

# Structural characterization of Heliobacter plyori proteins required for survival of the bacterium

---

Kekez, Ivana

Doctoral thesis / Disertacija

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:217:633407>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-12-31**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)





University of Zagreb  
FACULTY OF SCIENCE

Ivana Kekez

**STRUCTURAL CHARACTERIZATION OF  
*Helicobacter pylori* PROTEINS REQUIRED FOR  
SURVIVAL OF THE BACTERIUM**

DOCTORAL THESIS

Zagreb, 2016

---





Sveučilište u Zagrebu  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET

Ivana Kekez

**STRUKTURNA KARAKTERIZACIJA PROTEINA  
BAKTERIJE *Helicobacter pylori* VAŽNIH ZA  
NJEZINO PREŽIVLJAVANJE**

DOKTORSKI RAD

Zagreb, 2016

---





University of Zagreb  
FACULTY OF SCIENCE

Ivana Kekez

**STRUCTURAL CHARACTERIZATION OF  
*Helicobacter pylori* PROTEINS REQUIRED FOR  
SURVIVAL OF THE BACTERIUM**

DOCTORAL THESIS

Supervisors:

Dr. sc. Dubravka Matković-Čalogović, Professor

Dr. sc. Giuseppe Zanotti, Professor

Zagreb, 2016

---





Sveučilište u Zagrebu  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET

Ivana Kekez

**STRUKTURNA KARAKTERIZACIJA PROTEINA  
BAKTERIJE *Helicobacter pylori* VAŽNIH ZA  
NJEZINO PREŽIVLJAVANJE**

DOKTORSKI RAD

Mentori:  
prof. dr. sc. Dubravka Matković-Čalogović  
prof. dr. sc. Giuseppe Zanotti

Zagreb, 2016

---





---

## Acknowledgments

I would like to express special gratitude to my supervisor Prof. Dr. Dubravka Matković-Čalogović for her continuous support, encouragement and guidance throughout my PhD study. That being said, I would like to single out the enjoyable and memorable time we had spent together at synchrotron as the time and place where her patient explanations regarding crystallographic issues and her open-door policy allowed me to quickly discuss any results or problems.

I am sincerely grateful to Prof. Dr. Giuseppe Zanotti for accepting to be my second supervisor and for giving me the opportunity to carry out most of the experiments in his laboratory at the Department of Biomedical Sciences, University of Padua. The project described in this thesis is a continuation of his studies on the structural characterization of the *Helicobacter pylori* proteins involved in the pathogenicity of this bacterium. Fast and easy communication, along with his expertise in this field of research, made him an excellent tutor. Besides for Prof. Dr. Zanotti, I would also like to thank all his past and present laboratory staff for all the kindness and hospitality which I had felt during my stay in Padua.

Special thanks go to Dr. Laura Cendron for introducing me to the „world“ of protein crystallography and for guiding me during my first stay in Padua. I appreciate her unselfish will to share with me numerous helpful advice and practical experiences. Laura, thank you for encouraging me never to give up on protein crystals and to always look for other possibilities worth trying. I enjoyed our scientific discussions, all the nice times we spent together. All in all, thank you for being Laura!

I would also like to thank all the girls from Prof. G. Zanotti's group, especially Francesca, Paola, Valentina and Elena, for a friendly and relaxed atmosphere and for always being kind and willing to help – it was a pleasure to work with all of you!

My acknowledgments also go to members of the evaluation committee: Dr. Jasmina Rokov-Plavec and Dr. Marija Luić whose comments and suggestions greatly improved the thesis.

I am also grateful to Prof. Dr. Patrizia Polverino de Laureto (University of Padua, Department of Pharmaceutical Sciences) for giving me the opportunity to use the CD spectrometer and for helping me with the mass measurements and analysis. Further gratitude is also extended to Prof. Dr. Massimo Bellanda (University of Padua, Department of Chemistry) for providing access to the NMR spectrometer.

---

---

My acknowledgments also go to many people working at the Department of Chemistry, University of Zagreb. Here I would like to extend particular gratitude to the Division of Biochemistry staff for allowing me to use their instruments and equipment and for giving me a chance to continue my experimental work here in Zagreb. Also deserving of special acknowledgments are my colleagues from the Division of General and Inorganic Chemistry and Zoran Bojanić who provided me with IT support.

Irena, Željka and Biserka, thank you ever so much for giving me tremendous professional support, for all the nice times we spent together, as well as for visiting Padua!

Warm thanks go to Dalibor who always believed in my work and motivated me.

Many thanks go to my friends Ines, Martina, Ivona and Vlatka for all the nice days and pleasant evenings together.

A loving thank you goes to my husband Mario who always believed in me and gave me advice and encouragement.

Last but not least, I would like to thank my parents, Marija and Joso, and my sister Josipa, for invaluable support and encouragement during my life. I want to thank you for all the small things and all the care - I am just happy to have you all! This thesis I dedicate to you.

Ivana Kekez

---

## Contents

SAŽETAK.....	XV
ABSTRACT .....	XVII
PROŠIRENI SAŽETAK.....	XIX
§ 1. INTRODUCTION .....	1
§ 2. LITERATURE REVIEW .....	3
2.1. <i>Helicobacter pylori</i> – an interesting medical target.....	3
2.2. <i>H. pylori</i> virulence factors .....	4
2.3. <i>H. pylori</i> flagellum.....	10
2.3.1. <i>The hook</i> .....	14
2.3.1. <i>FlgD</i> .....	15
2.4. Detoxification and metal ion homeostasis in <i>H. pylori</i> .....	16
2.4.1. <i>Copper ion homeostasis</i> .....	17
2.4.2. <i>CrdA</i> .....	17
2.5. Heat shock proteins in <i>H. pylori</i> .....	19
2.5.1. <i>HP1026</i> .....	20
§ 3. MATERIALS AND METHODS .....	25
3.1. Materials .....	25
3.1.1. <i>Equipment</i> .....	25
3.1.2. <i>Chromatography supplies</i> .....	26
3.1.3. <i>Crystallographic supplies</i> .....	26
3.1.4. <i>Kits</i> .....	27
3.1.5. <i>Filters and membranes</i> .....	27
3.1.6. <i>Chemicals and buffers</i> .....	27
3.1.7. <i>Enzymes</i> .....	28
3.1.8. <i>Antibodies</i> .....	29
3.1.9. <i>Oligonucleotides</i> .....	29
3.1.10. <i>Vectors</i> .....	30
3.1.11. <i>Bacterial strains: Escherichia coli</i> .....	32
3.1.11.1. <i>Host cell for cloning</i> .....	32
3.1.11.2. <i>Strains for protein expression</i> .....	32

