potato beetle, *Leptinotarsa decemlineata*. *Cryobiology*, *88*, 54-63.

Grubor-Lajsic, G., Block, W., Palanacki, V., & Glumac, S. (1991). Cold hardiness parameters of overwintering diapause larvae of *Ostrinia nubilalis* in Vojvodina, Yugoslavia. *CryoLetters*, *12*(3), 177-182.

Gunning, P., O'neill, G., & Hardeman, E. (2008). Tropomyosin-based regulation of the actin cytoskeleton in time and space. *Physiological reviews*, *88*(1), 1-35.

Guo, Y., Guettouche, T., Fenna, M., Boellmann, F., Pratt, W. B., Toft, D. O., Smith, D. F., & Voellmy, R. (2001). Evidence for a mechanism of repression of heat shock factor 1 transcriptional activity by a multichaperone complex. *Journal of Biological Chemistry*, *276*(49), 45791-45799.

H

Habchi, J., Tompa, P., Longhi, S., & Uversky, V. N. (2014). Introducing protein intrinsic disorder. *Chemical reviews*, *114*(13), 6561-6588.

Hadley, N. F. (1994). Water relations of terrestrial arthropods. CUP Archive.

Hahn, D. A., & Denlinger, D. L. (2007). Meeting the energetic demands of insect diapause: nutrient storage and utilization. *Journal of insect physiology*, *53*(8), 760-773.

Hahn, D. A., & Denlinger, D. L. (2011). Energetics of insect diapause. *Annual review of entomology*, *56*, 103-121.

Hao, Y. J., Li, W. S., He, Z. B., Si, F. L., Ishikawa, Y., & Chen, B. (2012). Differential gene expression between summer and winter diapause pupae of the onion maggot *Delia antiqua*, detected by suppressive subtractive hybridization. *Journal of insect physiology*, *58*(11), 1444-1449.

Haslbeck, M., Franzmann, T., Weinfurtner, D., & Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nature structural & molecular biology*, *12*(10), 842-846.

Hatos, A., Hajdu-Soltész, B., Monzon, A. M., Palopoli, N., Álvarez, L., Aykac-Fas, B., Bassot, C., Benítez, G. I., Bevilacqua, M., Chasapi, A., Chemes, L., Davey, N. E., Davidović, R., Dunker, A. K., Elofsson, A., Gobeill, J., Foutel, N. S. G., Sudha, G., Guharoy, M., Horvath, T., Iglesias, V., Kajava, A. V., Kovacs, O. P., Lamb, J., Lambrughi, M., Lazar, T., Leclercq, J. Y., Leonardi, E., Macedo-Ribeiro, S., Macossay-Castillo, M., Maiani, E., Manso, A. M., Marino-Buslje, C., Martínez-Pérez, E., Mészaros, B., Mičetić, I., Minervini, G., Murvai, N., Necci, M., Ouzounis, C. A., Pajkos, M., Paladin, L., Pancsa, R., Papaleo, E., Parisi, G., Pasche, E., Barbosa Pereira, P. J., Promponas, V. J., Pujols, J., Quaglia, F., Ruch, P., Salvatore, M., Schad, E., Szabo, B., Szaniszló, T., Tamana, S., Tantos, A., Veljkovic, N., Ventura, S., Vranken, W., Dosztányi, Z., Tompa, P., Tosatto, S. C. E. & Piovesan, D. (2020). DisProt: intrinsic protein disorder annotation in 2020. *Nucleic acids research*, *48*(D1), D269-D276.

He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., & Dunker, A. K. (2009). Predicting intrinsic disorder in proteins: an overview. *Cell research*, *19*(8), 929-949.

Hibino, E., Inoue, R., Sugiyama, M., Kuwahara, J., Matsuzaki, K., & Hoshino, M. (2016). Interaction between intrinsically disordered regions in transcription factors Sp1 and TAF4. *Protein Science*, *25*(11), 2006-2017.

Hochachka, P. W., & Somero, G. N. (2002). *Biochemical adaptation: mechanism and process in physiological evolution*. Oxford university press.

Huang, L. H., & Kang, L. (2007). Cloning and interspecific altered expression of heat shock protein genes in two leafminer species in response to thermal stress. *Insect Molecular Biology*, *16*(4), 491-500.

Huang, L. H., Wang, C. Z., & Kang, L. (2009). Cloning and expression of five heat shock protein genes in relation to cold hardening and development in the leafminer, *Liriomyza sativa*. *Journal of insect physiology*, *55*(3), 279-285.

I

Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of molecular biology*, *323*(3), 573-584.

Imai, K., & Mitaku, S. (2005). Mechanisms of secondary structure breakers in soluble proteins. *Biophysics*, *1*, 55-65.

International Silkworm Genome Consortium. (2008). The genome of a lepidopteran model insect, the silkworm Bombyx mori. *Insect biochemistry and molecular biology*, *38*(12), 1036-1045.

Izumi, Y., Sonoda, S., & Tsumuki, H. (2007). Effects of diapause and cold-acclimation on the avoidance of freezing injury in fat body tissue of the rice stem borer, *Chilo suppressalis* Walker. *Journal of insect physiology*, *53*(7), 685-690.

J

Jakob, U., Kriwacki, R., & Uversky, V. N. (2014). Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chemical reviews*, *114*(13), 6779-6805.

Janke, C., Rogowski, K., & Van Dijk, J. (2008). Polyglutamylation: a fine-regulator of protein function?. *EMBO reports*, *9*(7), 636-641.

Jones, D. T., & Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*, *31*(6), 857-863.

Jorda, J., Xue, B., Uversky, V. N., & Kajava, A. V. (2010). Protein tandem repeats – the more perfect, the less structured. *The FEBS journal*, *277*(12), 2673-2682.

K

Kankare, M., Salminen, T., Laiho, A., Vesala, L., & Hoikkala, A. (2010). Changes in gene expression linked with adult reproductive diapause in a northern malt fly species: a candidate gene microarray study. *BMC ecology*, *10*(1), 1-9.

Kapur, I., Boulier, E. L., & Francis, N. J. (2022). Regulation of Polyhomeotic condensates by intrinsically disordered sequences that affect chromatin binding. *Epigenomes*, *6*(4), 40.

King, A. M., & MacRae, T. H. (2015). Insect heat shock proteins during stress and diapause. *Annual review of entomology*, *60*, 59-75.

Kleinman, H. K., & Sosne, G. (2016). Thymosin β4 promotes dermal healing. *Vitamins and hormones*, *102*, 251-275.

Kocmánková, E., Trnka, M., Žalud, Z., Semerádová, D., Dubrovský, M., Muška, F., & Možný, M. (2008). Comparison of two mapping methods of potential distribution of pests under present and changed climate. *Plant Protection Science*, *44*(2), 49.

Kojić, D. (2009). Cold hardiness and dehydration of European corn borer (*Ostrinia nubilalis*): molecular and cell mechanisms. Doctoral dissertation, University of Novi Sad, Faculty of Sciences.

Kojić, D., Popović, Ž. D., Orčić, D., Purać, J., Orčić, S., Vukašinović, E. L., Nikolić, T. V., & Blagojević, D. P. (2018). The influence of low temperature and diapause phase on sugar and polyol content in the European corn borer *Ostrinia nubilalis* (Hbn.). *Journal of insect physiology*, *109*, 107-113.

Koshland Jr, D. E. (1995). The key–lock theory and the induced fit theory. *Angewandte Chemie International Edition in English*, *33*(23-24), 2375-2378.

Koštál, V. (2006). Eco-physiological phases of insect diapause. *Journal of insect physiology*, *52*(2), 113-127.

Koštál, V., Štětina, T., Poupardin, R., Korbelová, J., & Bruce, A. W. (2017). Conceptual framework of the eco-physiological phases of insect diapause development justified by transcriptomic profiling. *Proceedings of the National Academy of Sciences*, *114*(32), 8532-8537.

Koštál, V., & Tollarova-Borovanska, M. (2009). The 70 kDa heat shock protein assists during the repair of chilling injury in the insect, *Pyrrhocoris apterus*. *PloS one*, *4*(2), e4546.

Koštál, V., Zahradníčková, H., Šimek, P., & Zelený, J. (2007). Multiple component system of sugars and polyols in the overwintering spruce bark beetle, *Ips typographus*. *Journal of Insect Physiology*, *53*(6), 580-586.

Kregel, K. C. (2002). Heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. *Journal of applied physiology*, *92*(5), 2177-2186.

Kristó, I., Bajusz, C., Borsos, B. N., Pankotai, T., Dopie, J., Jankovics, F., Vartiainen, M. K., Erdélyi, M., & Vilmos, P. (2017). The actin binding cytoskeletal protein Moesin is involved in nuclear mRNA export. *Biochimica et Biophysica Acta. Molecular Cell Research*, *1864*(10), 1589–1604.

Kuhar, T. P., Barlow, V. M., Hoffmann, M. P., Fleischer, S. J., Groden, E., Gardner, J., Hazzard, R., Wright, M. G., Pitcher, S. A., Speese, J., & Westgate, P. (2004). Potential of *Trichogramma ostriniae* (Hymenoptera: Trichogrammatidae) for biological control of European corn borer (Lepidoptera: Crambidae) in solanaceous crops. *Journal of Economic Entomology*, *97*(4), 1209-1216.

Kulkarni, P., & Uversky, V. N. (2018). Intrinsically disordered proteins: the dark horse of the dark proteome. *Proteomics*, *18*(21-22), 1800061.

L

Lazar, T., Schad, E., Szabo, B., Horvath, T., Meszaros, A., Tompa, P., & Tantos, A. (2016). Intrinsic protein disorder in histone lysine methylation. *Biology Direct*, *11*(1), 1-10.

Léger, T., Mally, R., Neinhuis, C., & Nuss, M. (2021). Refining the phylogeny of Crambidae with complete sampling of subfamilies (Lepidoptera, Pyraloidea). *Zoologica Scripta*, *50*(1), 84-99.

Li, A., & Denlinger, D. L. (2008). Rapid cold hardening elicits changes in brain protein profiles of the flesh fly, *Sarcophaga crassipalpis*. *Insect molecular biology*, *17*(5), 565-572.

Li, B., & Yang, Z. (2022). Multilocus evidence provides insight into the demographic history and asymmetrical gene flow between *Ostrinia furnacalis* and *Ostrinia nubilalis* (Lepidoptera: Crambidae) in the Yili area, Xinjiang, China. *Ecology and Evolution*, *12*(11), e9504.

Li, Z. W., Li, X., Yu, Q. Y., Xiang, Z. H., Kishino, H., & Zhang, Z. (2009). The small heat shock protein (sHSP) genes in the silkworm, *Bombyx mori*, and comparative analysis with other insect sHSP genes. *BMC Evolutionary Biology*, *9*(1), 1-14.

Liao, J., Zhang, C., Ru, W., Wang, D., & Zhang, W. (2018). Effects of overexpression and inhibited expression of thymosin, an actin-interacting protein from *Bombyx mori*, on BmNPV proliferation and replication. *Archives of insect biochemistry and physiology*, *98*(1), e21449.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods*, *25*(4), 402–408.

Lobley, A., Swindells, M. B., Orengo, C. A., & Jones, D. T. (2007). Inferring function using patterns of native disorder in proteins. *PLoS computational biology*, *3*(8), e162.

M

Ma, S., Kang, Z., Lü, P., Yang, Y., Yao, Q., Xia, H., & Chen, K. (2015). Molecular and physiological characterization of two novel multirepeat β-thymosins from silkworm, *Bombyx mori*. *Plos one*, *10*(10), e0140182.

MacRae, T. H. (2010). Gene expression, metabolic regulation and stress tolerance during diapause. *Cellular and Molecular Life Sciences*, *67*, 2405-2424.

Marden, J. H., Fescemyer, H. W., Saastamoinen, M., MacFarland, S. P., Vera, J. C., Frilander, M. J., & Hanski, I. (2008). Weight and nutrition affect pre-mRNA splicing of a muscle gene associated with performance, energetics and life history. *Journal of Experimental Biology*, *211*(23), 3653-3660.

Marinotti, O., Calvo, E., Nguyen, Q. K., Dissanayake, S., Ribeiro, J. M. C., & James, A. A. (2006). Genome-wide analysis of gene expression in adult *Anopheles gambiae*. *Insect molecular biology*, *15*(1), 1-12.

Marks, E. D., & Kumar, A. (2016). Thymosin β4: roles in development, repair, and engineering of the cardiovascular system. *Vitamins and Hormones*, *102*, 227-249.

Marqués, M. I., Borreguero, J. M., Stanley, H. E., & Dokholyan, N. V. (2003). Possible mechanism for cold denaturation of proteins at high pressure. *Physical review letters*, *91*(13), 138103.

Mayer, M. P., & Gierasch, L. M. (2019). Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *Journal of Biological Chemistry*, *294*(6), 2085-2097.

Mayer, M. P., Schröder, H., Rüdiger, S., Paal, K., Laufen, T., & Bukau, B. (2000). Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nature structural biology*, *7*(7), 586-593.

Mazur, P. (1970). Cryobiology: The Freezing of Biological Systems: The responses of living cells to ice formation are of theoretical interest and practical concern. *Science*, *168*(3934), 939-949.

Mazur, P. (1984). Freezing of living cells: mechanisms and implications. *American journal of physiology-cell physiology*, *247*(3), C125-C142.

McMahon, H. T., & Gallop, J. L. (2005). Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature*, *438*(7068), 590-596.

Minezaki, Y., Homma, K., Kinjo, A. R., & Nishikawa, K. (2006). Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *Journal of molecular biology*, *359*(4), 1137-1149.

Mirsky, A. E., & Pauling, L. (1936). On the structure of native, denatured, and coagulated proteins. *Proceedings of the National Academy of sciences*, *22*(7), 439-447.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419.

Morgan, A. A., & Rubenstein, E. (2013). Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PloS one*, *8*(1), e53785.

Mutuura, A., & Munroe, E. (1970). Taxonomy and distribution of the European corn borer and allied species: genus *Ostrinia* (Lepidoptera: Pyralidae). *The Memoirs of the Entomological Society of Canada*, *102*(S71), 1-112.

N

Narberhaus, F. (2002). α-Crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiology and Molecular Biology Reviews*, *66*(1), 64-93.

Nation, J. (2008) Insect physiology and biochemistry. 2nd Edition. CRC Press.

Necci, M., Piovesan, D., & Tosatto, S. C. (2016). Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Protein Science*, *25*(12), 2164-2174.

Necci, M., Piovesan, D., & Tosatto, S. C. (2021). Critical assessment of protein intrinsic disorder prediction. *Nature methods*, *18*(5), 472-481.

Nordin, J. H., Cui, Z., & Yin, C. M. (1984). Cold-induced glycerol accumulation by *Ostrinia nubilalis larvae* is developmentally regulated. *Journal of Insect Physiology*, *30*(7), 563-566.

Nylin, S. (2013). Induction of diapause and seasonal morphs in butterflies and other insects: knowns, unknowns and the challenge of integration. *Physiological Entomology*, *38*(2), 96-104.

O

Okada, A. K., Teranishi, K., Ambroso, M. R., Isas, J. M., Vazquez-Sarandeses, E., Lee, J.-Y., Melo, A. A., Pandey, P., Merken, D., Berndt, L., Lammers, M., Daumke, O., Chang, K., Haworth, I. S., & Langen, R. (2021). Lysine acetylation regulates the interaction between proteins and membranes. *Nature Communications*, *12*(1), 6466.

Oldfield, C. J., & Dunker, A. K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual review of biochemistry*, *83*, 553-584.

P

Pancsa, R., & Tompa, P. (2012). Structural disorder in eukaryotes. *PloS one*, *7*(4), e34687.

Park, S. M., Jung, H. Y., Kim, T. D., Park, J. H., Yang, C. H., & Kim, J. (2002). Distinct roles of the N-terminal-binding domain and the C-terminal-solubilizing domain of α-synuclein, a molecular chaperone. *Journal of Biological Chemistry*, *277*(32), 28512-28520.

Pavadai, E., Rynkiewicz, M. J., Ghosh, A., & Lehman, W. (2020). Docking troponin T onto the tropomyosin overlapping domain of thin filaments. *Biophysical journal*, *118*(2), 325-336.

Pavlides, S. C., Pavlides, S. A., & Tammariello, S. P. (2011). Proteomic and phosphoproteomic profiling during diapause entrance in the flesh fly, Sarcophaga crassipalpis. *Journal of insect physiology*, *57*(5), 635-644.

Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., Hu, G., Uversky, V. N., & Kurgan, L. (2015). Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cellular and Molecular Life Sciences*, *72*(1), 137–151.

Pentony, M. M., & Jones, D. T. (2010). Modularity of intrinsic disorder in the human proteome. *Proteins: Structure, Function, and Bioinformatics*, *78*(1), 212-221.

Petersen, J., Eriksson, S. K., Harryson, P., Pierog, S., Colby, T., Bartels, D., & Röhrig, H. (2012). The lysine-rich motif of intrinsically disordered stress protein CDeT11-24 from Craterostigma plantagineum is responsible for phosphatidic acid binding and protection of enzymes from damaging effects caused by desiccation. *Journal of experimental botany*, *63*(13), 4919-4929.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic acids research*, *29*(9), e45-e45.

Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C. J., Aspromonte, M. C., Davey, N. E., Davidović, R., Dosztányi, Z., Elofsson, A., Gasparini, A., Hatos, A., Kajava, A. V., Kalmar, L., Leonardi, E., Lazar, T., Macedo-Ribeiro, S., Macossay-Castillo, M., Meszaros, A., Minervini, G., Murvai, N., Pujols, J., Roche, D. B., Salladini, E., Schad, E., Schramm, A., Szabo, B., Tantos, A., Tonello, F., Tsirigos, K. D., Veljković, N., Ventura, S., Vranken, W., Warholm, P., Uversky, V. N., Dunker, A. K., Longhi, S., Tompa, P. & Tosatto, S. C. E. (2017). DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Research*, *45*(D1), D219–D227.

Popović, Ž. D. (2014). Molecular and biochemical basis of diapause of European corn borer *Ostrinia nubilalis* (Hbn.) (Lepidoptera: Pyralidae). Doctoral dissertation, University of Belgrade, Faculty of Biology.

Popović, Ž. D., Maier, V., Avramov, M., Uzelac, I., Gošić-Dondo, S., Blagojević, D., & Koštál, V. (2021). Acclimations to cold and warm conditions differently affect the energy metabolism of diapausing larvae of the european corn borer *Ostrinia nubilalis* (Hbn.). *Frontiers in physiology*, 2126.

Popović, Ž. D., Subotić, A., Nikolić, T. V., Radojičić, R., Blagojević, D. P., Grubor-Lajšić, G., & Koštál, V. (2015). Expression of stress-related genes in diapause of European corn borer (*Ostrinia nubilalis* Hbn.). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *186*, 1-7.

Portman, S. L., Felton, G. W., Kariyat, R. R., & Marden, J. H. (2020). Host plant defense produces species-specific alterations to flight muscle protein structure and flight-related fitness traits of two armyworms. *Journal of Experimental Biology*, *223*(16), jeb224907.

Portman, S. L., Kariyat, R. R., Johnston, M. A., Stephenson, A. G., & Marden, J. H. (2014). Cascading effects of host plant inbreeding on the larval growth, muscle molecular composition, and flight capacity of an adult herbivorous insect. *Functional Ecology*, *29*(3), 328-337.

Pörtner, H. O., & Knust, R. (2007). Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science*, *315*(5808), 95-97.

Privalov, P. L. (1990). Cold denaturation of protein. *Critical reviews in biochemistry and molecular biology*, *25*(4), 281-306.

Purać, J., Kojić, D., Popović, Ž. D., Vukašinović, E., Tiziani, S., Günther, U. L., & Grubor-Lajšić, G. (2015). Metabolomic Analysis of Diapausing and Noni-diapausing Larvae of the European Corn Borer *Ostrinia nubilalis* (Hbn.) (Lepidoptera: Crambidae). *Acta Chimica Slovenica*, *62*(4), 761–767.

Purać, J., Kojić, D., Petri, E., Popović, Ž. D., Grubor-Lajšić, G., & Blagojević, D. P. (2016). Cold adaptation responses in insects and other arthropods: an "omics" approach. *Short Views on Insect Genomics and Proteomics: Insect Proteomics, Vol. 2*, 89-112.

R

Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., & Dunker, A. K. (2007). Intrinsic disorder and functional proteomics. *Biophysical journal*, *92*(5), 1439-1456.

Reichmann, D., Xu, Y., Cremers, C. M., Ilbert, M., Mittelman, R., Fitzgerald, M. C., & Jakob, U. (2012). Order out of disorder: working cycle of an intrinsically unfolded chaperone. *Cell*, *148*(5), 947-957.

Rinehart, J. P., & Denlinger, D. L. (2000). Heat-shock protein 90 is down-regulated during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*, but remains responsive to thermal stress. *Insect molecular biology*, *9*(6), 641-645.

Rinehart, J. P., Li, A., Yocum, G. D., Robich, R. M., Hayward, S. A., & Denlinger, D. L. (2007). Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proceedings of the National Academy of Sciences*, *104*(27), 11130-11137.

Robich, R. M., Rinehart, J. P., Kitchen, L. J., & Denlinger, D. L. (2007). Diapause-specific gene expression in the northern house mosquito, *Culex pipiens* L., identified by suppressive subtractive hybridization. *Journal of Insect Physiology*, *53*(3), 235-245.

Röhl, A., Rohrberg, J., & Buchner, J. (2013). The chaperone Hsp90: changing partners for demanding clients. *Trends in biochemical sciences*, *38*(5), 253-262.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Structure, Function, and Bioinformatics*, *42*(1), 38-48.

Romero-Pérez, S. P., Covarrubias, A. A., & Campos, F. (2023). A simple method to purify intrinsically disordered proteins by adjusting trichloroacetic acid concentration. *Protein Expression and Purification*, *202*, 106183.

S

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature reviews Molecular cell biology*, *14*(10), 630-642.

Santner, A. A., Croy, C. H., Vasanwala, F. H., Uversky, V. N., Van, Y. Y. J., & Dunker, A. K. (2012). Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry*, *51*(37), 7250-7262.

Sarkar, T., Manna, T., Bhattacharyya, S., Mahapatra, P., Poddar, A., Roy, S., Pena, J., Solana, R., Tarazona, R., & Bhattacharyya, B. (2001). Role of the carboxy-termini of tubulin on its chaperone-like activity. *Proteins: Structure, Function, and Bioinformatics*, *44*(3), 262-269.

Schillaci, D., Vitale, M., Cusimano, M. G., & Arizza, V. (2012). Fragments of β-thymosin from the sea urchin Paracentrotus lividus as potential antimicrobial peptides against staphylococcal biofilms. *Annals of the New York Academy of Sciences*, *1270*(1), 79-85.

Sekulić, R., Spasić, R., & Kereši, T. (2008). Štetočine povrća i njihovo suzbijanje. *Poljoprivredni fakultet, Novi Sad i Beograd, Institut za ratarstvo i povrtarstvo, Novi Sad*.

Simon, M., & Hancock, J. M. (2009). Tandem and cryptic amino acid repeats accumulate in disordered regions of proteins. *Genome Biology*, *10*(6), 1-16.

Semenyuk, P. I. (2021). Effect of Polyphosphorylation on Behavior of Protein Disordered Regions. *International Journal of Molecular Sciences*, *22*(15), 7883.

Senanayake, D. G., Radcliffe, E. B., & Holliday, N. J. (2000). Oviposition and diapause behavior in Colorado potato beetle (Coleoptera: Chrysomelidae) populations from east central Minnesota and the valley of the Red River of the North. *Environmental Entomology*, *29*(6), 1123-1132.

Shabardina, V., Kashima, Y., Suzuki, Y., & Makalowski, W. (2020). Emergence and evolution of ERM proteins and merlin in metazoans. *Genome Biology and Evolution*, *12*(1), 3710-3724.

Siddiqui, K. S., Poljak, A., Guilhaus, M., De Francisci, D., Curmi, P. M. G., Feller, G., D'Amico, S., Gerday, C., Uversky, V. N., & Cavicchioli, R. (2006). Role of lysine versus arginine in enzyme cold-adaptation: Modifying lysine to homo-arginine stabilizes the cold-adapted α-amylase from *Pseudoalteramonas haloplanktis*. *Proteins: Structure, Function, and Bioinformatics*, *64*(2), 486–501.

Siemer, A. B. (2020). Advances in studying protein disorder with solid-state NMR. *Solid state nuclear magnetic resonance*, *106*, 101643.

Smith, H. E. (1920). Broom corn, the probable host in which *Pyrausta nubilalis* Hubn. reached America. *Journal of Economic Entomology*, *13*(5), 425-430.

Smith, M. D., & Jelokhani-Niaraki, M. (2012). pH-induced changes in intrinsically disordered proteins. *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools*, 223-231.

Sole, J., Sans, A., Riba, M., & Guerrero, A. (2010). Behavioural and electrophysiological responses of the European corn borer *Ostrinia nubilalis* to host-plant volatiles and related chemicals. *Physiological Entomology*, *35*(4), 354-363.

Solis, M. (2007). Phylogenetic studies and modern classification of the Pyraloidea (Lepidoptera). *Revista Colombiana de Entomología*, *33*(1), 1-8.

Sonoda, S., Fukumoto, K., Izumi, Y., Yoshida, H., & Tsumuki, H. (2006). Cloning of heat shock protein genes (hsp90 and hsc70) and their expression during larval diapause and cold tolerance acquisition in the rice stem borer, *Chilo suppressalis* Walker. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, *63*(1), 36-47.

Stetler, R. A., Gan, Y., Zhang, W., Liou, A. K., Gao, Y., Cao, G., & Chen, J. (2010). Heat shock proteins: cellular and molecular mechanisms in the central nervous system. *Progress in neurobiology*, *92*(2), 184-211.

Storey, J. M., & Storey, K. B. (2022). Chaperone proteins: universal roles in surviving environmental stress. *Cell Stress and Chaperones*, 1-12.

Storey, K. B., & Storey, J. M. (2004). Metabolic rate depression in animals: transcriptional and translational controls. *Biological Reviews*, *79*(1), 207-233.

Storey, K. B., & Storey, J. M. (2012). Insect cold hardiness: metabolic, gene, and protein adaptation. *Canadian Journal of Zoology*, *90*(4), 456-475.

T

Tanford, C. (1968). Protein denaturation. *Advances in protein chemistry*, *23*, 121-282.

Tantos, A., Friedrich, P., & Tompa, P. (2009). Cold stability of intrinsically disordered proteins. *FEBS letters*, *583*(2), 465-469.

Tantos, A., & Tompa, P. (2012). Identification of intrinsically disordered proteins by a special 2D electrophoresis. *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools*, 215-222.

Taški, K. J., Stanić, B. Đ., & Grubor-Lajšić, G. N. (2004). The presence of an arylphorin-type storage protein at different life stages of *Ostrinia nubilalis*-Lepidoptera: Pyralidae. *Zbornik Matice srpske za prirodne nauke*, (106), 5-13.

Teets, N. M., Kawarasaki, Y., Lee, R. E., & Denlinger, D. L. (2013). Expression of genes involved in energy mobilization and osmoprotectant synthesis during thermal and dehydration stress in the Antarctic midge, *Belgica antarctica*. *Journal of Comparative Physiology B*, *183*, 189-201.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends in biochemical sciences*, *37*(12), 509-516.

Tungjitwitayakul, J., Tatun, N., Singtripop, T., & Sakurai, S. (2008). Characteristic expression of three heat shock-responsive genes during larval diapause in the bamboo borer *Omphisa fuscidentalis*. *Zoological Science*, *25*(3), 321-333.

U

Ukmar-Godec, T., Hutten, S., Grieshop, M. P., Rezaei-Ghaleh, N., Cima-Omori, M.-S., Biernat, J., Mandelkow, E., Söding, J., Dormann, D., & Zweckstetter, M. (2019). Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nature Communications*, *10*(1).

Urbanek, A., Popovic, M., Morató, A., Estaña, A., Elena-Real, C.A., Mier, P., Fournet, A., Allemand, F., Delbecq, S., Andrade-Navarro, M.A. and Cortés, J., Sibille N. & Bernadó, P. (2020). Flanking regions determine the structure of the poly-glutamine in huntingtin through mechanisms common among glutamine-rich human proteins. *Structure*, *28*(7), 733-746.

Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein science*, *11*(4), 739-756.

Uversky, V. N. (2009). Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The protein journal*, *28*, 305-325.

Uversky, V. N. (2011). Intrinsically disordered proteins from A to Z. The International Journal of Biochemistry & Cell Biology, 43(8), 1090–1103.

Uversky, V. N. (2013a). The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically disordered proteins*, *1*(1), e24684.

Uversky, V. N. (2013b). The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS letters*, *587*(13), 1891-1901.

Uversky, V. N. (2015). The intrinsic disorder alphabet. III. Dual personality of serine. *Intrinsically disordered proteins*, *3*(1), e1027032.

Uversky, V. N. (2019). Intrinsically disordered proteins and their "mysterious"(meta) physics. *Frontiers in Physics*, *7*, 10.

Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions?. *Proteins: structure, function, and bioinformatics*, *41*(3), 415-427.

Uversky, V. N., Li, J., & Fink, A. L. (2001). Evidence for a partially folded intermediate in α-synuclein fibril formation. *Journal of Biological Chemistry*, *276*(14), 10737-10744.

Uversky, V. N., Oldfield, C. J., & Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: introducing the D^2 concept. *Annu. Rev. Biophys.*, *37*, 215-246.

Uzelac, I., Avramov, M., Čelić, T., Vukašinović, E., Gošić-Dondo, S., Purać, J., Kojić, D., Blagojević, D., & Popović, Ž. D. (2020). Effect of Cold Acclimation on Selected Metabolic Enzymes During Diapause in The European Corn Borer *Ostrinia nubilalis* (Hbn.). *Scientific Reports*, *10*(1), 9085.