---

<b>3.2. Microbiological methods .....</b>	<b>33</b>
3.2.1. <i>Media for cultivation of E. coli cells.....</i>	33
3.2.2. <i>Transformation of chemically competent E. coli cells.....</i>	33
<b>3.3. Molecular biological methods .....</b>	<b>33</b>
3.3.1. <i>Amplification and isolation of plasmid DNA from E. coli.....</i>	33
3.3.2. <i>Determination of DNA concentration.....</i>	34
3.3.3. <i>Polymerase chain reaction (PCR).....</i>	34
3.3.4. <i>Agarose gel electrophoresis.....</i>	35
3.3.5. <i>Cloning and sequencing.....</i>	36
3.3.5.1. <i>Standard cloning: restriction and ligation of DNA.....</i>	36
3.3.5.2. <i>Enzyme free-cloning.....</i>	37
<b>3.4. Protein biochemical methods .....</b>	<b>38</b>
3.4.1. <i>Protein quantification .....</i>	38
3.4.2. <i>Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).....</i>	38
3.4.3. <i>Western blotting .....</i>	40
3.4.4. <i>Proteolytic cleavage.....</i>	41
3.4.4.1. <i>Removal of the hexa His-tag from HP1026 with thrombin.....</i>	41
3.4.4.2. <i>Removal of the GST-tag from GST_HP1326 with PreScission protease.....</i>	42
3.4.4.3. <i>Limited proteolysis of the FlgD (HP0907) protein .....</i>	42
<b>3.5. Expression of recombinant proteins in E. coli .....</b>	<b>43</b>
3.5.1. <i>Expression of CrdA (HP1236), HP1026 and FlgD (HP0907).....</i>	43
3.5.2. <i>Expression of selenomethionine- HP1026 and FlgD (HP0907).....</i>	44
<b>3.6. Purification of CrdA, HP1026 and FlgD .....</b>	<b>45</b>
3.6.1. <i>Affinity chromatography .....</i>	45
3.6.1.1. <i>Affinity chromatography using a Strep-Tactin.....</i>	45
3.6.1.2. <i>Affinity chromatography with Ni-NTA agarose .....</i>	46
3.6.1.3. <i>Affinity chromatography with glutathione sepharose .....</i>	47
3.6.2. <i>Size-exclusion chromatography (SEC).....</i>	48
3.6.3. <i>Ultrafiltration.....</i>	49
<b>3.7. Refolding of the CrdA protein solubilized from inclusion bodies.....</b>	<b>50</b>
<b>3.8. Protein characterization.....</b>	<b>51</b>
3.8.1. <i>Circular dichroism.....</i>	51
3.8.2. <i>One-dimensional NMR spectroscopy (1D NMR).....</i>	53
3.8.3. <i>Light scattering .....</i>	53

---

3.8.3.1. <i>Dynamic light scattering (DLS)</i> .....	53
3.8.3.2. <i>Static light scattering (SLS)</i> .....	54
3.8.4. <i>Thermofluor assay</i> .....	55
3.8.5. <i>Mass spectrometry</i> .....	58
3.8.6. <i>ATPase activity assay by reversed-phase high-performance liquid chromatography (RP-HPLC)</i> .....	59
3.8.7. <i>Ultraviolet-visible (UV-VIS) titration</i> .....	60
3.8.8. <i>Crystallographic studies</i> .....	60
3.8.8.1. <i>Crystallization</i> .....	60
3.8.8.2. <i>Data processing and structure determination</i> .....	62
3.8.9. <i>Bioinformatics methods and software tools</i> .....	62
<b>§ 4. RESULTS AND DISCUSSION</b> .....	<b>65</b>
<b>4.1. FlgD from <i>H. pylori</i> (HpFlgD)</b> .....	<b>65</b>
4.1.1. <i>Cloning, expression and purification</i> .....	65
4.1.2. <i>Crystallization and data collection</i> .....	67
4.1.3. <i>Size of HpFlgD in solution</i> .....	69
4.1.4. <i>Size of HpFlgD in the crystal</i> .....	71
4.1.5. <i>Structure of the FlgD monomer</i> .....	73
4.1.6. <i>Quaternary structure and interfaces</i> .....	76
4.1.7. <i>Comparison with other structures</i> .....	81
4.1.8. <i>Proposed physiological role of truncated HpFlgD</i> .....	84
<b>4.2. CrdA from <i>H. pylori</i></b> .....	<b>86</b>
4.2.1. <i>Cloning, expression and purification of StrepCrdA, CrdAHis and GSTCrdA</i> .....	86
4.2.2. <i>Size of CrdA in solution</i> .....	89
4.2.3. <i>Affinity of CrdA towards copper(II) ions</i> .....	91
4.2.4. <i>Crystallization of CrdA</i> .....	92
<b>4.3. HP1026 from <i>H. pylori</i></b> .....	<b>93</b>
4.3.1. <i>Expression, purification and size of HP1026 in solution</i> .....	93
4.3.2. <i>Activity of HP1026</i> .....	94
4.3.3. <i>Stability of HP1026</i> .....	96
4.3.4. <i>Crystallization of HP1026</i> .....	98
<b>§ 5. CONCLUSIONS</b> .....	<b>103</b>
<b>§ 6. LIST OF ABBREVIATIONS</b> .....	<b>105</b>
<b>§ 7. REFERENCES</b> .....	<b>107</b>
<b>§ 8. CURRICULUM VITAE</b> .....	<b>XXVIII</b>





Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
**Kemijski odsjek**

Doktorska disertacija

## SAŽETAK

### STRUKTURNA KARAKTERIZACIJA PROTEINA BAKTERIJE *Helicobacter pylori* VAŽNIH ZA NJEZINO PREŽIVLJAVANJE

Ivana Kekez

Horvatovac 102a, Zagreb, Hrvatska  
Via Ugo Bassi 58/B, Padova, Italija

U sklopu ove disertacije strukturno su okarakterizirani proteini iz bakterije *H. pylori* koji su važni za njezino preživljavanje (*HpFlgD*, *CrDA*, *HP1026*). Riješena kristalna struktura skraćenog proteina *HpFlgD* ukazala je na drugačiju međusobnu orijentaciju dviju domena nego što je nađeno u ostalim homolozima proteina *FlgD*. Također je pokazano da je skraćeni *HpFlgD* i u otopini i kristalu prisutan kao tetramer što ukazuje na značajnu razliku u molekularnoj organizaciji biča u različitim bakterijskih vrsta. Uočeno je da dodatak iona  $\text{Cu}^{2+}$  u otopinu *CrDA*, proteina koji pretpostavljeno veže bakrove ione, potiče stvaranje monomernih vrsta u otopini te da *CrDA* veže ione  $\text{Cu}^{2+}$  sa slabim afinitetom što je karakteristika proteina koji prenose bakrove ione. Funkcionalna karakterizacija proteina *HP1026* je po prvi put demonstrirala da *HP1026* obavlja ATPaznu aktivnost. Iako se proteini koji spadaju u skupinu  $\text{AAA}^+$  proteina najčešće udružuju kao heksameri, za *HP1026* je pokazano da u otopini stvara dimerne vrste.

(115 + xxxii stranica, 49 slika, 39 tablica, 147 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj kemijskoj knjižnici, Horvatovac 102a, Zagreb i Nacionalnoj i sveučilišnoj knjižnici, Hrvatske bratske zajednice 4, Zagreb.

Ključne riječi: *CrDA* / *FlgD* / *HP1026* / *H. pylori* / strukturna karakterizacija

Mentori: prof. dr. sc. Dubravka Matković-Čalogović  
prof. dr. sc. Giuseppe Zanotti

Rad prihvaćen: 01.06.2016.

Ocjenitelji: doc. dr. sc. Jasmina Rokov-Plavec  
prof. dr. sc. Dubravka Matković-Čalogović  
dr. sc. Marija Luić, zn. savj.