V

Vajgand, D. (2010). Flight dynamic of Lepidoptera of economic importance in Sombor during 2010 and forecast for 2011. *Acta entomologica Serbica*, *15*(2), 205-219.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., & Babu, M. M. (2014). Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews*, *114*(13), 6589–6631.

Vicedo, E., Schlessinger, A., & Rost, B. (2015). Environmental pressure may change the composition protein disorder in prokaryotes. *PloS one*, *10*(8), e0133990.

Vinterhalter, G., Kovačević, J. J., Uversky, V. N., & Pavlović-Lažetić, G. M. (2021). Bioinformatics analysis of correlation between protein function and intrinsic disorder. *International journal of biological macromolecules*, *167*, 446-456.

Vukašinović, E. L., Pond, D. W., Worland, M. R., Kojić, D., Purać, J., Blagojević, D. P., & Grubor-Lajšić, G. (2013). Diapause induces changes in the composition and biophysical properties of lipids in larvae of the European corn borer, *Ostrinia nubilalis* (Lepidoptera: Crambidae). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *165*(4), 219-225.

Vukašinović, E. L., Pond, D. W., Worland, M. R., Kojić, D., Purać, J., Popović, Ž. D., & Grubor-Lajšić, G. (2015). Diapause induces remodeling of the fatty acid composition of membrane and storage lipids in overwintering larvae of *Ostrinia nubilalis*, Hubn.(Lepidoptera: Crambidae). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *184*, 36-43.

Vukašinović, E. L., Pond, D. W., Grubor-Lajšić, G., Worland, M. R., Kojić, D., Purać, J., Popović, Ž. D. & Blagojević, D. P. (2018). Temperature adaptation of lipids in diapausing *Ostrinia nubilalis*: an experimental study to distinguish environmental versus endogenous controls. *Journal of Comparative Physiology B*, *188*, 27-36.

Vuzman, D., Azia, A., & Levy, Y. (2010). Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *Journal of molecular biology*, *396*(3), 674-684.

W

Wei, B., & Jin, J. P. (2016). TNNT1, TNNT2, and TNNT3: Isoform genes, regulation, and structure–function relationships. *Gene*, *582*(1), 1-13.

Weissgerber, T. L., Milic, N. M., Winham, S. J., & Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLOS Biology*, *13*(4), e1002128.

Williams, R. M., Obradovic, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J. & Dunker, A. K. (2000). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. In *Biocomputing 2001* (pp. 89-100).

Willett, C. S., & Harrison, R. G. (1999). Insights into genome differentiation: pheromone-binding protein variation and population history in the European corn borer (*Ostrinia nubilalis*). *Genetics*, *153*(4), 1743-1751.

Windisch, H. S., Lucassen, M., & Frickenhaus, S. (2012). Evolutionary force in confamiliar marine vertebrates of different temperature realms: adaptive trends in zoarcid fish transcriptomes. *BMC genomics*, *13*(1), 1-16.

Winter, D. L., Abeygunawardena, D., Hart-Smith, G., Erce, M. A., & Wilkins, M. R. (2015). Lysine methylation modulates the protein–protein interactions of yeast cytochrome C Cyc1p. *Proteomics*, *15*(13), 2166-2176.

Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, *293*(2), 321-331.

Wright, P. E., & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews Molecular cell biology*, *16*(1), 18-29.

Wu, H. (1995). Studies on denaturation of proteins XIII. A theory of denaturation. In *Advances in protein chemistry* (Vol. 46, pp. 6-26). Academic Press.

Wu, S., Hong, L., Wang, Y., Yu, J., Yang, J., Yang, J., Zhang, H. & Perrett, S. (2020). Kinetics of the conformational cycle of Hsp70 reveals the importance of the dynamic and heterogeneous nature of Hsp70 for its function. *Proceedings of the National Academy of Sciences*, *117*(14), 7814-7823.

X

Xue, B., Blocquel, D., Habchi, J., Uversky, A. V., Kurgan, L., Uversky, V. N., & Longhi, S. (2014). Structural disorder in viral proteins. *Chemical reviews*, *114*(13), 6880-6911.

Xue, B., Dunker, A. K., & Uversky, V. N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure and Dynamics*, *30*(2), 137-149.

Y

Yang, C. L., Meng, J. Y., Zhou, L., Yao, M. S., & Zhang, C. Y. (2021). Identification of five small heat shock protein genes in *Spodoptera frugiperda* and expression analysis in response to different environmental stressors. *Cell Stress and Chaperones*, *26*, 527-539.

Yang, W. J., Xu, K. K., Cao, Y., Meng, Y. L., Liu, Y., & Li, C. (2019). Identification and expression analysis of four small heat shock protein genes in cigarette beetle, *Lasioderma serricorne* (Fabricius). *Insects*, *10*(5), 139.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1).

Young, V. R., & Ajami, A. M. (2000). Glutamate: an amino acid of particular distinction. *The Journal of nutrition*, *130*(4), 892S-900S.

Z

Zachariassen, K. E., & Kristiansen, E. (2000). Ice nucleation and antinucleation in nature. *Cryobiology*, *41*(4), 257-279.

Zhang, C., Wei, D., Shi, G., Huang, X., Cheng, P., Liu, G., Guo, X., Liu, L., Wang, H., Miao, F. & Gong, M. (2019). Understanding the regulation of overwintering diapause molecular mechanisms in *Culex pipiens pallens* through comparative proteomics. *Scientific Reports*, *9*(1), 6485.

Zhang, F. X., Shao, H. L., Wang, J. X., & Zhao, X. F. (2011). β-Thymosin is upregulated by the steroid hormone 20-hydroxyecdysone and microorganisms. *Insect molecular biology*, *20*(4), 519-527.

Zhang, Q., & Denlinger, D. L. (2010). Molecular characterization of heat shock protein 90, 70 and 70 cognate cDNAs and their expression patterns during thermal stress and pupal diapause in the corn earworm. *Journal of Insect Physiology*, *56*(2), 138-150.

Zhang, Y., Launay, H., Schramm, A., Lebrun, R., & Gontero, B. (2018). Exploring intrinsically disordered proteins in *Chlamydomonas reinhardtii*. *Scientific Reports*, *8*(1), 1-11.

Zhao, J. Y., Zhao, X. T., Sun, J. T., Zou, L. F., Yang, S. X., Han, X., Zhu, W.C., Yin, Q. & Hong, X. Y. (2017). Transcriptome and proteome analyses reveal complex mechanisms of reproductive diapause in the two-spotted spider mite, *Tetranychus urticae*. *Insect Molecular Biology*, *26*(2), 215-232.

Zheng, X., Krakowiak, J., Patel, N., Beyzavi, A., Ezike, J., Khalil, A. S., & Pincus, D. (2016). Dynamic control of Hsf1 during heat shock by a chaperone switch and phosphorylation. *elife*, *5*, e18638.

Zou, J., Guo, Y., Guettouche, T., Smith, D. F., & Voellmy, R. (1998). Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell*, *94*(4), 471-480.

Żwirowski, S., Kłosowska, A., Obuchowski, I., Nillegoda, N. B., Piróg, A., Ziętkiewicz, S., Bukau, B., Mogk, A. & Liberek, K. (2017). Hsp70 displaces small heat shock proteins from aggregates to initiate protein refolding. *The EMBO journal*, *36*(6), 783-796.

8. Supplementary materials

Table 27. Primer pairs used for qPCR analysis.

	Gene	Primer sequence (5'-3')	Accession
1.	*actin*	F: CAGAAGGAAATCACAGCTCTAGCC R: ATCGTACTCCTGTTTCGAGATCCA	EL928709.1
2.	*rps3*	F: GTCGCAGAATTCGTGAACTAACCT R: ATGATGAACCTCAGCACACCATAG	EL929086.1
3.	*hsp90*	F: CAAGATCGTTCTTCACATCAAGGAG R: CGTCCTCTTTCTTCTCTTCTTCAGC	EL929806.1
4.	*hsc70*	F: CCCCACGAAGCAGACGCAGA R: TCGACGGCGGACACGTTGAG	JF708084
5.	*hsp70*	F: GCACAGGCCGCAGCAAGAAC R: AGGGCTTGTCGCACGCTGAA	XM_028309302
6.	*hsp20.4*	F: CGAAGAAAGTATCAGACGTGTCCAA R: TAAATGCAACGCATCACGAGATTAC	AB568467.1
7.	*hsp20.1*	F: CAGCGCTAAAGAATGAAAGGTCTGT R: TAGGTATCTCTCATTTCGCCTGTCC	AB568468.1
8.	*tropmy2*	F: GTGCCGAGAAGGCCGAGGAG R: GGATGCGTCGGTTCAGGGCA	XM_028309660
9.	*tnt*	F: AGCCCTCACAGGCAAGCACC R: CGGCTCTCTCCTGCCACACC	XM_028316365
10.	*thym*	F: ATGAAGCACGCCACCACGGA R: GTCCTTGGTGGGGAGGGGGT	XM_028305412
11.	*moe*	F: CCGCGAGGAGTGGGAGCAAA R: GTCGACGCCCAGCCAGAGTT	XM_028309570

Table 28. Results of gene expression analysis in *O. nubilalis* larvae. The results are presented as means of calculated ΔCt values ± standard error of mean obtained from three independent biological pools per experimental group repeated in technical replicates, where one pool was comprised of 3 larvae.

	hsp90	*hsc70*	*hsp70*	*hsp20.4*	*hsp20.1*
ND	-0.14 ± 0.06	-0.11 ± 0.07	-2.21 ± 0.02	-2.22 ± 0.06	-1.87 ± 0.06
D(15)	-0.63 ± 0.05	0.18 ± 0.06	-1.63 ± 0.07	-1.97 ± 0.08	-4.13 ± 0.11
D(5)	0.04 ± 0.01	0.17 ± 0.002	-1.01 ± 0.01	-1.73 ± 0.03	-4.19 ± 0.03
D(-3)	0.20 ± 0.01	0.34 ± 0.05	-0.71 ± 0.03	-1.44 ± 0.04	-3.53 ± 0.05
D(-16)	0.18 ± 0.02	0.28 ± 0.04	-0.55 ± 0.02	-1.74 ± 0.02	- 3.64 ± 0.02

	tropmy2	*tnt*	*thym*	*moe*
ND	-0.92 ± 0.06	-0.79 ± 0.03	-0.99 ± 0.09	-1.26 ± 0.04
D(15)	-0.43 ± 0.08	-0.83 ± 0.06	-1.91 ± 0.05	-1.51 ± 0.05
D(5)	-0.61 ± 0.03	-1.03 ± 0.006	-2.22 ± 0.002	-1.44 ± 0.017
D(-3)	-0.55 ± 0.03	-1.01 ± 0.01	-2.15 ± 0.04	-1.33 ± 0.004
D(-16)	-0.77 ± 0.01	-1.34 ± 0.03	-1.90 ± 0.07	-1.15 ± 0.02

9. Extended abstract in Serbian

Uvod: Strukturno-neuređeni proteini (*engl.* intrinsically disordered proteins, IDPs) su proteini koji pri fiziološkim uslovima sredine ne poseduju sekundarnu i tercijarnu strukturu, nego se nalaze u relativno denaturisanom obliku. Iz tog razloga oni poseduju povišenu otpornost na takve sredinske uslove koji bi doveli do gubitka strukture globularnih proteina (Uversky i sar., 2001; Uversky, 2009). Takođe, strukturna neuređenost može biti prisutna u ograničenim količinama i kod tipično strukturno-uređenih proteina, u vidu strukturno-neuređenih regiona (*engl.* intrinsically disordered regions, IDRs) (Necci i sar., 2016). Takvi regioni si veoma česti u sekvencama proteina koji učestvuju u regulaciji signalnih puteva (Iakoucheva i sar., 2002; Xue i sar., 2012), a često se nalaze na C- i N- krajevima proteina (Lobley i sar., 2007; Pentony i Jones, 2010; Vuzman i sar., 2010). Osobenost IDP-ova da ne poseduju strukturu pri fiziološkim uslovima sredine je zapisana u njihovoj primarnoj, aminokiselinskoj strukturi. Analize aminokiselinskog sastava ovih proteina su pokazale da su oni obogaćeni u sadržaju polarnih i naelektrisanih aminokiselina poput lizina i glutaminske kiseline, a imaju smanjen udeo nepolarnih i hidrofobnih aminokiselina poput triptofana i fenilalanina (Uversky i sar., 2000; Williams i sar., 2001; Campen i sar., 2008). Kao posledica ovakvog aminokiselinskog sastava, IDP-ovi poseduju visoko ukupno naleketrisanje i sniženu hidrofobnost, što im onemogućava da se saviju i zadobiju više nivoe strukture pri fiziološkim uslovima (Uversky i sar., 2000; Uversky, 2011, 2019). U nepovoljnim uslovima životne, kao što su visoka/niska temperatura ili izrazito kisela/bazna pH, globularni proteini mogu denaturisati i izgubiti svoju funkciju (Tanford, 1968). Strukturno-neuređeni proteini, sa druge strane, pokazuju drugačiji odgovor na takve uslove. Pri visokim temperaturama, povećavaju se hidrofobne interakcije između aminokiselina koje dovode do delimičnog savijanja IDP-ova i privremenog zadobijanja sekundarnih struktura (Uversky, 2009, 2013a). U uslovima niskih temperatura, aktivnost IDP-ova je u velikoj meri očuvana u poređenju sa globularnim proteinima kod kojih se ona može izgubiti uzastopnim zamrzavanjem i odmrzavanjem (Tantos i sar., 2009). U uslovima izrazito kisele ili bazne pH sredine, kod IDP-ova dolazi do maskiranja naelektrisanja čime se uklanjaju elektrostatička odbijanja između susednih aminokiselina i protein može da se privremeno savije u sekundarnu strukturu (Uversky, 2009, 2013a; Smith i Jelokhani-Niaraki, 2012). Različitim *in silico* analizama je utvrđeno da su IDP-ovi veoma zastupljeni u živom svetu, kao i kod virusa. Postoji povezanost između stepena kompleksnosti organizma, u smislu da li je prokariotski ili eukariotski, i sadržaja IDP-ova u proteomu. Prisutni su u većoj meri kod eukariota, kod kojih i do 50% proteina može sadržati barem jedan IDR (Uversky, 2011; Pancsa i Tompa,

2012; Xue i sar., 2012; Peng i sar., 2015; Kulkarni i Uversky, 2018). Funkcije koje obavljaju IDP-ovi i proteini sa IDR-ovima se zasnivaju na njihovom nedostatku strukture. U neuređenom obliku oni mogu da prepoznaju različite ciljne molekule i da potom zadobiju strukturu nakon što stupe u kontakt sa njima. Iz toga razloga često su uključeni u procese regulacije signalnih puteva, transkripcije, translacije i ćelijske diferencijacije (Tompa, 2012; Wright i Dyson, 2015; Bondos i sar., 2021). Strukturna neuređenost je korelisana i sa proteinima koji obavljaju funkciju molekularnih šaperona (Reichmann i sar., 2012).