University of Zagreb  
Faculty of Science  
**Department of Chemistry**

Doctoral Thesis

## ABSTRACT

### STRUCTURAL CHARACTERIZATION OF *Helicobacter pylori* PROTEINS REQUIRED FOR SURVIVAL OF THE BACTERIUM

Ivana Kekez  
Horvatovac 102a, Zagreb, Croatia  
Via Ugo Bassi 58/B, Padua, Italy

Within this thesis several proteins from *H. pylori*, important for survival of the bacterium, were structurally characterized (*HpFlgD*, *CrdA*, *HP1026*). Crystal structure of the truncated form of the *HpFlgD* protein revealed that spatial orientation of the two domains differs from that of the homologous *FlgD* family members. This fact together with the observation that truncated *HpFlgD* assembles into tetramers, both in the solution and in the crystal form, strongly suggests that significant differences exist in the molecular organization of the flagella in different bacterial species. It was shown that incubation of the putative copper binding *CrdA* protein with  $\text{Cu}^{2+}$  ions favours formation of monomeric species in solution and that *CrdA* binds  $\text{Cu}^{2+}$  with very low affinity which is a property of copper trafficking proteins. Functional assays of the *HP1026* protein demonstrated for the first time its ATPase activity. While proteins that belong to the class of AAA<sup>+</sup> proteins usually form hexamers, *HP1026* was found to form dimers.

(115 + xxxii pages, 49 figures, 39 tables, 147 references, original in English)

Thesis deposited in Central Chemical Library, Horvatovac 102A, Zagreb, Croatia and National and University Library, Hrvatske bratske zajednice 4, Zagreb, Croatia.

Keywords: *CrdA* / *FlgD* / *HP1026* / *H. pylori* / structural characterization

Supervisors: Dr. Dubravka Matković-Čalogović, Professor  
Dr. Giuseppe Zanotti, Professor

Thesis accepted: 01.06.2016.

Reviewers: Dr. Jasmina Rokov-Plavec, Assistant Professor  
Dr. Dubravka Matković-Čalogović, Professor  
Dr. Marija Luić, Senior Scientist





Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
**Kemijski odsjek**

Doktorska disertacija

## PROŠIRENI SAŽETAK

*Helicobacter pylori* je patogena bakterija nađena na sluznici čovjekova želuca. Danas se zna da je infekcija povezana s različitim ishodima bolesti. Iako su mnoge infekcije neprimjetne i bez ikakvih simptoma bolesti, većina rezultira s različitim stupnjevima gastritisa. U 10 % inficirane populacije zaraza će se razviti u teško želučano oboljenje poput peptičkog ulkusa (čira) te gastritisa. U 1 % zaražene populacije doći će do razvoja adenokarcinoma želuca te limfoma mukoznog tkiva (MALT). Stoga je 1994. Svjetska zdravstvena organizacija proglasila bakteriju *H. pylori* karcinogenom I. reda [R. M. Jr Peek and M. J. Blaser, *Nature Rev. Cancer* **2** (1) (2002) 28–37; S. Suerbaum and P.N. Michetti, *Engl. J. Med.* **347** (2002) 1175–1186]. Genom sedamnaest sojeva bakterije *H. pylori* je u potpunosti sekvenciran, poput 26695, J99, G27 i HPAG1 [J. F. Tomb, O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter, *Nature* **388** (6642) (1997) 539–547; R. A. Alm, L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust, *Nature* **397** (1999) 176–180; J. D. Oh, H. Kling-Backhed, M. Giannakis, J. Xu, R. S. Fulton, L. A. Fulton, H. S. Cordum, C. Wang, G. Elliott, J. Edwards, E. R. Mardis, L. G. Engstrand, and J. I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **103** (2006) 9999–10004; D. A. Baltrus, M. R. Amieva, A. Covacci, T. M. Lowe, D. S. Merrell, K. M. Ottemann, M. Stein, N. R. Salama, and K. Guillemin. *J. Bacteriol.* **191** (1) (2009) 447–448]. Svi genomi imaju sličnu prosječnu duljinu i gustoću kodirajućih sekvenci. Ekstremno brza rekombinacija zajedno sa značajnom razlikom

sekvenci upućuje na panmiktičku strukturu populacije [S. Suerbaum and C. Josenhans, *Nat. Rev. Microbiol.* **5** (2007) 441–452].

Prema patološkoj važnosti bakterija *H. pylori* je u posljednjih dvadeset godina postala važan cilj istraživanja. Napravljene su i analizirane brojne interaktivne mape proteina iz bakterije *H. pylori* [J. C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne, and P. Legrain, *Nature* **409** (2001) 211–215] koji joj omogućavaju preživljavanje u kiselom mediju [L. Cendron and G. Zanotti *Febs J.* **278** (8) (2011) 1223–1231]. Suprotno prethodnom, broj podataka o riješenim strukturama proteina ove bakterije je malen: do ožujka 2016. u *Protein Data Bank* pohranjene su 439 datoteke s koordinatama atoma (<http://www.pdb.org>) dok su 33 datoteke u postupku pohranjivanja.

U okviru ove disertacije strukturno je okarakterizirano nekoliko proteina važnih za preživljavanje bakterije *H. pylori*: FlgD, CrdA, HP1026.

Pokretljivost stanice bakterije *H. pylori* uvjetovana je pravilnom strukturnom organizacijom biča. Bič se sastoji od otprilike 50 različitih proteina važnih za pravilnu regulaciju i udruživanje proteina, neophodnih za pokretanje stanice u bolje uvjete za preživljavanje [H. Zhou, M. Luo, X. Cai, J. Tang, S. Niu, W. Zhang, Y. Hu, Y. Yin, A. Huang, and D. Wang, *Proteins* **79** (7) (2011) 2346–2351]. Glavne komponente koje izgrađuju bič su: filament, kuka (engl. *hook*) i bazalno tijelo (engl. *basal body*) [O. A. Soutourina and P. N. Bertin, *FEMS Microbiol. Rev.* **27** (2003) 505–523]. FlgD pridonosi virulenciji bakterije *H. pylori* budući da je nužan za kontrolu pravilne strukturne organizacije kape biča (engl. *hook cap*) [W. T. Kuo, K. H. Chin, W. T. Lo, A. H. Wang, and S. H. Chou, *J. Mol. Biol.* **381** (1) (2008) 189–199]. Do sada su riješene dvije kristalne strukture proteina FlgD iz bakterija *Xanthomonas campestris* [W. T. Kuo, K. H. Chin, W. T. Lo, A. H. Wang, and S.H. Chou, *J. Mol. Biol.* **381** (1) (2008) 189–199] i *Pseudomonas aeruginosa* [H. Zhou, M. Luo, X. Cai, J. Tang, S. Niu, W. Zhang, Y. Hu, Y. Yin, A. Huang, and D. Wang, **79** (7) *Proteins* (2011) 2346–2351] međutim sličnost proteinskih sekvenci između navedenih homologa je manja od 20 %.

CrdA (HP1326) je protein iz bakterije *H. pylori* važan za održavanje razine bakrovih iona u citoplazmi ispod toksične razine [B. Waidner, K. Melchers, F. N. Stähler, M. Kist, and S. Bereswill, *J. Bacteriol.* **187** (2005) 4683–4688]. Sve što se do danas zna o spomenutom proteinu su studije o RNA profiliranju [B. Waidner, K. Melchers, I. Ivanov, H. Loferer, K. W.

Bensch, M. Kist, and S. Bereswill, *J. Bacteriol.* **184** (2002) 6700–6708] koje upućuju na uključenost proteina u homeostazu bakrovih iona, međutim nedostaje stuktorna karakterizacija.