Kukuruzni plamenac *Ostrinia nubilalis* (Hbn, 1796) je polifagna vrsta moljca koja je rasprostranjena na prostorima Evrope, severne Afrike i zapadne Azije. Takođe, slučajno je introdukovana na severnoamerički kontinent početkom 20. veka, verovatno putem intenzivne trgovine žitaricama između Sjedinjenih Američkih Država i određenih evropskih zemalja (Smith, 1920; Caffrey i Worthley, 1927; Bethenod i sar., 2005). Ova vrsta je ekonomski značajna jer se njene gusenice hrane listovima i drugim delovima biljaka kao što su kukuruz, krompir, paradajz i preko 200 drugih biljnih vrsta (Capniera, 2000; Kuhar i sar., 2004; Sole i sar., 2010). Kukuruzni plamenac je holometabolni insekt i tokom svog životnog ciklusa prolazi kroz potpunu metamorfozu ili preobražaj. Životni ciklus kukuruznog plamenca se, s toga, sastoji od četiri razvojna stupnja: jaje, larva odnosno gusenica, lutka i odrasla jedinka odnosno moljac ili imago (Caffrey i Worthley, 1927; Capinera, 2000). U toku jedne godine uglavnom se javljaju dve generacije ovog moljca – jedna letnja i jedna zimska, ali su zabeleženi slučajevi kada ih je bilo tri i više. Iz tog razloga, kukuruzni plamenac se smatra multivoltnom vrstom (Capinera, 2000; Vajgand, 2010). Odrasle jedinke poslednje letnje generacije koja se javlja u toku godine polažu jaja iz kojih će se izleći gusenice zimske generacije. Nakon izleganja gusenice se intenzivno hrane i uglavnom prolaze kroz četiri presvlačenja i pet razvojnih stupnjeva ili instara (Caffrey i Worthley, 1927; Capinera, 2000). U ovom periodu godine dolazi do skraćenja dužine dana, što služi kao okidač gusenicama petog stupnja da se pripreme za dolazak zime. Taj period je praćen padom temperature i smanjenjem količine dostupne hrane. Da bi sačuvale ograničene energetske rezerve, usled prestanka hranjenja, gusenice ulaze u specifično stanje hipometabolizma koje se naziva dijapauza (Beck, 1962; Hahn i Denlinger, 2007). U okviru ove faze mirovanja dolazi do opšteg smanjenja stope metabolizma gusenica, kao i do promena na biohemijskom, fiziološkom i molekularnom nivou u cilju očuvanja energetskih rezervi i razvijanja otpornosti na hladnoću (Danks, 1987; Rinehart i sar., 2000; Denlinger, 2009). Među tim promenama su ekspresija

specifičnih gena u vezi sa otpornošću na stres, prestanak kretanja, sinteza krioprotektivnih molekula, povećanje količine glicerola u hemolimfi, izmene u lipidnom sastavu ćelijskih membrana, prelazak sa aerobnog na anaerobni metabolizam i mnoge druge (Grubor-Lajšić i sar., 1991; Kojić, 2009; MacRae, 2010; Vukašinović i sar., 2013, 2015, 2018; Popović, 2014, Popović i sar., 2015; Uzelac i sar., 2020). Pošto je životni ciklus kukuruznog plamenca dobro opisan, ova vrsta je pogodna za indukovanje ulaska u stanje mirovanja i sticanje otpornosti na hladnoću, usled čega se ona koristi kao pouzdan model-sistem za proučavanje molekularnih mehanizama koji upravljaju pomenutim procesima.

Ciljevi istraživanja: Dosadašnje naučne studije pokazale su da IDP-ovi imaju izraženu termalnu stabilnost kako na visokim tako i na niskim temperaturama. Takođe, pokazale su i da pojedini IDP-ovi bolje održavaju svoju funkciju na niskim temperaturama u poređenju sa uređenim proteinima. Zbog toga, budući da je *O.nubilalis* organizam čije larve petog stupnja stiču otpornost na niske temperature tokom dijapauze, ova vrsta predstavlja idealan model-sistem za proučavanje uloge IDP-ova u molekularnoj ekofiziologiji otpornosti na niske temperature u *in vivo* uslovima.

Hipoteza ovog rada je da će u uslovima niskih temperatura u fazi mirovanja gusenica *O. nubilalis* biti povišen sadržaj IDP-ova, kao posledica postepenog razvoja otpornosti na niske temperature. Uzimajući u obzir jedinstvenost primarne strukture IDP-ova i proteina sa dugačkim IDR-ovima, njene posledice na termalnu stabilnost ovih proteina kao i nedostatak informacija o njihovom prisustvu i značaju kod organizama koji imaju razvijene sposobnosti za preživljavanje na niskim temperaturama životne sredine, kako bi se dokazala hipoteza ciljevi istraživanja bili su da se: 1) utvrdi njihovo prisustvo kod *O. nubilalis* optimizacijom postojećih metoda za njihovu detekciju; 2) da se analizira njihov sadržaj u različitim eksperimentalnim uslovima u kojima je na kontrolisan način organizam uveden u stanje mirovanja u kome je indukovan postepen razvoj otpornosti na hladnoću; 3) da se bioinformatičkim analizama odrede kvalitativni i kvantitativni pokazatelji strukturne neuređenosti identifikovanih IDP-ova i proteina sa dugačkim IDR-ovima, i da se proceni stepen te neuređenosti; 4) da se bioinformatičkim analizama odredi (potencijalna) molekularna funkcija identifikovanih IDP-ova i proteina sa dugačkim IDR-ovima; 5) da se procene postoje li značajne promene u kvalitativnom i kvantitativnom sastavu ovih proteina u različitim eksperimentalnim uslovima koji su manje ili više pogodovali razvoju otpornosti na niske temperature; 6) da se analizira ekspresija gena onih IDP-ova i proteina sa dugačkim IDR-ovima koji imaju ili značajan stepen strukturne neuređenosti i/ili poznatu značajnu biološku ulogu u uslovima termalnog stresa; 7) da se na osnovu dobijenih podataka svih gore navedenih analiza dobije uvid u ulogu IDP-ova i

proteina sa dugačkim IDR-ovima u ekofiziološkoj osnovi dijapauze i razvoja otpornosti na niske temperature kod *O.nubilalis*.

Materijal i metode: Za potrebe istraživanja u ovoj disertaciji izvedene su dve eksperimentalne postavke – Pilot i Glavna eksperimentalna postavka. U okviru Pilota, dijapauzirajuće gusenice 5. instara su sakupljene sa oglednih polja Instituta za kukuruz „Zemun Polje“ tokom zimske sezone 2018/2019. godine. Oformljene su dve eksperimentalne grupe i svaka se sastojala od 3 biološka ponavljanja sa po 5 gusenica: 1) Dnca – neaklimatizovana dijapauzirajuća grupa i 2) Dca – dijapauzirajuća grupa aklimatizovana na niske temperature. Nakon uzorkovanja, gusenice su držane dve nedelje na 15°C i deo je zamrznut u tečnom azotu na –80°C do daljih analiza (Dnca grupa). Preostale gusenice su držane u dijapauzirajućim uslovima (12h dan: 12h noć) i postepeno hlađene spuštanjem temperature za 1°C svakog dana, uz dvonedeljno zadržavanje na specifičnim temperaturama (5, –3 i –16°C) da bi se aklimatizovale (Slika 3). Nakon poslednje aklimatizacije gusenice su zamrznute u tečnom azotu na –80°C do daljih analiza (Dca grupa). U okviru Glavnog eksperimenta, dijapauzirajuće gusenice 5. instara su sakupljene iz zaraženih stabljika kukuruza dobijenih od Instituta za ratarstvo i povrtarstvo u Novom Sadu tokom zimske sezone 2019/2020. godine. Deo gusenica je ostavljen u stabljikama da završi svoj životni ciklus i produkuje generaciju nedijapauzirajućih larvi koje predstavljaju ND – nedijapauzirajuću kontrolnu grupu. Ostatak gusenica je stavljen u staklene posude sa hranljivom podlogom i prenet u inkubatore da se aklimatizuju na niske temperature. Proces aklimatizacije je tekao po istom principu i u istim dijapauzirajućim uslovima kao i u Pilotskoj postavci, uz dodatno uzorkovanje nakon svake dvonedeljne aklimatizacije (Slika 4). Nakon svakog uzorkovanja gusenice su zamrznute u tečnom azotu na –80°C i formirane su još četiri eksperimentalne grupe: 1) D(15) – dijapauzirajuće gusenice aklimatizovane dve nedelje na 15°C, 2) D(5) – dijapauzirajuće gusenice aklimatizovane dve nedelje na 5°C, 3) D(–3) – dijapauzirajuće gusenice aklimatizovane dve nedelje na –3°C i 4) D(–16) – dijapauzirajuće gusenice aklimatizovane dve nedelje na –16°C. Svaka eksperimentalna grupa se sastojala od 4 biološka ponavljanja sa po 3 gusenice – 1 biološki replikat za proteinske analize i 3 biološka replikata za ispitivanje relativne genske ekspresije.

Cele gusenice su homogenizovane pomoću tučka i avana u ledeno hladnom 50 mM fosfatnom puferu pH 7.5 sa dodatim 1 mM DTE i dodatno homogenizovani sonikacijom pomoću Qsonica Q500 uređaja. Grubi homogenati su prečišćeni centrifugiranjem i supernatanti su podeljeni u po dve nove tubice po uzorku. Po jedna tubica od svakog uzorka je iskorišćena za termalnu obradu

da se uklone globularni proteini i obogati sadržaj IDP-ova (zagrevani uzorci). Ostale tubice nisu termalno obrađene radi kontrole i poređenja (nezagrevani uzorci). Nakon toga, svi uzorci su dodatno prečišćeni još jednim centrifugiranjem i koncentracija proteina u njima je određena metodom po Bradfordu. Da bi se potvrdila uspešnost termalne obrade uzoraka i obogaćivanja frakcije IDP-ova, uzorci su razdvojeni na specijalizovanoj dvodimenzionalnoj elektroforezi na poliakrilamidnom gelu (*engl.* polyacrylamide gel electrophoresis, PAGE) u denaturišućim uslovima kako bi se razdvojili IDP-ovi od uređenih proteina. Nakon razdvajanja, uzorci su obojeni srebrom korišćenjem Pierce™ Silver Stain Kit (Thermo Scientific, Valtam, MA, SAD) po uputstvu proizvođača radi detektovanja proteina.

Za identifikaciju proteina korišćena je tečna hromatografija visoke efikasnosti kuplovana sa tandemskom masenom spektrometrijom (*engl.* liquid chromatography tandem mass spectrometry, LC-MS/MS). Uzorci su pre masene spektrometrije pripremljeni za analizu digestijom pomoću proteinaza tripsina i Lys-C. Digerirani peptidi su analizirani LC-MS/MS-om, a potom Mascot softverom za pretraživanje. S obzirom na ograničenu količinu podataka o proteinskim sekvencama *O. nubilalis* u bazama podataka, pretraga je proširena i na sekvence koje pripadaju srodnoj vrsti *O. furnacalis*, kao i svim ostalim sekvencama lepidoptera dostupnim u NCBI bazi podataka (*engl.* National Center for Biotechnology Information). Nakon identifikacije proteina, preuzete su FASTA sekvence za sve identifikovane proteine i njihova strukturna neuređenost je predviđena pomoću IUPred platforme (https://iupred3.elte.hu/) (Dosztányi i sar., 2005). Na osnovu njihove predviđene strukturne neuređenosti, proteini su klasifikovani u četiri kategorije: 1) OP-ovi – globularni proteini, procenat neuređenosti do 10% (*engl.* ordered proteins), 2) NOP-ovi – skoro uređeni proteini, procenat neuređenosti 10-30% (*engl.* nearly ordered proteins), 3) PDP-ovi – delimično neuređeni proteini, procenat neuređenosti 30-70% (*engl.* partially disordered proteins) i 4) MDP-ovi – većinski neuređeni proteini, procenat neuređenosti preko 70% (*engl.* mostly disordered proteins). Proteini su takođe analizirani u pogledu sadržaja dugačkih strukturno-neuređenih regiona (*engl.* long IDRs) od barem 20 uzastopnih aminokiselina za koje se smatra da promovišu neuređenost (*engl.* disorder-promoting). Pored toga, analiziran je aminokiselinski sastav ovih proteina, kao i njihove funkcije korišćenjem podataka dostupnih u bazama kao što su Interpro (https://www.ebi.ac.uk/interpro/, Blum i sar., 2021), UniProt (www.uniprot.org, Consortium, 2021), Pfam (http://pfam.xfam.org/, Mistry i sar., 2021) i GeneOntology (http://geneontology.org/, Ashburner i sar., 2000).

Za potrebne analize relativne ekspresije gena, ukupna RNK je izolovana iz uzoraka Glavne eksperimentalne postavke korišćenjem TRIzol reagensa (Invitrogen, Valtam, MA, SAD) prema uputstvu proizvođača. Nakon određivanja koncentracije izolovane ukupne RNK, komplementarna DNK je sintetisana korišćenjem kompleta High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Valtam, MA, SAD) prema uputstvu proizvođača. Za analizu relativne genske ekspresije geni su odabrani na osnovu procenjenog sadržaja strukturne neuređenosti proteina koje kodiraju, kao i njihovih funkcija. Odabrani su sledeći geni: 1) *actin* i *rps03* kao kontrolni geni i 2) *hsp90*, *hsp70*, *hsc70*, *hsp20.4*, *hsp20.1*, *tropomyosin-2* (*tropmy2*), *troponin t* (*tnt*), *thymosin beta* (*thym*) i *moesin/ezrin/radixin homolog 1* (*moe*). Početnice za ove gene su dizajnirane pomoću programa Primer-BLAST (Ye i sar., 2012), a poručene od kompanije Vivogen LLC (Beograd, Srbija). Specifičnost početnica je proverena na agaroznom gelu, a efikasnost utvrđena metodom kvantitativne lančane reakcije polimeraze (*engl.* quantitative polymerase chain reaction, qPCR) na seriji razblaženja komplementarne DNK (1:1, 1:10, 1:100, 1:1000 i 1:10000). Za qPCR analizu korišćen je reagens 2X Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific, Valtam, MA, SAD) i aparat CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, CA, SAD). Relativna genska ekspresija odabranih gena je određena u svim eksperimentalnim grupama i izračunata prema metodi koju su opisali Ganger i sar. (2017), a statistička obrada dobijenih podataka je urađena pomoću programa Statistica verzija 14.0 (StatSoft, Inc., Tulsa, OK, SAD). Statističke značajne razlike su utvrđene jednofaktorskom analizom varijanse (*engl.* one-way analysis of variance, ANOVA) i *post-hoc* Tuckey-evim testom za nivo značajnosti $p < 0.05$.

Rezultati: Kako bi se proverila uspešnost postupka termalne obrade uzoraka radi obogaćivanja sadržaja IDP-ova, proteini iz zagrevanih i nezagrevanih uzoraka su razdvojeni pomoću specifične 2D PAG elektroforeze. Po rezultatima (Slika 5) može se videti da je termalna obrada bila uspešna, kao i da je zaista neophodan korak pripreme pre daljih analiza. Kod nezagrevanih uzoraka (Slika 5A) nije došlo do razdvajanja u drugoj dimenziji, već su proteini ostali uglavnom u prvom gelu. Kod zagrevanih uzoraka, došlo je do prelaska proteina u drugi gel (Slika 5B) i IDP-ovi se mogu videti duž crne dijagonale. Može se uočiti i određeni broj termostabilnih uređenih proteina koji se nalaze iznad crne dijagonale.