Geni za proteine koji se sintetiziraju uslijed izlaganja *H. pylori* termičkom šoku (engl. *heat shock proteins*; HSPs) organizirani su kao tri višecistronska operona. Njihova transkripcija kontrolirana je trima promotorima ( $P_{gro}$ ,  $P_{hrc}$ ,  $P_{cbp}$ ). Infekcija i prilagodba bakterije *H. pylori* ekstremnim uvjetima želuca omogućena je i spomenutim HSP-ima koji su uključeni u regulaciju aktivnosti ureaze i prijanjanje bakterijskih stanica za stijenku epitela želuca [D. Roncarati, A. Danielli, and V. Scarlato, *J. Bacteriol.* **193** (20) (2011) 5629–5636]. HP1026 pripada ovoj skupini proteina i u potpunosti je neokarakteriziran stoga je razumijevanje funkcije ovog proteina iznimno važno.

Cilj ovog istraživanja je strukturno karakterizirati proteine FlgD, CrdA i HP1026 iz bakterije *H. pylori* i potencijalno objasniti njihovu ulogu s naglasakom na razumijevanju funkcije spomenutih proteina pomoću podataka dobivenih strukturnom i biofizikalnom karakterizacijom. Prema dosadašnjim istraživanjima pretpostavke su: (1) FlgD kontrolira pravilnu polimerizaciju proteina FlgE koji izgrađuje kuku biča, (2) CrdA regulira razinu bakrovih iona u stanici ispod toksične razine te (3) HP1026 posjeduje ATP-veznu domenu. Rezultati istraživanja pridonijet će boljem razumijevanju uloge esencijalnih metala u strukturi i funkciji homeostaze, kontrole sinteze strukturnih elemenata biča bakterije te uloge proteina sintetiziranog uslijed izlaganja stanice termičkom šoku.

Istraživani proteini (FlgD, CrdA i HP1026 iz organizma *Helicobacter pylori*) proizvedeni su u laboratoriju primjenom tehnologije rekombinantne DNA, ugradnjom u plazmidne vektore i prekomjernom ekspresijom genskih produkata u bakteriji *Escherichia coli*. Ispravnost ugradnje gena za željenu makromolekulu provjerena je sekvenciranjem, a identitet proteina s privjeskom utvrđen je *western* analizom. U slučaju kad je prisutnost His-privjesaka ometala ispravno formiranje kvaterne strukture željenog proteina, za kloniranje su korišteni drugi privjesci (GST-privjesak, Strep-privjesak) ili je privjesak bio uklonjen. GST-privjesak također pomaže i u topljivosti proteina koji bi u protivnom formirao inkluzijska tijela.

Izolacija proteina iz bakterije *E. coli* započeta je razaranjem bakterijskih stanica sonikacijom ili razbijačem stanica (engl. *cell cracker*), te pročišćavanjem različitim kromatografskim metodama (afinitetna kromatografija, gel-filtracija), ovisno o prisustvu, odnosno odsustvu privjeska na proteinu. Čistoća završnog uzorka analizirana je korištenjem diskontinuirane denaturirajuće elektroforeze (SDS-PAGE). Koncentracija dobivenih proteina

---

određena je spektrofotometrijski, mjerenjem apsorbancije pri 280 nm, a prema teorijskim vrijednostima ekstinkcijskih koeficijenata, te izračunata prema Beer-Lambertovom zakonu. U slučaju velike razine ekspresije ali male topljivosti primjenjena je denaturacija i renaturacija proteina.

Kružni dikroizam (engl. *circular dichroism*, CD) i jednodimenzijski NMR korišteni su radi dobivanja informacije o prisutnosti elemenata sekundarne strukture u proteinu što je iznimno važno nakon denaturacije i ponovne renaturacije proteina. Kvaliteta uzorka za kristalizaciju ispitana je metodom dinamičkog raspršivanja svjetlosti (engl. *dynamic light scattering*, DLS). Interakcija CrdA i bakrovih iona ispitana je UV-VIS titracijom dok je aktivnost proteina HP1026 (sposobnost hidrolize ATP-a) istražena praćenjem promjene apsorbancije produkta reakcije hidrolize ATP-a.

Testiranje optimalnih uvjeta za nastanak kristala i optimizacija kristalizacije provedeni su metodom difuzije pare otapala iz sjedeće kapi upotrebom robota za automatsku kristalizaciju (Oryx8 robot) u nanolitarskoj skali (minimalna veličina kapi je 200 nL). Za kristalizaciju su korištene komercijalne kristalizacijske otopine, a nakon dobivanja prvih kristala kristalizacijski uvjeti su optimizirani u svrhu dobivanja kristala dobrih difrakcijskih sposobnosti variranjem različitih čimbenika (cijepljenje mikrokristalima, promjena temperature kristalizacije, dodatak aditiva, promjena omjera precipitanta i proteina u kristalizacijskoj kapljici, dehidracija kristala). Dobiveni kristali, ukoliko je bilo potrebno, namakani su u krio otopini, a potom skladišteni u tekućem dušiku. Snimanje kristala proteina obavljeno je na sinkrotronu ESRF (Grenoble, Francuska) ili SLS (Villigen, Švicarska) metodom rentgenske difrakcije na jediničnom kristalu. Kristalna struktura proteina riješena je bilo metodom molekulske zamjene (u slučaju postojanja dovoljno sličnog modela) ili metodom anomalnog raspršenja teškog elementa (u tu svrhu pripremljen je derivat proteina sa selenometioninom).

Gen *hp0907* ukloniran je iz dva različita bakterijska soja *H. pylori* (26695, G27). Proteini su pročišćeni do zadovoljavajuće čistoće te je gel-filtracijom i MALS eksperimentom (engl. *multi angle light scattering*) određeno da se cjeloviti proteini u otopini nalaze kao dimerne molekule. Jedinični kristali su pripremljeni metodom difuzije pare otapala iz sjedeće kapi korištenjem Oryx8 robota te su dobivene dvije različite kristalne forme skraćenog proteina FlgD. FlgD\_G27 kristalizira u monoklinskom sustavu, *P2* prostornoj grupi, a difrakcijski podatci su sakupljeni na sinkrotronskoj liniji ID14-4 (ESRF, Grenoble, Francuska) do razlučivanja od 2,75 Å. FlgD\_26695 i SeMetFlgD\_26695 kristaliziraju u tetragonskom sustavu,

---

I422 prostornoj grupi, a podaci su sakupljeni na sinkrotronskoj liniji PXIII (SLS, Villigen, Švicarska) do razlučivanja od 2,17 Å u slučaju FlgD\_26695 te do razlučivanja od 2,8 Å u slučaju SeMetFlgD\_26695. Iako je u oba slučaja postavljena kristalizacija čitavog proteina (FlgD\_G27 – 316 aminokiselina; FlgD\_26695 – 301 aminokiselina), riješene kristalne strukture zajedno s rezultatima masene spektrometrije ukazuju da se u kristalnoj strukturi proteina FlgD\_G27 i FlgD\_26695 nalazi skraćen protein FlgD (Asn127 – Lys272). Kristalne strukture proteina FlgD\_G27 (*HpFlgD\_m*) i FlgD\_26695 (*HpFlgD\_t*) razlikuju se samo u dvije aminokiseline (257 i 268) i njihovim preklapanjem se dobije standardna devijacija između 146 jednakih C<sup>α</sup> atoma od 0,38 Å. U asimetričnoj jedinici *HpFlgD\_m* nalaze se 4 molekule monomera i zauzimaju  $V_m = 2,63 \text{ \AA}^3 \text{ Da}^{-1}$ . Struktura sadrži oko 53 % otapala. Struktura *HpFlgD\_t* ima u asimetričnoj jedinici samo 1 molekulu monomera i zauzima  $V_m = 3,26 \text{ \AA}^3 \text{ Da}^{-1}$  te udio otapala iznosi oko 62 %. Nadalje, kvaterna struktura u obje kristalne forme je tetramer koji je u strukturi *HpFlgD\_t* generiran preko kristalografske osi četvrtog reda dok je u strukturi *HpFlgD\_m* generiran nekristalografskom simetrijom. Dimeri se stvaraju preko kontakata β-lanca 9 i β-lanca 12 tudor domena susjednih monomera, a takva situacija se ponavlja četiri puta stvarajući mrežu vodikovih veza između 4 molekule monomera. Rezultati dobiveni korištenjem PISA programa (engl. *Protein interfaces, surfaces and assemblies*) ukazuju da je takav tetramer vjerojatno postoji i u otopini. Razlika u oligomerizaciji cjelovitog proteina *HpFlgD* u otopini (dimer) i skraćenog proteina *HpFlgD* u kristalu (tetramer) može biti prisustvo N-kraja proteina blizu centra tetramera gdje se nalazi os četvrtog reda. Dulji N-kraj bi onemogućio stvaranje kontakata nužnih za tetramerizaciju. Kristalna struktura proteina FlgD je pokazala da se monomer sastoji od dvije domene, tudor i Fn-III domene, kao što je slučaj i u ostala dva homologa proteina FlgD čije su strukture riješene. Fn-III domenu izgrađuju aminokiseline 141 – 234 dok tudor domenu čini 14 aminokiselina na N-kraju proteina (127 – 140) i ostatak aminokiselinskog slijeda na C- kraju proteina (235 – 272). Lys272 je posljednja vidljiva aminokiselina u strukturama monoklinske i tetragonske forme proteina *HpFlgD*. Međutim, značajna razlika je uočena u međusobnom položaju tudor domene u odnosu na Fn-III domenu u *HpFlgD* spram ostalim homolozima (*XcFlgD* i *PaFlgD*). Različita orijentacija domena između *HpFlgD* i ostalih homologa može upućivati na prisutnost dva prepoznatljiva orijentacijska modula koji predstavljaju dio mehanizma aktivacije proteina.

Pretraživanjem *Uniprot* banke podataka uočeno je da sve sekvence homologa proteina FlgD u različitim vrstama roda *Helicobacter* sadrže prepoznatljive i neuređene regije na C-kraju



proteina redoslijeda XQK(X)<sub>2</sub> (X = D / E / N; X<sub>2</sub> = PI / PL / LS / PQ). Očuvane regije variraju samo u broju ponavljanja, a uvijek završavaju slijedom TPPKETA. Ovakav ponavljajući motiv nije pronađen u ostalim homolozima proteina FlgD, ali su slični motivi pronađeni u ostalim bakterijskim proteinima što može ukazivati na važnost takvog motiva u prepoznavanju proteinskih partnera. Štoviše, FliC protein (gradi filament biča) stvara interakcije preko svog C-kraja sa šaperonom FliS te se na taj način sprječava preuranjena polimerizacija proteina FliC. Također, nedostatak N-kraja u strukturi *HpFlgD* nije neobičan budući da ga i strukture ostalih homologa ne sadrže, a poznato je da FlgD iz bakterije *Escherichia coli* sadrži 71 aminokiselinu kao signalnu sekvencu dok FlgD iz bakterije *Salmonella enterica* sadrži čak i 100 aminokiselina kao signal za izlazak iz stanice. U prilog tomu da N-kraj proteina predstavlja signalnu sekvencu i da je zbog fleksibilnosti sklon degradaciji govori i analiza CD spektroskopijom koja je potvrdila da cjelovit protein *HpFlgD* sadrži  $\alpha$  zavojnice (12,8 %) što je obilježje signalnih sekvenca. U riješenoj kristalnoj strukturi *HpFlgD*  $\alpha$ -zavojnice nisu prisutne.

Priređeni su različiti genski konstrukti za ekspresiju proteina CrdA (Strep\_CrdA, CrdA\_His<sub>6</sub> i GST\_CrdA). Optimalna ekspresija proteina dobivena je pri nižoj temperaturi (16 °C), međutim u slučaju CrdA\_His<sub>6</sub> i Strep\_CrdA dobivene su velike količine proteina u obliku inkluzijskih tijela dok je protein GST\_CrdA prekomjerno eksprimiran s velikim udjelom u topljivim frakcijama. Kako bi se povećala topljivost proteina, uspješno su primjenjene metode za ponovno smatanje proteina, uključujući denaturaciju te renaturaciju polaganim razrjeđivanjem. Različite varijante rekombinantnog proteina su nadalje pročišćene afinitetnom kromatografijom i gel-filtracijom. Metodama kružnog dikroizma te jednodimenzijskog NMR-a potvrđeno je postojanje elementa sekundarne strukture u renaturiranom proteinu. Na temelju analitičke gel-filtracije izračunata je molekulska masa proteina CrdA s His- privjeskom u prisutnosti iona Cu<sup>2+</sup> (14,8 kDa). Dobivena vrijednost se podudara s predviđenom molekulskom masom (12,4 kDa) i odgovara postojanju monomernih vrsta u otopini. Dodatkom reagensa za keliranje metala (EDTA) dolazi do značajne promjene u ponašanju proteina CrdA, te tendencije proteina ka različitom stupnju oligomerizacije. Protein CrdA je ugušćen do maksimalne koncentracije,  $\gamma = 9 \text{ mg mL}^{-1}$ . Kvaliteta uzorka za kristalizaciju ispitana je DLS-om koji je ukazao na polidisperznost uzorka dok je dodatak deterdženta n-oktil- $\beta$ -D-glukopiranozida (nOG) deterdženta i iona Cu<sup>2+</sup> povećao monodisperznost uzorka. Afinitet iona Cu<sup>2+</sup> prema

proteinu CrdA ispitan je UV-VIS spektroskopijom te je izračunata  $K_d = 1,076 \times 10^{-3} \text{ mol L}^{-1}$  što upućuje na nizak afinitet prema ionima  $\text{Cu}^{2+}$  što je karakteristika proteina koji prenose bakrove ione. Međutim, budući da aminokiselinski slijed CrdA obiluje metioninima, a ne sadrži histidin i cistein, potrebno je ispitati afinitet iona  $\text{Cu}^+$  spram CrdA. Za kristalizacijske eksperimente korišten je robot za automatsku kristalizaciju (Oryx8 robot). Pritom je korištena metoda difuzije pare iz sjedeće kapi te je ispitano 686 različitih kristalizacijskih uvjeta. Kristalizacijski eksperiment je u slučaju proteina CrdA u prisutnosti iona  $\text{Cu}^{2+}$  i nOG-a rezultirao formiranjem malih kristala u uvjetu s PEG'S II SUITE (Qiagen) (0,1 mol  $\text{L}^{-1}$ , Tris pH 8,5, (v/v) 16 % PEG 4000) pri 293 K. Difrakcijska mjerenja dobivenih kristala provedena su na liniji ID23-1 ESRF sinkrotrona (Grenoble, Francuska) te su potvrdila da su nastali kristali proteina. Nažalost, kristali su slabo difraktirali (do razlučivanja od 20 Å), te je potrebna daljna optimizacija kristalizacije.