U ovom istraživanju identifikovano je ukupno 2103 proteina – 820 u Pilotu i 1283 u Glavnoj postavci. Kada se odbace dvostruki unosi, u Pilotu je identifikovano 608 jedinstvenih proteina, a u Glavnoj postavci 600. I u jednoj i u drugoj postavci skoro 50% unosa je pripadalo proteinima čije je postojanje samo predviđeno pomoću postojećih nukleotidnih sekvenci. Među eksperimentalnim grupama u Pilotu, 294 proteina je identifikovano samo u Dnca grupi, 102 samo u Dca grupi, a 212 je nađeno u obe (Slika 6A, Tabela 11). Kod Glavne postavke, 218 je nađeno samo u ND grupi, 142 samo u D(15) grupi, 171 samo u D(5) grupi, 158 samo u D(–3) grupi, 164 samo u D(–16) grupi, a preostalih 68 u svim grupama (Slika 6B, Tabela 11). Kada se sagledaju rezultati na nivou cele studije, identifikovano je 892 jedinstvena proteina, od čega 292 samo u Pilotu, 284 samo u Glavnoj postavci, a 316 je nađeno u obe postavke. Termalna obrada uzoraka je dovela do identifikacije dodatnih jedinstvenih proteina u svim eksperimentalnim grupama koji nisu pronađeni u nezagrevanim uzorcima (Slika 7, Tabela 12). Takođe, u zagrevanim uzorcima je sveukupno identifikovano manje proteina nego u nezagrevanim uzorcima. U proseku, u nezagrevanim uzorcima je identifikovano oko 50% više jedinstvenih proteina u odnosu na zagrevane uzorke. Jedini izuzetak je ND grupa iz Glavne postavke, gde je broj identifikovanih uzoraka bio sličan u oba tipa uzorka – 138 u nezagrevanom i 129 u zagrevanom uzorku.

Procentualna strukturna neuređenost identifikovanih proteina je predviđena korišćenjem IUPred platforme i na osnovu nje su svrstani u jednu od četiri grupe – OP-ovi, NOP-ovi, PDP-ovi ili MDP-ovi. U svim eksperimentalnim grupama najveći broj proteina je svrstan u grupu OP-ova. Najveći broj proteina iz ove kategorije je utvrđen u Dnca grupi Pilota, kao i u ND grupi Glavne postavke – 319, odnosno 194 proteina (Slika 8, Tabela 13). Dijapauzirajuće grupe u Glavnoj postavci su se uglavnom razlikovale po broju OP proteina, dok su po broju preostalih grupa – NOP-ova, PDP-ova i MDP-ova bile relativno slične (Tabela 13). Gledajući procentualni udeo pojedinačnih grupa proteina u ukupnom broju (Slika 9), može se videti da su njihovi odnosi relativno stabilni između eksperimentalnih grupa kako Pilota (Slika 9A), tako i Glavne postavke (Slika 9B). Termalna obrada uzoraka je imala izuzetno pozitivan efekat na udeo onih proteina koji se smatraju više neuređenim (PDP-ovi i MDP-ovi). I u jednom i u drugoj eksperimentalnoj postavci, termalnom obradom je skoro prepolovljen udeo proteina iz OP kategorije (sa oko ~80% na oko ~40% udela) u ukupnom broju proteina, dok je značajno porastao udeo PDP i MDP proteina (Slike 10 i 11), a delimično je obogaćen i udeo proteina iz NOP kategorije. Identifikovani proteini su analizirani i po pitanju sadržaja dugačkin IDR-ova, regiona koji sadrže makar 20 uzastopnih neuređenih

aminokiselina, pošto ovi regioni često imaju specifične molekularne funkcije. Proteini iz kategorije OP-ova prednjače po odsustvu dugačkih IDR-ova (preko 90%) u svim eksperimentalnim grupama (Slike 12 i 13). Sa druge strane, udeo proteina iz ostale tri kategorije (NOP-ovi, PDP-ovi i MDP-ovi) koji sadrže makar jedan dugačak IDR je u proseku viši kod onih eksperimentalnih grupa koje su izlagane niskim temperaturama (Slika 12A nasuprot B i Slika 13A nasuprot B, C, D i E).

Analiziran je aminokiselinski sastav svih identifikovanih proteina kako bi se utvrdio udeo pojedinačnih uređenih i neuređenih aminokiselina u njihovim primarnim strukturama. Proteini iz kategorije MDP-ova su u proseku bili sastavljeni od 68.9% neuređenih i 31.1% uređenih aminokiselina. Slično je bilo i sa proteinima iz kategorije PDP-ova – 64.9% neuređenih i 35.1% uređenih. Taj odnos je nastavio da se smanjuje kada se pogledaju sastavi proteina iz NOP i OP kategorija, ali je uvek bio viši udeo neuređenih aminokiselina (Slika 14). Od pojedinačnih aminokiselina, proteini iz MDP i PDP kategorija su posebno bili obogaćeni u udelu glutamata (11.91%, odnosno 11.64%), lizina (9.34%, odnosno 9.26%) i glutamina (6.51%, odnosno 5.26%) u odnosu na udele tih aminokiselina kod proteina iz NOP i OP kategorija. Takođe, MDP proteini su bili izrazito obogaćeni u sastavu prolina u odnosu na sve ostale kategorije, dok su PDP proteini bili obogaćeni u udelu leucina koji se smatra uređenom aminokiselinom (Slika 15).

Kako bi se stekao uvid u biološki značaj IDP-ova za procese adaptacije na niske temperature kod kukuruznog plamenca, identifikovani IDP-ovi su podvrgnuti bioinformatičkoj analizi koristeći podatke iz baza podataka UniProt, Pfam, Interpro i Gene Ontology. Ova analiza je pokazala da postoje podaci za samo 458 proteina iz ove disertacije, od čega je 143 proteina koji su ili iz kategorije PDP-ova i MDP-ova ili sadrže makar 1 dugački IDR. Najviše proteina je bilo povezano sa funkcijama citoskeletalnih elemenata (40), zatim slede molekularni šaperoni (32) i proteini uključeni u metabolizam proteina i aminokiselina (15) (Slika 16). Manji deo proteina je povezan sa ulogama u elektron-transportnom lancu, metabolizmu lipida i razvojnim procesima (Slika 17).

Analiza relativne ekspresije gena koji kodiraju odabrane proteine sa strukturnom neuređenošću je izvedena na uzorcima iz Glavne postavke, a kao kontrola je korišćena ND eksperimentalna grupa. Relativna ekspresija skoro svih analiziranih gena je generalno bila povišena kod dijapauzirajućih grupa u odnosnu na kontrolnu grupu, posebno gena za proteine toplotnog stresa (*engl.* heat shock proteins. Kada su u pitanju ovi geni, njihova ekspresija u svim grupama prati sličan trend, a to je da raste kako su gusenice izlagane sve nižim temperaturama i što su duže bile u dijapauzi (Slike

19, 21 i 22). Jedino je relativna ekspresija *hsc70* gena bila na konstantno visokom nivou u svim dijapauzirajućim grupama (Slika 20), dok je ekspresija *hsp20.1* imala obrnut trend u odnosu na ostale HSP gene. Ovaj gen je imao povišenu ekspresiju u ND grupi u odnosu na dijapauzirajuće grupe (Slika 23). Takođe, ekspresija ovog gena je bila niža na početku dijapauze kada gusenice nisu još izložene temperaturama ispod nule (Slika 23, grupe D(15) i D(5)), a potom raste kako dijapauza odmiče i temperatura još više opada (grupe D(–3) i D(–16). Relativna ekspresija dva strukturna gena koji su analizirani (*tropmy2* i *tnt*) se dijametralno razlikuju. Ekspresija *tropmy2* je generalno visoka u dijapauzirajućim i hlađenim gusenicama u odnosu na ND grupu (Slika 24), dok je ekspresija *tnt* obrnutog karaktera – povišena u nedijapauzi, a zatim se snižava sa padom temperature i odmicanjem programa dijapauze (Slika 25). Relativna ekspresija *thym* gena je takođe snižena u svim dijapauzirajućim grupama u odnosu na ND kontrolu (Slika 26), dok je ekspresija *moe* gena niža na početku dijapauze u odnosu na ND grupu, ali raste sa izlaganjem niskim temperaturama i progresijom dijapauze, te je u grupi D(–16) ekspresija ovog gena viša od ND kontrole (Slika 27).

Diskusija: Usled svoje strukturne neuređenosti, IDP-ovi nisu najpogodniji za strukturne analize tipičnim metodama strukturne biologije kao što su kristalografija X-zracima i kriogenična elektronska mikroskopija (Piovesan i sar., 2017; Bari i Prakashchand, 2021; Avramov i sar., 2022). Međutim, upravo ta strukturna neuređenost im pruža otpornost na denaturišuće uslove sredine (Uversky, 2009), što se može iskoristiti za obogaćivanje uzoraka ovim proteinima i njihovu dalju analizu. Obogaćivanje sadržaja IDP-ova se može postići na nekoliko načina, kao što su korišćenje precipitirajućih agenasa ili zagrevanje uzoraka. Nakon bilo kojeg od ova dva tretmana, denaturisani proteini se mogu ukloniti centrifugiranjem, a u uzorku ostaje povišena frakcija IDP-ova (Cortese i sar., 2005, Galea i sar,. 2006, 2009; Romero-Pérez i sar., 2023). U ovoj disertaciji korišćen je pristup zagrevanja uzorka zbog jednostavnosti postupka i zadovoljavajućeg uklanjanja globularnih proteina usled čega se mogu otkriti drugi proteini koji bi u suprotnom bili zamaskirani od strane mnogobrojnijih uređenih proteina (Galea i sar., 2006; Zhang i sar., 2018; Avramov i sar., 2022). Kako bi se potvrdila uspešnost postupka termalne obrade uzoraka, korišćena je modifikovana metoda 2D PAG elektroforeze koja služi za razlikovanje IDP-ova od uređenih proteina. Ona se zasniva na korišćenju 8M uree tokom razdvajanja u drugoj dimenziji, koja utiče na kretanje globularnih proteina, a nema efekta na kretanje IDP-ova (Csizmók i sar., 2006; Tantos and Tompa, 2012). Strukturno-neuređeni proteini se u drugoj dimenziji raspoređuju po dijagonali

na gelu, dok je kretanje globularnih proteina otežano i obično se nađu iznad dijagonale. Na Slici 5 su efekti termalne obrade uzorka odmah uočljivi. Nezagrevani uzorci su bili bogati globularnim proteinima i nisu uspeli da pređu na drugi gel (Slika 5A). Ovi uzorci su verovatno bili bogati proteinima za skladištenje visoke molekulske mase zvanim arilforinima koji se u velikoj meri nalaze u hemolimfi insekata kao što je kukuruzni plamenac (Taški i sar., 2004). Zagrevanjem uzoraka ovi proteini su uklonjeni i razdvajanje proteina iz tih uzoraka je bilo uspešnije. Može se videti nekoliko grupa proteina koji su se rasporedili po dijagonali nakon razdvajanja u drugoj dimenziji (Slika 5B), što ukazuje na to da su to ili delimično ili većinski strukturno-neuređeni proteini (Csizmók i sar., 2006; Tantos and Tompa, 2012). Takođe se mogu primetiti i neki uređeni proteini koji su termostabilni, iznad dijagonale (Slika 5B).

Identifikacija proteina je urađena poređenjem spektara peptida dobijenih nakon tandemske masene spektrometrije sa sekvencama proteina kod *O. nubilalis* koje su dostupne u NCBI bazi podataka. Zbog ograničenog broja dostupnih sekvenci za ovu vrstu (~1600 unosa), pretraga je proširena i na srodnu vrstu *O. furnacalis* (~28 500 unosa) kao i sve druge vrste lepidoptera. *O. furnacalis* je odabrana zbog svoje veoma bliske srodnosti sa *O. nubilalis*, što se ogleda u zabeleženom protoku gena između ove dve vrste (Domingue i sar., 2008; Li i Yang, 2022). Na nivou celog istraživanja identifikovano je 2103 proteina i većina je identifikovana na osnovu sekvenci drugih vrsta osim *O. nubilalis*, što se i očekivalo. U ovu brojku su uključeni i duplikati iz različitih eksperimentalnih grupa, a nakon uklanjanja duplikata dolazi se do brojke od 892 jedinstvena proteina na nivou celog istraživanja. Od toga 292 proteina su nađena samo u postavci Pilota, 284 u Glavnoj postavci, a 361 u obe postavke. Eksperimentalni uslovi su uticali na broj identifikovanih proteina. U proseku manje proteina je identifikovano kod onih grupa koje su bile izlagane niskim temperaturama (Slika 6, Dca, D(15), D(5), D(–3) i D(–16) grupe). Ove razlike su usled eksperimentalne postavke, s obzirom na to da ulaskom u dijapauzu i sticanjem otpornosti na hladnoću dolazi do proteomskih promena kod ove vrste (Grubor-Lajšić i sar., 1991; Hahn and Denlinger, 2011; Kojić i sar., 2018; Uzelac i sar., 2020; Popović i sar., 2021; Avramov i sar., 2022), čime je takođe i omogućeno identifikovanje mnogih proteina koji su nađeni samo u tim grupama (Slika 6). Takođe, treba naglasiti da je skoro polovina proteina identifikovana na osnovu sekvenci koje su bile vezane samo za predviđene proteine, čime je predviđanje njihovog postojanja potvrđeno. Zagrevanje uzoraka je dovelo do identifikacije dodatnih jedinstvenih proteina koji su bili prisutni samo u zagrevanim uzorcima, iako je zagrevanje očekivano smanjilo ukupan broj proteina koji su mogli biti

identifikovani u takvim uzorcima (Slika 7). Verovatno je da su ti proteini u nezagrevanim uzorcima bili zaklonjeni drugim proteinima koji su bili prisutni u mnogo većoj meri, što je onemogućilo detekciju manje prisutnih proteina putem LC-MS/MS. Neki proteini su potencijalno komponente višeproteinskih kompleksa, što takođe otežava njihovu detekciju. Zagrevanjem uzoraka se mogu ukloniti oni proteini koji su prisutni u velikom broju i rastaviti višeproteinski kompleksi, što omogućuje identifikaciju još jedinstvenih proteina (Avramov i sar., 2022).

Usled specifičnog aminokiselinskog sastava IDP-ova i proteina sa IDR-ovima, kao i njihove sklonosti ka određenim aminokiselinama, razvijene su različite *in silico* metode na osnovu tih aminokiselinskih specifičnosti za analiziranje različitih aspekata IDP-ova i proteina sa IDR-ovima (He i sar., 2009; Deng i sar., 2012; Eickholt i Cheng, 2013; Jones and Cozzetto, 2015; Necci i sar., 2021). U ovoj disertaciji korišćeni su IUPred platforma (Dosztányi i sar., 2005) kao i specifične kompjuterske skripte za analizu ukupne i lokalizovane strukturne neuređenosti identifikovanih proteina, njihovog aminokiselinskog sastava, kao i uradila njihova funkcionalna karakterizacija. Analizirani proteini su na osnovu predviđene strukturne neuređenosti, izražene u procentima podeljeni u četiri kategorije. Proteini sa makar 70% neuređenosti su svrstani među većinski-neuređene proteine (MDP-ovi), oni sa neuređenošću između 70% i 30% su smešteni među delimično neuređene proteine (PDP-ovi), između 30% i 10% u skoro uređene proteine (NOP-ovi), dok su oni čija je neuređenost ispod 10% svrstani među uređene proteine (OP-ovi). Sagledavajući celokupne rezultate, može se videti da više od 30% identifikovanih proteina u obe postavke pripadaju ili MDP-ovima i PDP-ovima ili je strukturna neuređenost kod njih lokalizovana (Slike 8, 12 i 13). Ovi rezultati su u saglasnosti sa prethodnim istraživanjima koja su pokazala da je makar jedna trećina eukariotskih proteoma sa značajnim sadržajem strukturne neuređenosti (Pancsa and Tompa, 2012; Habchi i sar., 2014; Xue i sar., 2014; Peng i sar., 2015; Uversky, 2015; Bondos i sar., 2021). To što je za velik broj proteina koji su svrstani u kategoriju OP-ova izračunato da imaju procenat neuređenosti iznad 0% nije iznenađujuće, pošto su potpuno uređeni proteini izuzetno retki (Gall i sar., 2007). Potreba za termalnom obradom uzoraka pre ovakvih analiza je i ovde potvrđena, pošto je u zagrevanim uzorcima udeo strukturno-neuređenih proteina mnogostruko uvećan, pogotovo onih iz kategorije MDP-ova (Slike 10 i 11). Takođe, zabeležen je i značajan broj uređenih proteina (OP-ovi) upravo u zagrevanim uzorcima, što ukazuje na to da su ovi proteini potencijalno termostabilni. Ovakva poređenja zagrevanih i nezagrevanih uzoraka se često izostavlja u istraživanjima baziranim na obogaćivanju sadržaja IDP-ova (Cortese i sar., 2005;

Galea i sar., 2006; Galea i sar., 2009; Zhang i sar., 2018; Romero-Pérez i sar., 2023), dok rezultati ove disertacije ukazuju na prednosti ovakvog pristupa u povećanju broja proteina koji mogu biti podvrgnuti daljim analizama.

Pošto i uređeni proteini mogu posedovati neuređene segmente, svi proteini su analizirani za sadržaj dugih IDR-ova. To su nizovi od makar 20 uzastopnih neuređenih aminokiselina i za koje je utvrđeno da poseduju značajne biološke funkcije, kao što su molekularno prepoznavanje i vezivanje ili omogućavanje kretanja proteinskih domena (Oldfield i Dunker, 2014; Necci i sar., 2016; Davey, 2019; Deiana i sar., 2019). Prvo su analizirani proteini iz kategorija PDP-ova, NOP-ova i OP-ova jer oni čine većinu identifikovanih proteina u obe eksperimentalne postavke. Za preko 2/3 ovih proteina je pokazano da ne poseduju dugačke IDR-ove, od čega je većina iz kategorije OP-ova (Slike 12 i 13). Više od polovine NOP-ova nije sadržalo nijedan dugački IDR, ali ni 20% PDP-ova takođe, što je isprva bilo iznenađujuće. Vrlo je verovatno da ti PDP-ovi poseduju regione koji imaju manje od 20 uzastopnih neuređenih aminokiselina, čime su tačno ispod praga detekcije, a opet su dovoljno dugački da putem IUPred platforme budu klasifikovani kao PDP-ovi čija je neuređenost između 30% i 70%. Slično je bilo i sa proteinima iz kategorije MDP-ova koji imaju preko 70% neuređenosti. Oko 20% MDP-ova iz Pilota i 14% iz Glavne postavke takođe nije posedovalo nijedan dugački IDR po ovim kriterijumima (Slika 12 i 13), verovatno iz istog razloga kao i oni PDP-ovi kod kojih nisu zabeleženi neuređeni regioni.

Analiziran je i aminokiselinski sastav svih proteina, s obzirom na to da od primarne strukture proteina zavisi da li će moći da se saviju u uređene proteine ili ne (Uversky i sar., 2000). Na osnovu njihovih osobina aminokiseline mogu biti svrstane u grupu onih koje pogoduju strukturnoj uređenosti (W, F, Y, I, M, L, V, N, C i T) ili strukturnoj neuređenosti (A, G, R, D, H, Q, K, S, E i P). Za triptofan se smatra da je aminokiselina koja najviše pogoduje strukturnoj uređenosti, a za prolin da najviše pogoduje strukturnoj neuređenosti (Campen i sar., 2008). Iz tog razloga, analize aminokiselinskog sastava identifikovanih proteina može pružiti vredne informacije o raspodeli pojedinačnih aminokiselina u ovoj skupini podataka i doprineti karakterizaciji proteina. Kada su u pitanju MDP-ovi i PDP-ovi, sadržaj neuređenih aminokiselina u njihovim primarnim strukturama je izuzetno visok, 68.9%, odnosno 64.9% (Slika 14). Pogotovo je visok udeo polarnih aminokiselina glutamata (E, 11.91%, odnosno 11.64%), lizina (K, 9.34% odnosno 9.26%) i glutamina (Q, 6.51%, odnosno 5.26%) u poređenju sa preostale dve grupe proteina koje su više

uređene. Takođe, MDP-ovi imaju izrazito povišen sadržaj prolina (7.95%) u odnosu na sve ostale kategorije proteina (Slika 15A). Sa druge strane, u primarnoj strukturi PDP-ova je prisutna uređena aminokiselina leucin u visokom procentu koji je skoro isto kao i kod uređenih proteina (Slika 15B). Verovatan razlog za ovakav rezultat je to kod mišićnih proteini, kao što su npr. tropomiozini za koje je utvrđeno da poseduju značajnu strukturnu neuređenost, aminokiselina leucin može činiti makar 10% od ukupnog broja aminokiselina u primarnoj strukturi, pošto ovi proteini poseduju leucinske cipzare na njihovim C-krajevima (Brown i sar., 2005). Takođe, postoje indikacije da sklonost ka određenim aminokiselinama može biti i specifična za datu vrstu (Zhang i sar., 2018), ali je potrebno još istraživanja u tom pravcu da bi se izveo konkretan zaključak.

Odmah nakon funkcionalne analize proteina pokazalo se da i dalje postoje velike rupe po pitanju informacija o anotaciji kako uređenih tako i neuređenih proteina. Od 892 jedinstvena proteina, jedva 50% njih je prisutno u UniProt bazi podataka i anotirano informacijama iz barem jedne od preostalih baza podataka, kao što je npr. Pfam baza. Kada se sagledaju proteini koji poseduju strukturnu neuređenost, samo 143 njih je moglo biti pokriveno ovom analizom. Slaba anotacija proteina kod vrsta leptira i moljaca je verovatno i posledica toga što za ove vrste ne postoje reprezentativni model organizmi osim svilene bube (International Silkworm Genome Consortium, 2008). Na osnovu onih proteina sa strukturnom neuređenošću koji su mogli biti funkcionalno analizirani, utvrđeno je da oni obavljaju širok dijapazon molekularnih funkcija i da su uključeni u veoma raznolike biološke procese. Velika većina je uključena u procese u vezi sa citoskeletalnom mrežom (40), a zatim su brojni oni koji deluju kao molekularni šaperoni (32). Veliku brojnost imaju i proteini koji su uključeni ili u metabolizam aminokiselina i proteina uopšte (18) ili u translacione procese (15). Ostatak proteina obavlja funkcije u vezi sa metabolizmom ugljenih hidrata i lipida, vezivanje nukleinskih kiselina, kao i u oksidoredukcionim procesima koji se odvijaju u elektron-transportnom lancu (Slike 16 i 17). Proteini koji imaju funkcije vezane za formiranje kutikule insekata su bili pronađeni samo u onim uzorcima koji su bili tretirani, a udeo proteina sa citoskeletalnim funkcijama ili onih koji učestvuju u vezivanju nukleinskih kiselina je povećan upravo nakon termalne obrade uzoraka (Slika 16, Heated). Ovi rezultati ponovo ukazuju na značaj takve pripreme uzoraka pre identifikacije i funkcionalne analize strukturno-neuređenih proteina.

Proteini sa strukturnom neuređenošću poseduju određene jedinstvene fizičko-hemijske osobine po kojima se značajno razlikuju od uređenih, globularnih proteina. Usled posedovanja tih osobina, IDP-ovi su posebno otporni na negativne uslove životne sredine i mogu da zadrže svoju aktivnost i funkcionalnost ukoliko budu izloženi takvim uslovima. Iz toga razloga, velika je verovatnoća da ovi proteini igraju važne uloge u adaptacijama organizama na niske temperature koje im omogućavaju da prežive i kada temperatura padne daleko ispod tačke mržnjena vode. Za ispitivanje ovog aspekta IDP-ova odabrana je vrsta kukuruznog plamenca *O. nubilalis* koja se već decenijama koristi u istraživanjima molekularnih osnova adaptacija koje omogućavaju organizmima da prežive hipometaboličke uslove života na niskom temperaturama. Pošto adaptacije na takve uslove mogu dovesti do promene sastava proteoma organizma, prvo je provereno da li usled aklimatizacije dijapauzirajućih gusenica na niske temperature dovodi izmena u količini IDP-ova koji su mogli biti identifikovani. I kod Pilota i kod Glavne postavke, aklimatizacija je dovela do smanjenja ukupnog broja proteina koji su mogli biti detektovani i identifikovani (Slika 6), što se i odrazilo na broj PDP-ova i MDP-ova u tim eksperimentalnim grupama (Slika 7). Međutim, ako se sagleda udeo pojedinačnih kategorija proteina u ukupnom broju, vidi se da je odnos svih kategorija proteina relativno konstantan između eksperimentalnih grupa (Slika 9). Drugim rečima, nije došlo do promene u količini strukturno-neuređenih proteina usled stresa izazvanog niskim temperaturama. Ono što se jeste promenilo je udeo proteina koji sadrže dugačke IDR-ove, koji ujedno i mogu imati izraženu biološku značajnost. Kada se pogleda raspodela proteina sa dugačkim IDR-ovima u različitim eksperimentalnim grupama, onda uticaj dijapauze i niskih temperatura na sadržaj IDP-ova postaje izražen (Slike 12 i 13). Ako se izuzmu uređeni proteini (OP-ovi), udeo proteina sa strukturno-neuređenim regionima je povišen kod dijapauzirajućih gusenica koje su izlagane niskim temperaturama. Ovakvi rezultati su u skladu sa ranijim istraživanjima koja su pokazala da izražena strukturna neuređenost proteinima omogućava da budu fleksibilniji, čime mogu da održe svoju aktivnost i funkcionalnost na niskim temperaturama (Siddiqui i sar., 2006). Takođe, u istraživanjima rađenim na bakterijama koje nastanjuju ekstremna staništa pokazano je da tip staništa može uticati na količinu strukturno-neuređenih proteina u organizmu. Posebno izražen sadržaj IDP-ova su imale bakterije koja nastanjuju slane sredine (*Halobacterium sp. NRC-1* i *Haloarcula marismortui ATCC 43049*), dok je povišen udeo proteina sa dugačkim IDR-ovima zabeležen kod psihrofilne bakterije *Colwellia pszychrerythraea* u odnosu na hipertermofilnu *Pyrococcus horikoshii* (Vicedo i sar., 2015). Među

eukariotima, izlaganje biljke *Arabidopsis thaliana* povišenim temperaturama dovodi do ekspresije gena za proteine koji imaju povišen sadržaj naelektrisanih aminokiselina, a smanjen udeo hidrofobnih aminokiselina, a takođe su bili obogaćeni u pogledu dugačkim IDR-ova (Alvarez-Ponce, 2018). Kod arktičke ribe *Pachychara brachycephalum*, koja je prilagođena životu na konstantnih 0°C, pokazano je da dolazi do specifičnih izmena u aminokiselinskom sastavu proteina u odnosu na njoj srodnu vrstu *Zoarces viviparus* koja živi u toplijim vodama. Zabeležene promene u aminokiselinskom sastavu se povezuju sa povećanom fleksibilnošću proteina baš kod vrste koja nastanjuje konstantno hladne vode – *Pachychara brachycephalum* (Brodte i sar., 2006; Pörtner i Knust, 2007; Windisch i sar., 2012). Svi ovi rezultati, kao i rezultati iz ove disertacije, ukazuju na to stresni uslovi životne sredine mogu dovesti do promena na nivou strukturne neuređenosti proteina, kao i na to da je pri niskim temperaturama poželjno da proteini budu što fleksibilniji kako bi zadržali svoju funkcionalnost.

Kako bi se ispitalo da li i kako rezultati proteomskih analiza u ovoj disertaciji koreliraju sa promenama na translacionom, analizirana je relativna ekspresija gena koji kodiraju različite proteine u čijoj strukturi je zabeležena strukturna neuređenost, a koji bi mogli biti uključeni u procese adaptacije na niske temperature i život u dijapauzi. Ukupno je analizirano 9 gena – pet za HSP proteine (*engl.* heat shock proteins, HSPs, *hsp90, hsc70, hsp70, hsp20.4* i *hsp20.1*), dva za strukturne proteine (*tropmy2* i *tnt*), jedan za protein uključen u imuni odgovor insekata (*thym*) i jedan protein iz ERM porodice proteina (*engl.* ezrin/radixin/moesin, *moe*). Za sve proteine koji su kodirani odabranim genima je utvrđeno da poseduju minimum 10% strukturne neuređenosti, kao i barem jedan dugački IDR.