Slijed aminokiselina proteina HP1026 sadrži očuvane sljedove aminokiselina 36 – 295 što svrstava protein u skupinu  $\text{AAA}^+$  proteina. Proteini iz spomenute skupine se odlikuju sposobnošću vezanja i hidrolize ATP-a, te sudjelovanjem u različitim staničnim procesima (transport i degradacija proteina, transkripcija, replikacija i sl.). Protein HP1026 uspješno je eksprimiran pri 28 °C, te je zatim pročišćen afinitetnom kromatografijom i gel-filtracijom. Molekulska masa monomera HP1026 iznosi 43,8 kDa te je analitičkom gel-filtracijom ustanovljeno da su i Apo\_HP1026 i kompleks HP1026:ATP- $\gamma$ -S prisutni kao dimerne vrste (96,9 kDa) u otopini, te da dodatak ATP- $\gamma$ -S ne potiče oligomerizaciju. Nadalje, ispitana je ATP-azna aktivnost HP1026, te su dobiveni sljedeći kinetički parametri:  $K_m = 344 \text{ } \mu\text{mol L}^{-1}$ ,  $V_{\text{max}} = 0,6019 \text{ } \mu\text{mol L}^{-1} \text{ s}^{-1}$ ,  $k_{\text{cat}} = 0,02 \text{ s}^{-1}$ . Dobivene vrijednosti se podudaraju s kinetičkim parametrima u slučaju aktivnosti proteina MgsA iz bakterije *Escherichia coli*. Analiza sekundarne strukture proteina CD metodom je pokazala da se HP1026 glavninom sastoji od  $\alpha$ -zavojnica (44,5 %). Protein se pokazao kao izuzetno topljiv te je ugušćen do koncentracije veće od 40 mg  $\text{mL}^{-1}$ , a mjerenja DLS-om su pokazala da je uzorak proteina monodisperzan i pogodan za kristalizacijske eksperimente. Također je ispitana termička stabilnost proteina uz dodatak različitih aditiva, te je uočeno da ADP i ATP- $\gamma$ -S značajno povećavaju temperaturu mekšanja ( $T_m$ ) proteina HP1026 i to za +3,52 °C u slučaju dodatka ATP- $\gamma$ -S te za +5,18 °C u slučaju dodatka ADP-a. Kristalizacijski eksperimenti postavljeni su korištenjem Oryx8 robota metodom difuzije pare otapala iz sjedeće kapi. Kristali heksagonskog oblika su dobiveni u

različitim uvjetima PACT SUITE (Qiagen), te su snimljeni na sinkrotronskoj liniji PXIII (SLS, Villigen, Švicarska) i pokazali slabu difrakciju (do razlučivanja od 15 Å). Uzevši u obzir rezultate o termičkoj stabilnosti proteina pokušala se optimizacija kristalizacije uz dodatak aditiva, te su dobiveni veći kristali u uvjetu 0,2 mol L<sup>-1</sup> NaBr, 0,1 mol L<sup>-1</sup> Bis-tris propan pH 8,5, 20 % PEG 3350, 0,015 mmol L<sup>-1</sup> CYMAL-7 pri 20 °C. Iako su kristali bili veći i uređeniji, difraktirali su samo do 7 Å što još uvijek nije dovoljno razlučivanje za rješavanje kristalne strukture proteina.

---

## § 1. INTRODUCTION

The bacterium *Helicobacter pylori* colonizes the stomach of more than half of the world's population.<sup>1</sup> According to the National Center for Biotechnology Information (NCBI) the genomes of seventeen strains have been completely sequenced like 26695, J99, HPAG1 and G27.<sup>2-5</sup> All of the genomes have similar average lengths of coding sequences and coding density. The extremely high recombination rate, in combination with a remarkable sequence diversity, has suggested a panmictic population structure.<sup>6</sup> Approximately 10 % of the infected subjects will develop severe gastric pathologies, like peptic ulcer disease and atrophic gastritis. Approximately 1 % of infected individuals develop gastric adenocarcinoma and lymphoma of the mucosa-associated lymphoid tissue (MALT). The World Health Organization has declared *H. pylori* a class 1 carcinogen in 1994.<sup>7-9</sup>

Due to its pathological relevance, *H. pylori* has become an important target for research in the last twenty years, since its first discovery: a large amount of literature has been produced on it, either from the medical or from the biological point of view; interaction maps of its proteome have been created<sup>10</sup> and analysed and its ability to survive in the acidic environment has been investigated.<sup>11</sup> On the contrary, structural data on the proteins of this bacterium are relatively limited: 439 files of atomic coordinates were present in the *Protein Data Bank* (PDB) in March 2016 (<http://www.pdb.org>), and 33 are waiting for release. Many proteins have not been characterized at all.

Within this research several proteins from *H. pylori*, important for its survival, were characterized (FlgD, CrdA, HP1026).

The flagellum is a protein complex composed of about 50 different proteins needed for its proper regulation and assembly. It provides the cell a chance to move to a more favorable environment and to avoid stressful conditions.<sup>12</sup> Major sections that define the flagellum are: the filament, the hook and the basal body.<sup>13</sup> FlgD plays an important role in *H. pylori* virulence, owing to its role in forming the hook cap component.<sup>14</sup> Until now only two crystal structures of FlgD have been solved (from *Xanthomonas campestris* and *Pseudomonas aeruginosa*) but the protein sequence identity between FlgD from *H. pylori* and FlgD from the previously mentioned bacteria is less than 20 %.

*H. pylori* copper resistance determinant CrdA (HP1326) is required for keeping the concentration of free copper ions below toxic levels.<sup>15</sup> At this moment there are no structural data on the CrdA protein and the only available studies are on RNA profiling.<sup>16</sup>

Genes for heat shock proteins (HSPs) of *H. pylori* are organised in three multicistronic operons, transcriptionally controlled by three upstream promoters ( $P_{gro}$ ,  $P_{hrc}$ ,  $P_{cbp}$ ). Successful infection and adaption to the extreme gastric environment is also enabled by the heat shock proteins (HSPs) which are involved in regulation of the urease activity and adhesion to the epithelial cells.<sup>17</sup> Since HP1026 is part of one of the heat shock operons controlled under the HspR regulator with an unclear function, investigation of its role is very important.

Biophysical and structural characterization of the described proteins is crucial for understanding the function of these *H. pylori* vital proteins for survival in a highly acidic environment. The starting hypotheses of this work were: (1) FlgD controls the proper polymerization of FlgE inside the hook of the flagellum, (2) the CrdA protein is involved in maintaining the concentration of free copper ions below toxic levels and (3) the HP1026 protein possesses an ATP-binding domain.

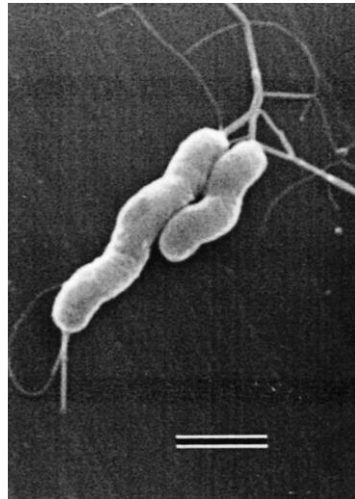
The goal of the doctoral dissertation was to structurally characterize FlgD, CrdA and HP1026, and taking into account their biophysical and structural properties to potentially demonstrate their function. The structural characterization was done by biophysical and structural analysis utilizing methods like thermal stability assay, circular dichroism, X-ray diffraction and 1D-NMR. To understand if CrdA exerts function of copper transport the interaction between the copper ions and CrdA was investigated by UV-VIS titration. To obtain experimental evidence on the role of HP1026, the ATP hydrolysis assay was performed using the reversed phase HPLC by measuring the change of the absorbance of the reactions products.

The results obtained in this research widen our knowledge and contribute to better understanding of the role of the essential metal ions in the structure and function of metalloproteins, the control of proper flagellum construction, as well as the function of the heat shock protein.

## § 2. LITERATURE REVIEW

### 2.1. *Helicobacter pylori* – an interesting medical target

*Helicobacter pylori* (*H. pylori*) is a small (3.5 µm x 0.5 µm), slow growing, microaerophilic, non-spore forming and spiral-shaped Gram-negative rod bacteria (Fig. 1). The bacterium became an important and interesting target of research since it is known that it can colonize human gastric cells and lead to development of severe gastrointestinal diseases such as chronic gastritis, peptic ulcer disease, gastric mucosa associated lymphoid tissue (MALT) lymphoma, and gastric cancer.<sup>18,19</sup> The bacterium was isolated and characterized at the beginning of the '80s by scientists Marshall and Warren.<sup>20</sup> It is known that more than half of the human population is affected with this bacterium.<sup>21</sup> Moreover, this bacterium has accompanied humans for tens of thousands of years during migrations in the absence of a targeted antimicrobial therapy.<sup>7</sup> It has been hypothesized that *H. pylori* colonization could have provided benefits to its human carriers and hence provided a selective advantage during long periods of human history.<sup>22,6,9</sup> The infection with this bacterium is mainly transmitted from human-to-human possibly by the fecal–oral or oral–oral route. The infection can be spread vertically within families, generally in early childhood. While many *H. pylori* infected individuals are clinically asymptomatic, most will exhibit some degree of gastritis. Approximately 10 % of the infected subjects will develop more severe gastric pathologies, like peptic ulcer disease and atrophic gastritis. Approximately 1 % of infected individuals develop gastric adenocarcinoma and lymphoma of the mucosa-associated lymphoid tissue (MALT), a correlation that prompted the World Health Organization to declare *H. pylori* the first bacterial class 1 carcinogen in 1994.<sup>8,9</sup>



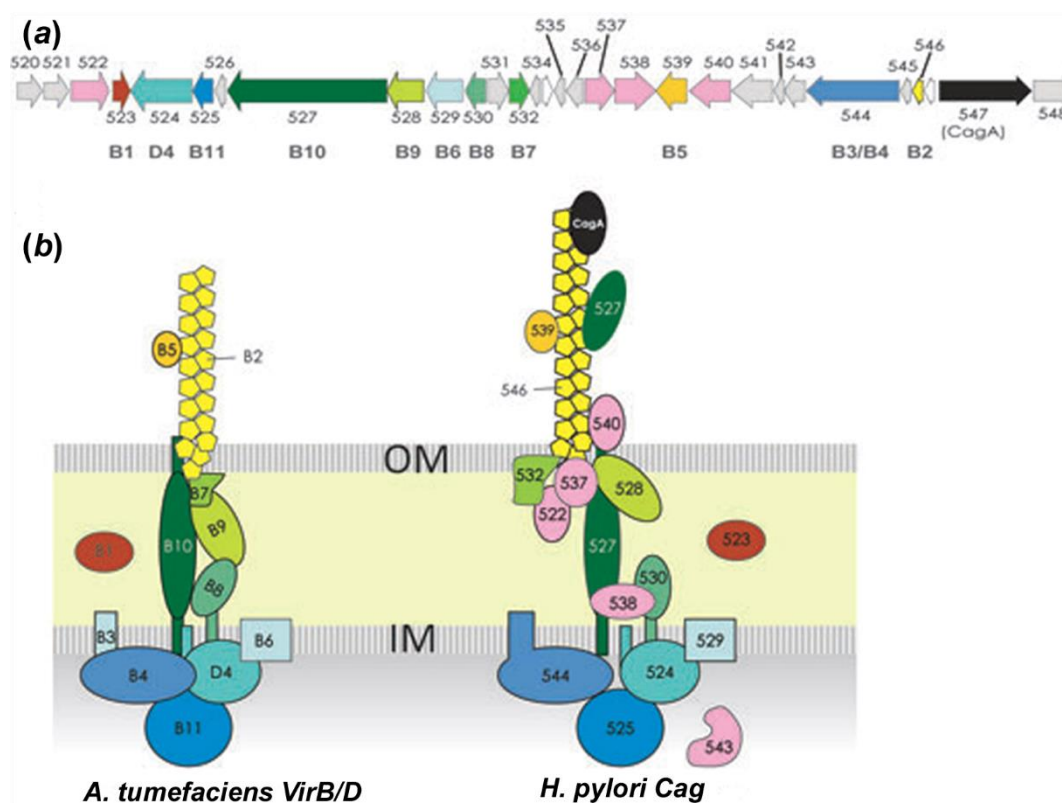
**Figure 1.** Electron micrograph of *H. pylori*. Spiral bacterial bodies with a bundle of sheathed flagella; bar = 1  $\mu\text{m}$ , (taken from Yoshiyama and Nakazawa, 2002).<sup>23</sup>

*H. pylori* genome is relatively small containing less than 1600 genes,<sup>2</sup> in comparison to other Gram-negative prokaryotes, such as *Escherichia coli* and *Salmonella enterica* subsp. *enterica* serovar, whose genomes include more than 4000 genes.<sup>24,25</sup> The genomes of seventeen strains of *H. pylori* have been completely sequenced, among them 26695, J99, HPAG1 and G27.<sup>2-5</sup> Genome of the strain 26695 contains 1,667,867 base pairs. It is 24 kb larger than the J99 strain, which contains 1,643,831 base pairs, 71 kb larger than HPAG1, which contains 1,596,366 base pairs and 15 kb larger than the G27 strain, which contains 1,652,983 base pairs. All the genomes have similar average lengths of coding sequences, coding density and the bias of initiation codons. Plausible cause of a wide genetic diversity could be explained by the need of *H. pylori* to adapt to the gastric conditions of its host as well as to the distinct patterns of the host immune response.

## 2.2. *H. pylori* virulence factors

Several pathogenicity factors like *cytotoxin associated gene* pathogenicity island (*cag* PAI), VacA vacuolating cytotoxin, urease, adhesins and outer membrane proteins are involved in virulence of *H. pylori*.<sup>26</sup> Large diversity of *H. pylori* strains indicated that some of the strains are more virulent and can induce morphological changes and successive degeneration of in vitro-cultured cells.<sup>27</sup>

CagA was identified as a main marker for the evaluation of the strain pathogenicity and the presence of a genomic PAI. *Cag* PAI is a DNA fragment about 40 kb long which encodes for about 27 to 31 proteins, depending on the analyzed strain (Fig. 2).<sup>28,29</sup> Eighteen proteins of *cag* PAI are involved in the structural organization of the type IV secretion system (T4SS) which is the main apparatus used in the infection process. T4SS enables penetration to the gastric epithelial cells and translocation of CagA, peptidoglycan and other bacterial effectors to the host cell.<sup>30–32</sup> After the injection of CagA into the host cells the protein becomes phosphorylated and starts to interact with more than 20 different human proteins included in signal transduction.<sup>33</sup> This influences the proper function of epithelial cells like cell–cell adhesion, signalling, adherence and proliferation.<sup>34</sup> The bacterial peptidoglycan delivered by T4SS triggers in the host cell the Nod1-response and induction of the nuclear factor- $\kappa$ B pathway.<sup>35</sup>