Proteini toplotnog stresa su molekularni šaperoni koji su zaduženi za održavanje homeostaze tako što obezbeđuju ispravno savijanje novosintetisanih proteina, uklanjanju oštećenih proteina, kao i transportu nascentnih proteina do krajnjeg mesta njihovog sazrevanja (MacRae, 2010; Saibil, 2013; King i MacRae, 2015; Storey i Storey, 2022). U uslovima normometabolizma i odsustva stresa njihova ekspresija je umerena, ali se brzo indukuje nakon izlaganja organizma stresnim faktorima poput niskih i visokih temperatura, UV zračenja, nedostatka kiseonika ili vode i drugih (Hochachka i Somero, 2002; Kregel, 2002; King i MacRae, 2015). Relativna ekspresija svih analiziranih HSP gena, izuzev *hsp20.1*, je imala sličan obrazac – veoma povišena ekspresija u dijapauzirajućim gusenicama izlaganim niskim temperaturama u odnosu na ND kontrolu. Kod

hsp20.1 situacija je bila obrnuta. Ekspresija ovog gena je bila povišena kod ND kontrole u odnosu na ostale grupe. Takođe, ekspresija svih gena je rasla tokom dijapauze i izlaganja niskim temperaturama, izuzev *hsc70* čija je ekspresija bila visoka u svim dijapauzirajućim grupama bez obzira na temperaturu kojoj su gusenice bile izlagane. Ekspresija *hsp90* gena je bila izrazito visoka u svim eksperimentalnim grupama, osim u grupi D(15) gde je bila veoma snižena (Slika 19). Ovi rezultati su u skladu sa prethodnim istraživanjima u kojima je analizirana ekspresija *hsp90* kod insekata koji prolaze kroz fazu zastoja razvoja i/ili bivaju izloženi niskim temperaturama tokom svog životnog ciklusa. Kod dve vrste diptera iz roda *Liriomyza*, izlaganje niskim temperaturama je dovelo do povišene ekspresije *hsp90* (Huang i Kang, 2007; Huang i sar., 2009). Takođe, povišena ekspresija ovog gena zabeležena je i tokom dijapauze na stadijumu lutke vrste lukove muve *Delia antiqua* (Chen i sar., 2005), kao i kod pirinčanog moljca *Chilo suppressalis* (Sonoda i sar., 2006). HSP90 je uključen u mnogobrojne procese vezane za pravilnu sintezu i savijanje proteina (MacRae, 2010; Röhl i sar., 2013), tako da je njegova povišena ekspresija tokom dijapauze i razvijanja otpornosti na hladnoću kod gusenica kukuruznog plamenca sasvim očekivana kako bi se obezbedila strukturna celovitost proteina u nepovoljnim uslovima. Ekspresija *hsc70* i *hsp70* je takođe bila povišena kod dijapauzirajućih gusenica izlaganih niskim temperaturama u odnosu na ND kontrolu (Slike 20 i 21). Takođe, za razliku od *hsp90* i *hsp70*, ekspresija *hsc70* nije bila pod uticajem tretmana hlađenjem. Ovo može ukazivati na to da je ekspresija *hsc70* više pod kontrolom programa dijapauze, nego spoljašnjih uslova. Rezultati dobijeni ovim analizama su u skladu sa prethodnim istraživanjima rađenim na kukuruznom plamencu (Popović, 2014; Popović i sar., 2015). Gledajući istraživanja na drugim vrstama, ekspresija *hsc70* ne pokazuje neki određen obrazac. Tokom dijapauze na stadijumu adulta kod voćne mušice *Drosophila montana* ekspresija *hsc70* je snižena, dok je tokom larvalne dijapauze kod moljca *Sesamia nonagrioides* bila povišena (Gkouvitsas i sar., 2009). Povišena ekspresija *hsc70* zabeležena je u masnom telu nedijapauzirajućih gusenica svilene bube *Bombyx mori* nakon izlaganja niskim temperaturama (Fang i sar., 2021), dok je hlađenje dovelo do smanjenja ekspresije kod nedijapauzirajućih adulta vatrene stenice *Pyrrhocoris apterus* (Koštál i Tollarová-Borovanska, 2009). Ekspresija *hsp70* je doslednija po pitanju odgovora na abiotske stresore, što je u skladu sa inducibilnom prirodom ovog gena. Kod pomenute vatrene stenice izlaganje niskim temperaturama je dovelo do povećanja ekspresije *hsp70* (Koštál i Tollarová-Borovanska, 2009), kao i kod pripadnika roda *Liriomyza* (Huang i Kang., 2007; Huang i sar., 2009). Takođe je

povišena i tokom larvalne dijapauze kod muve *Eurosta solidaginis* (Zhang i sar., 2011) i antarktičke vrste *Belgica antarctica* (Rinehart i sar., 2006, 2007). Veliki proteini toplotnog stresa – HSP90, HSC70 i HSP70 pripadaju grupi ATP-zavisnih HSP-ova, tako da se postavlja pitanje njihove tačne uloge u uslovima hipometabolizma kao što je dijapauza. Kada je ATP dostupan, ovi HSP-ovi aktivno pomažu u pravilnom savijanju aberantnih proteina ili interaguju sa različitim receptorima i kinazama kako bi regulisali ćelijski signaling (King i MacRae, 2015). Međutim, u uslovima kada je količina ATP-a ograničena, kao što je slučaj sa hipometaboličkim stanjem dijapauze (Popović i sar., 2021), proteini toplotnog stresa se verovatno pre nalaze u formi koja favorizuje prosto vezivanje ciljnih proteina i njihovo sklanjanje kako bi ih zaštitili dok se rezerve ATP-a ne obnove i HSP-ovi ponovo budu mogli da pomažu u pravilnom savijanju proteina (King i MacRae, 2015). Mali proteini toplotnog stresa (*engl.* small heat shock proteins, sHSPs) pripadaju grupi molekularnih šaperona veličine između 12 i 43 kDa i u čijoj strukturi se nalaze tipična α-kristalinski domen neophodan za šaperonsku funkciju i β-naborana ploča neophodna za oligomerizaciju ovih proteina (Haslbeck i sar., 2005; Stetler i sar., 2010). Raznovrsni su po broju funkcija koje obavljaju, veličini i brojnosti, a takođe su ATP-nezavisni šaperoni. Predstavljaju prvu liniju odbrane u uslovima stresa i sprečavaju oštećenja proteina vezujući se za njih i predavajući ih nekim od većih članova HSP porodice proteina (Li i sar., 2009; King i MacRae; 2015; Żwirowski i sar., 2017). Analize relativne ekspresije dva mala HSP proteina – *hsp20.4 hsp20.1* u ovoj disertaciji su pokazale da ovi geni imaju suprotstavljene odgovore na uslove dijapauze i izlaganje niskim temperaturama. Ekspresija *hsp20.4* je bila povišena u svim dijapauzirajućim grupama izlaganim niskim temperaturama u odnosu na nedijapauzirajuću kontrolu (Slika 22). Sa druge strane, ekspresija *hsp20.1* je bila povišena u ND grupi u odnosu na sve D grupe (Slika 23). Ovakvi suprotstavljeni obrasci ekspresije mogu ukazivati na podelu uloga ova dva gena u različitim životnim fazama i uslovima sredine. Treba istaći da je samo za protein HSP20.4, čija je ekspresija gena bila povišena tokom dijapauze i izlaganja niskim temperaturama, sa sigurnošću utvrđeno da poseduje jedan dugački IDR. Za protein HSP20.1 je predviđeno da poseduje visok nivo strukturne neuređenosti, ali ne i da poseduje dugačke IDR-ove. Razlog za tu diskrepancu je verovatno da je ovaj protein tik ispod praga od 20 uzastopnih neuređenih aminokiselina za detekciju dugačkih IDR-ova. Kao što je rečeno, mali HSP-ovi su raznovrsni po pitanju brojnosti i molekularnim procesima u koje su uključeni, te se i njihova ekspresija veoma razlikuju među različitim vrstama i ne može se uočiti neki obrazac. Tokom adultne dijapauze kod

komarca *Culex pipiens* ekspresija jednog malog HSP-a je povišena pri kraju dijapauze (Robich i sar., 2007), dok je kod gusenica moljca *S. nonagrioides* ekspresija *hsp20.8* bila promenljiva tokom čitave dijapauze uz blagi porast pri njenom kraju (Gkouvitsas i sar., 2008). Ekspresija malih HSP-ova se razlikuje čak i između jedinki iste vrste, a različitog pola. Kod odraslih moljaca *Spodoptera frugiperda* ekspresija *SfsHsp19.3*, *SfsHsp20*, *SfsHsp21.3* i *SfsHsp29* je bila povišena kod termalno-stresiranih mužjaka u odnosu na ženke, dok je kod ženki to bio slučaj sa *SfsHsp20.1* (Yang i sar., 2021). Ovakva raznolikost u obrascima ekspresije ukazuje na značaj malih HSP-ova u različitim fazama razvoja, kao i adaptacijama na abiotske izazivače stresa.

Transkriptomskim istraživanjima je pokazano da strukturni proteini često imaju promene u ekspresiji kod insekata i drugih zglavkara koji ulaze u dijapauzu i stiču otpornost na hladnoću (Li i sar., 2009; Pavlides i sar., 2011; Teets i sar., 2013; Bryon i sar., 2017). Međutim, značaj ovih promena u ekspresiji gena za strukturne proteine u pomenutim procesima nije u potpunosti razjašnjena. Strukturni proteini bi mogli biti uključeni u sticanje otpornosti na hladnoću, kao i biološke procese tokom dijapauze, na nekoliko nivoa. Mogu regulisati pokretljivost organizma usled promena u sredini i potrebe za očuvanjem energetskih rezervi, obezbeđivati elastičnost membrana pri niskim temperaturama i potpomagati procese u ćelijskom signalingu (Rinehart i sar., 2007; Clark i sar., 2012; Bryon i sar., 2017). U ovoj disertaciji ispitivana je relativna ekspresija dva gena za strukturne proteine, tropomiozin-2 (*engl.* tropomyosin-2, *tropmy2*) i troponin T (*tnt*) koji kontrolišu mišićnu kontrakciju. Tropomiozin-2 reguliše interakciju između aktina i miozina i uključen je u mnoge procese u vezi sa aktinskom citoskeletalnom mrežom (Gunning i sar., 2008). Troponin T je jedna od tri komponente troponinskog kompleksa koji povezuje preostale dve komponente, kalcijum-vezujući troponin C i inhibitorni troponin I. On takođe interaguje sa tropomiozinom i povezuje kako njega tako i ostale članove troponinskog kompleksa sa tankim aktinskim filamentima (Wei i Jin, 2016; Cao i sar., 2019). Tropomiozin blokira miozin-vezujuća mesta na aktinu i sprečava kontrakciju mišića, dok troponinski kompleks uklanja blokadu nastalu tropomiozinom i omogućava aktinu i miozinu da interaguju kako bi se mišić kontrahovao (Gunning i sar., 2008; Pavadai i sar., 2020). Obrasci ekspresije gena za ova dva strukturna proteina su bili u potpunoj suprotnosti jedan u odnosu na drugi (Slike 24 i 25). Kod aktivnih gusenica iz ND grupe *tropmy2* gen je imao veoma sniženu ekspresiju, dok je ekspresija *tnt* gena bila vidno povišena. Kod dijapauzirajućih grupa izlaganim niskim temperaturama situacija je bila obrnuta. Ekspresija *tropmy2* je bila povišena već na početku dijapauze u grupi D(15) i tako se zadržala do

pred kraj dijapauze kada se delimično snizila. Ekspresija *tnt* gena se snižavala tokom toka trajanja dijapauze i aklimatizacije na niske temperature. Ovi rezultati mogu ukazati na molekularne osnove toga kako gusenice prestaju da se kreću tokom dijapauze kako bi štedele energiju. Povišena ekspresija *tropmy2* gena i snižena ekspresija *tnt* gena tokom dijapauze bi doveli do toga da mišićne kontrakcije ne mogu da se odigraju i da se energija ne troši bez potrebe. Povišena ekspresija različitih *tropmy* gena je zabeležena i kod drugih vrsta koja ulaze u dijapauzu i stiču otpornost na hladnoću. Tako je slučaj kod lutaka lukove muve *D. antiqua* tokom letnje dijapauze (Hao i sar., 2012), kao i kod dijapauzirajućih odraslih ženki grinje *Tetranychus urticae* (Bryon i sar., 2017; Zhao i sar., 2017). Slično ja zabeleženo i kod prezimljujućih odraslih komaraca *Culex pipiens pallens* (Zhang i sar., 2019). Na osnovu pregleda dostupne literature, čini se da ekspresija *tnt* gena do sada nije ispitivana u kontekstu dijapauze i sticanja otpornosti na hladnoću. Međutim, promene u nivou *tnt* iRNK su dovedene u vezu sa nutritivnim statusom jedinke. Kod moljca duvana *Manduca sexta* poboljšana ishrana gusenica dovodi do boljih parametara letenja kasnijih odraslih jedinki, što je korelisano sa alternativnim iskrajanjem troponina T (Portman i sar., 2014). Sa druge strane, loša ishrana gusenica moljca *S. frugiperda* dovodi do povišene ekspresije onih izoformi troponina T koji su povezani sa smanjenom mišićnom funkcijom, dok je ekspresija onih troponin T izoformi koji su povezani sa pojačanom mišićnom funkcijom, kao što je *tnt* iz ove disertacije, snižena (Marden i sar., 2018; Portman i sar., 2020). S obzirom na to da se gusenice kukuruznog plamenca ne hrane tokom dijapauze, ovi rezultati mogu ukazivati na molekularne osnove prestanka kretanja jedinki tokom faze mirovanja.

Timozin beta (*engl.* thymosin beta) pripada porodici multifunkcionalnih polipeptida koji su prvi put izolovani iz grudne žlezde teleta (Goldstein i sar., 1966). Kod beskičmenjaka značajna je uloga timozina u humoralnom odgovoru na mikrobijalne infekcije (Schillaci i sar., 2012; Feng i sar., 2018; Liao i sar., 2018). Relativna ekspresija *thym* gena je bila povišena u nedijapauzirajućim gusenicama koje su bile u aktivnom razvoju, u odnosu na sve dijapauzirajuće grupe (Slika 26). Na osnovu ovih zapažanja, pretpostavlja se da timozin beta ima važniju ulogu u imunskom odgovoru kukuruznog plamenca u aktivnoj fazi životnog ciklusa ove vrste tokom koje se gusenice hrane, u odnosu na dijapauzu. Njegov značaj u fazi aktivnog razvoja se ogleda u tome što je količina timozina beta na proteinskom nivou bila povišena u lutkama svilene bube (Ma i sar., 2015), kao i u tome da je ekspresija gena za timozin betu bila povišena nakon izlaganja larvi kukuruzne sovice

Helicoverpa armigera hormonu 20-hidroksiekdizonu koji kontroliše presvlačenje i metamorfozu kod insekata (Zhang i sar., 2011).

Moezin (*engl.* moesin) pripada ERM porodici tri veoma srodna i izuzetno očuvana proteina (*engl.* ezrin/radixin/moesin). ERM proteini su citoskeletalni elementi koji posreduju u regulaciji brojnih signalnih puteva tako što potpomažu sklapanje višeproteinskih kompleksa na tačkama susretanja ćelijskih membrana i citoskeleta (Fehon i sar., 2010). Većina beskičmenjaka poseduje gen za samo jedan ERM protein, kao što je recimo gen za moezin kod voćne mušice *D. melanogaster* (Shabardina i sar., 2019; Bajusz i sar., 2021). Analiza relativne ekspresije *moe* gena u ovoj disertaciji je pokazala da je ona povišena kako kod nedijapauzirajućih gusenica, tako i kod dijapauzirajuće grupe D(–16). Ovo ukazuje na značaj *moe* gena kako u fazi aktivnog razvoja tako i u procesima sticanja otpornosti na hladnoću u metaboličkim uslovima. Kada se sagleda ekspresija ovog gena u dijapauzirajućim grupama, najniža je bila na početku dijapauze u grupi D(15) i postepeno je rasla sa opadanjem temperature i ulaskom gusenica dublje u dijapauzu (Slika 28), što pokazuje da je *moe* gen potencijalno uključen i u sam proces adaptacije na niske temperature. Kod insekata je utvrđeno da je moezin bitna komponenta ćelijskog jedra jer učestvuje u transportu iRNK (Kristó i sar., 2017) i regulaciji stotine različitih gena, među kojima su i geni za HSP-ove. U istraživanju na *D. melanogaster* je pokazano da narušavanje funkcije moezina u značajnoj meri smanjuje ekspresiju *hsp* gena (Bajusz i sar., 2021). Rezultati ekspresije *moe* gena na gusenicama kukuruznog plamenca su u skladu sa ovom utvrđenom povezanošću moezina i ekspresije *hsp* gena, što se može i uočiti praćenjem obrazaca ekspresije nekoliko *hsp* gena u dijapauzirajućim i hlađenim grupama (Slike 19, 21 i 22) i ekspresije *moe* gena (Slika 27) gde se vidi da imaju sličan trend.

Dijapauza, kao stanje mirovanja, je period tokom koga je organizam u stanju hipometabolizma i tokom koga svoje energetske rezerve usmerava kako na održavanje bazalnog metabolizma tako i na sintezu zaštitnih molekula za razvoj otpornosti na niske temperature i druge stresne faktore (oksidativni stress, dehidracija i sl.). Međutim, termodinamički uslovi u kojima se ovi procesi odvijaju su kinetički nepovoljni, jer kinetička energija opada sa temperaturom. Zbog toga, u uslovima niskih temperatura, od ključnog značaja je zaštititi organizam na ćelijskom nivou od negativnih efekata pada kinetičke energije sistema, koji bi se mogli ispoljiti kako u strukturnim aberantnim promenama, tako i u biohemijsko-fiziološkim zastojima procesa. Stoga molekuli koji

zadržavaju svoju fleksibilnost u takvim nepovoljnim uslovima mogu biti jedan od vidova adaptacija organizama na niske temperature na molekulskom nivou. IDP-ovi i proteini sa dugačkim IDR-ovima pripadaju baš takvim biomolekulima, konkretno proteinima, koji osim značajne strukturne imaju i brojne regulatorne i zaštitne funkcije u ćeliji i organizmu. Rezultati dobijeni u ovoj disertaciji pokazali su da, iako udeo strukturno-neuređenih proteina ne raste tokom dijapauze i razvijana otpornosti na hladnoću, već naprotiv ostaje stabilan, ipak raste udeo proteina sa većim brojem strukturno neuređenih regiona. Takođe, na ovaj porast udela proteina koji sadrže strukturno-neuređene regione tokom nepovoljnih uslova života ukazuju i rezultati relativne ekspresije gena u ovoj disertaciji, gde je ekspresija većine gena koji kodiraju proteine sa određenim stepenom strukturne neuređenosti bila povišena usled ulaska i ostajanja u hipometaboličkom stanju dijapauze, kao i izlaganja niskim temperaturama. Na ovaj način, organizam kompenzuje ograničene energetske resurse za opštu sintezu *de novo* proteina, štedeći energiju, a ipak dobijajući deo proteoma koji zadržava visoku fleksibilnost u nepovoljnim uslovima i time omogućuje kontinuum strukturne i funkcionalne homeostaze proteina i svih procesa u kojima su oni uključeni (Slika 28).

Zaključci: Na osnovu postavljene hipoteze, definisanih ciljeva i rezultata dobijenih analizama na larvama 5. stupnja *O. nubilalis* koje su podvrgavane različitim eksperimentalnim uslovima za razvoj otpornosti na niske temperature mogu se izvesti sledeći zaključci: 1) uspešno su uspostavljene i optimizovane metode za izolaciju IDP-ova i proteina sa dugačkim IDR-ovima kod *O. nubilalis*, a metodom termalne obrade grubog homogenata uspešno je obogaćena frakcija ovih proteina, što je potvrđeno specifičnom 2D PAGE elektroforezom i LC-MS/MS metodom; 2) bioinformatičkom analizom je potvrđeno postojanje brojnih proteina čije prisustvo je bilo samo predviđeno na osnovu nukleotidnih sekvenci u postojećim bazama podataka. 3) bioinformatička analiza je pokazala postojanje IDP-ova i proteina sa dugačkim IDR-ovima u svim eksperimentalnim grupama, kao i da metoda termalne pripreme uzorka pre LC-MS/MS u velikoj meri obogaćuje sadržaj ovih proteina; 4) postoji značajna grupa termostabilnih uređenih proteina, budući da je nakon metode termalne pripreme uzorka otkriven i veliki broj proteina sa visokim stepenom strukturne uređenosti; 5) bioinformatičkom analizom su identifikovani i potvrđeni proteini sa različitim stepenom strukturne neuređenosti; 6) sadržaj specifičnih aminokiselina koje pogoduju strukturnoj neuređenosti je izraženiji kod proteina sa višim stepenom strukturne neuređenosti u sekvenci, što je u saglasnosti sa dosadašnjom literaturom; 7) udeo IDP-ova ne raste

tokom postepenog razvoja otpornosti na hladnoću, čime je pretpostavljena hipoteza delimično opovrgnuta; 8) udeo proteina koji sadrže dugačke IDR-ove raste tokom razvoja otpornosti na hladnoću, čime je pretpostavljena hipoteza delimično potvrđena; 9) bioinformatička analiza je pokazala da su IDP-ovi i proteini sa dugačkim IDR-ovima uključeni u brojne biološke procese, jer imaju veliki broj molekularnih funkcija od kojih prednjače citoskeletni proteini, molekularni šaperoni, proteini uključeni u translaciju i metabolizam proteina, regulatorni proteini koji vezuju nukleinske kiseline i drugi; 10) deo proteina nije mogao biti identifikovan, a kod mnogih nisu opisane funkcije jer ne postoje podaci u dostupnim bazama podataka; 11) ekspresija većine odabranih gena koji kodiraju IDP-ove i proteine sa dugačkim IDR-ovima je povišena u uslovima dijapauze i razvoja otpornosti na hladnoću; 12) IDP-ovi i proteini sa dugačkim IDR-ovima učestvuju u održavanju strukturne i funkcionalne homeostaze organizma tokom hipometaboličke faze mirovanja (dijapauza) i doprinose sveukupnom potencijalu stečene otpornosti na hladnoću.

Prema mojim saznanjima ovo je prvo istraživanje koje se bavi IDP-ovima i proteinima sa dugačkim IDR-ovima u *in vivo* model-sistemu kod kojeg se indukuje razvijanje otpornosti na hladnoću tokom faze mirovanja. Dobijena saznanja obogaćuju razumevanje ekofiziološke osnove povezanosti dijapauze sa otpornošću na hladnoću, kao i molekularnih mehanizama koji omogućavaju kako razvoj tako i održavanje otpornosti na hladnoću tokom dugotrajne faze mirovanja. Takođe, nedvosmisleno je pokazano da postoje i termostabilni uređeni proteini kojima se treba posvetiti u daljim istraživanjima. Ovi rezultati mogu imati potencijalnu primenu u različitim oblastima – od krioprezervacije u medicini i veterini, do prehrambene i farmaceutske industrije.

Author biography

Miloš Avramov was born on October 21st 1991 in Novi Sad. He attained his primary education from „Đura Daničić“ elementary school as a recipient of the Vukova diploma distinguishment, and secondary education from the Karlovci Grammar School in Sremski Karlovci. In 2010 he began his Bachelor academic studies in biology at the Faculty of Sciences in Novi Sad, Department of Biology and Ecology and attained his Bachelor's degree in 2014 with a 9,41 GPA. During this time he was honored with the University and Faculty awards for GPA, as well as the Scholarship for exceptionally gifted students by the Ministry of Science, Education and Technological Development of the Republic of Serbia. Immediately after attaining his B.Sc. he enrolled in Master academic studies of molecular biology and joined the Laboratory for biochemistry and molecular biology, with **Prof. Jelena Purać, Ph.D.** as his advisor. Under her supervision, Miloš wrote and defended his Master's thesis entitled „Effect of resveratrol and proanthocyanidins on NF-kB gene expression in human lung fibroblast cell line“ in 2015, completing the Master level studies with a GPA of 10,00. He enrolled in the Doctoral academic studies of biology right after attaining his M.Sc., with **Prof. Željko D. Popović** as his advisor. At the start of this level of studies he was admitted as a Research Trainee and received the Doctoral scholarship from the Ministry of Science, Education and Technological Development. In 2017 he applied for and was awarded the International Visegrad Fund Scholarship Grant, allowing him to spend one year as a guest researcher at the Research Centre for Natural Sciences in Budapest, Hungary, under the supervision of **Senior Research Fellow Ágnes Tantos, Ph.D.**, where he was able to complete a significant portion of his planned dissertation experiments. Upon return from Budapest he is employed full-time as a Research Assistant at the Department of Biology and Ecology of the Faculty of Sciences in Novi Sad. During his doctoral studies, Miloš was extensively involved in both teaching and research activities. He has taught practical courses for „Biochemistry“ and „Molecular biology of eukaryotes“ for students enrolled in the Bachelor level studies of biology, ecology and professor of biology, as

well as for „Medical biochemistry“ for Master level studies students of molecular biology. Miloš has assisted more than a dozen Master students in carrying out experiments for their final theses, as well as with the preparation and writing of those manuscripts. He has co-authored three papers that have been published in distinguished journals (M21 category), of which one as the first author, seven papers presented at international conferences (M34 category) and twelve papers presented at national conferences (M63 and M64 categories). Miloš was also involved in preparing proposals for several provincial, national and international project grants, one of which was selected for funding by the Provincial Secretariat for Higher Education and Scientific Research of the Autonomous Province of Vojvodina in 2022 – „Pest insect species as alternative sources of proteins with antiviral and antimicrobial functions (AlterNATIVE)“.

План третмана података

Назив пројекта/истраживања
„Intrinsically disordered proteins of the European corn borer *Ostrinia nubilalis* (Hbn, 1796)“ „Структурно-неуређени протеини кукурузног пламенца *Ostrinia nubilalis* (Hbn, 1796)“
Назив институције/институција у оквиру којих се спроводи истраживање
Природно-математички факултет, Универзитет у Новом Саду Истраживачки центар за природне науке, Истраживачка мрежа „Етвеш Лоранд“, Будимпешта, Мађарска
Назив програма у оквиру ког се реализује истраживање
„Молекуларни механизми редокс сигналинга у хомеостази, адаптацији и патологији“, ОИ 173014 “Intrinsically disordered proteins (IDPs) in adaptation to low temperatures”, ев. број #51700031, Међународни вишеградски фонд Програм научноистраживачког рада Министарства науке, технолошког развоја и иновација републике Србије, ев. број 451-03-47/2023-01/200125
1. Опис података
1.1 Врста студије Студија користи *in vivo* (увођење експерименталних јединки у стање хипометаболизма и стицање отпорности на хладноћу), *in vitro* (раздвајање протеина на полиакриламидном гелу, раздвајање нуклеинских киселина на агарозном гелу, идентификација протеина помоћу течно-масене спектрометрије и анализа релативне експресије гена) и *in silico* (предвиђање структурне неуређености протеина и њихова карактеризација) технике како би се у испитиваној врсти утврдило присуство и заступљеност протеина са различитим степеном структурне неуређености и окарактерисала њихова улога у процесима адаптирања на ниске температуре и преживљавање дугог периода хипометаболизма. 1.2 Врсте података: **квалитативни** и **квантитативни** 1.3. Начин прикупљања података секвенце протеина и нуклеотида доступне у НЦБИ бази података визуелизација протеина након полиакриламидне гел електрофорезе записи са спектрофотометарских мерења Excel фајлови са подацима након *in silico* анализа

записи након мерења промене релативне експресије гена на одговарајућем уређају

преглед доступне литературе

1.3.1 Употребљени софтвер и формат датотеке:

Excel фајл, датотека .xlsx

PDF фајл, датотека .pdf

Текст фајл, датотека .docx

JPG фајл, датотека .jpg, .png

1.3.2. Број записа (код квантитативних података)

број варијабли: од 150 до 2000, у зависности од методе

број мерења: од 50 до 150, у зависности од методе

1.3.3. Поновљена мерења

не

Да ли формати и софтвер омогућавају дељење и дугорочну валидност података?

да

2. Прикупљање података

2.1 Методологија за прикупљање/генерисање података

2.1.1. У оквиру ког истраживачког нацрта су подаци прикупљени?

експерименти, спектрофотометријско одређивање концентрације протеина и нуклеинских киселина, идентификација протеина течно-масеном хроматографијом и спектрометријом, раздвајање протеина на полиакриламидном гелу

остало, *in silico* предвиђања структурне неуређености протеина и њихова квалитативна и функционална карактеризација

2.1.2 Навести врсте мерних инструмената или стандарде података специфичних за одређену научну дисциплину (ако постоје).

спектрофотометар, апаратура за електрофорезу на полиакриламидном и агарозном гелу, уређај за течно-масену хроматографију и спектрометрију, уређај за мерење релативне експресије гена (qPCR), комора за визуелизацију нуклеинских киселина на агарозном гелу, програмски пакети за *in silico* обраду резултата

2.2 Квалитет података и стандарди

2.2.1. Третман недостајућих података

Да ли матрица садржи недостајуће податке? Не.

2.2.2. На који начин је контролисан квалитет података? Описати

статистичким анализама и коришћењем контролних експеримената

2.2.3. На који начин је извршена контрола уноса података у матрицу?

контролом са другог корисничког налога

3. Третман података и пратећа документација

3.1. Третман и чување података

3.1.1. Подаци ће бити депоновани у Google Drive репозиторијум.

3.1.2. URL *drive.google.com/drive/folders/1nkctTsyBBGevO88MfCxX7QYegWopFMoS?*

3.1.3. DOI није додељен

3.1.4. Да ли ће подаци бити у отвореном приступу? Да.

3.2 Метаподаци и документација података

нема метаподатака

3.3 Стратегија и стандарди за чување података

3.3.1. До ког периода ће подаци бити чувани у репозиторијуму? ∞

3.3.2. Да ли ће подаци бити депоновани под шифром? Да.

3.3.3. Да ли ће шифра бити доступна одређеном кругу истраживача? Да.

3.3.4. Да ли се подаци морају уклонити из отвореног приступа после извесног времена? Не.

4. Безбедност података и заштита поверљивих информација

4.1 Формални стандарди за сигурност информација/података

4.1.2. Да ли је истраживање одобрено од стране етичке комисије? Не.

4.1.2. Да ли подаци укључују личне податке учесника у истраживању? Не.

5. Доступност података

5.1. Подаци ће бити јавно доступни

5.4. Навести лиценцу под којом ће прикупљени подаци бити архивирани.

ауторство – некомерцијално – делити под истим условима

6. Улоге и одговорност

6.1. Навести име и презиме и мејл адресу власника (аутора) података

Милош Аврамов, milos.avramov@dbe.uns.ac.rs

6.2. Навести име и презиме и мејл адресу особе која одржава матрицу с подацима

Милош Аврамов, milos.avramov@dbe.uns.ac.rs

6.3. Навести име и презиме и мејл адресу особе која омогућује приступ подацима другим истраживачима

Милош Аврамов, milos.avramov@dbe.uns.ac.rs