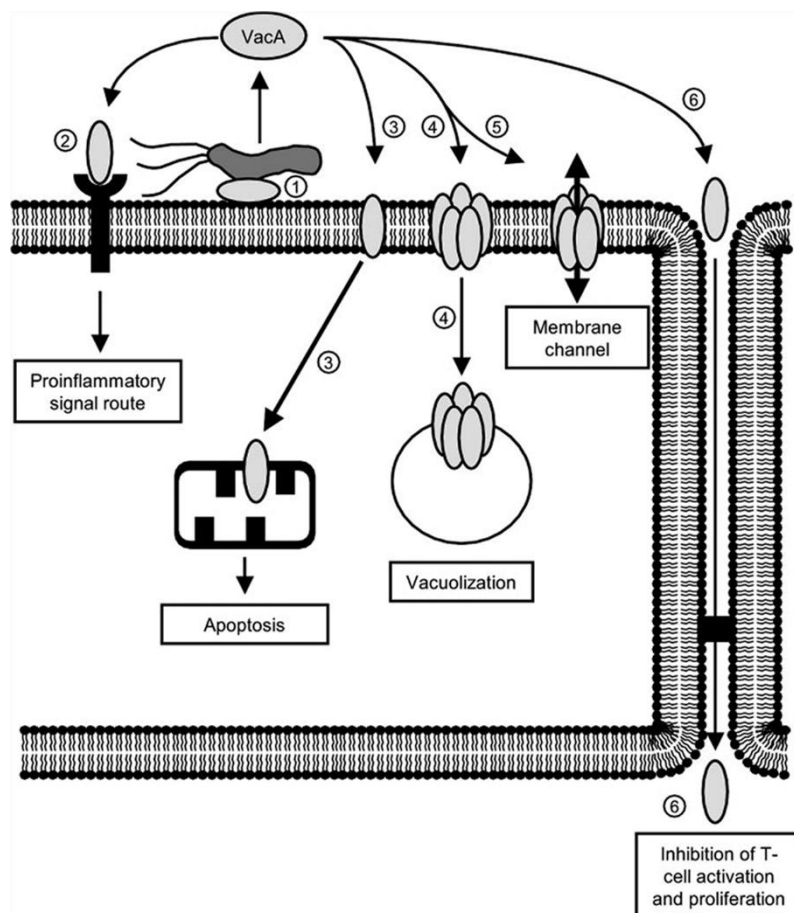


**Figure 2.** (a) Genes of the *H. pylori* strain 26695 that encode for the *cag* PAI. Numbers correspond to the HP0XXX number of the ORF represented by arrows;<sup>2</sup> (b) schematic overview of the prototypal T4SS VirB/D from *A. tumefaciens* (left) and comparison with components of the Cag-T4SS from *H. pylori* (right). Cytoplasmic NTPases, proteins forming the core trans-



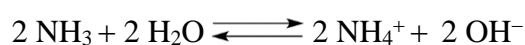
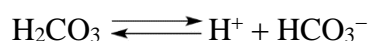
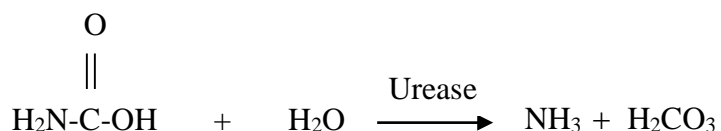
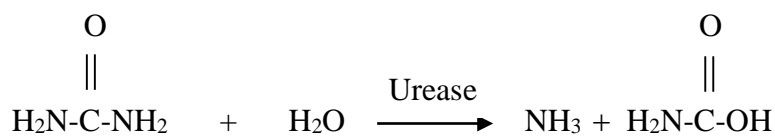
membrane complex and pilus components are coloured in blue, various shades of green and yellow/orange, respectively. Integral transmembrane segments or proteins are shown as squares. The additional components that have been shown to participate in the Cag-T4SS complex are coloured in pink while the effector CagA that has been located at the tip of the pilus is coloured in black. The figure is adapted from Terradot and Waksman, 2011.<sup>32</sup>

VacA is another protein classified as a pathogenic factor of *H. pylori*.<sup>26</sup> In its active form it is a 95 kDa toxin responsible for massive vacuolization in gastric epithelial cells in vitro.<sup>36</sup> Moreover, it is shown that peptic ulceration and gastric cancer can be caused by secretion of VacA.<sup>37,38</sup> Firstly, VacA is produced as a 140 kDa proprotein that is further cleaved into the 95 kDa mature protein during the secretion. There is a significant difference in the pathogenicity of this toxin encoded by the *vacA* gene among the different *H. pylori* strains. *VacA* gene can encode for several types of VacA protein that has different composition of signal and middle regions (s1/m1, s1/m2, s2/m2). S1/m2 genotypes have the highest vacuolating activity and are more involved in the peptic ulceration and gastric carcinoma, while the s2/m2 genotypes lack the vacuolating activity.<sup>37</sup> It is also interesting to highlight that even the specific *vacA* genotype of the same patient-strain combination shows a different expression level of VacA over time, showing the ability of rapid evolution of this bacterium to adapt its genetic material for persistent infection.<sup>39,40</sup> When secreted, VacA forms pores in the membrane of the epithelial host cells allowing the release of urea, anions, cations and nutrients from the host cells. The VacA protein is also included in several cellular processes like induction of apoptosis (through the activation of the endogenous mitochondrial channels) and inhibition of T-cells activation and proliferation (Fig. 3).



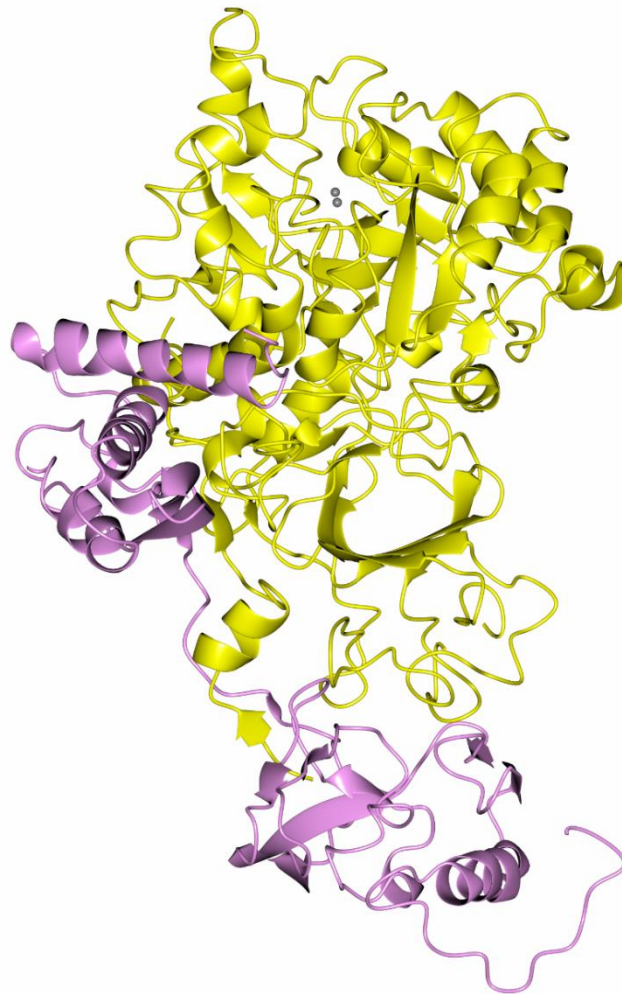
**Figure 3.** The influence of VacA on several cellular processes and its involvement in the infection process of *H. pylori*. This figure is taken from Kusters *et al.*, 2006.<sup>26</sup>

Since *H. pylori* colonizes an acidic niche like gastric epithelial cells, regulation of acidity is an important mechanism of survival of this bacterium in these conditions. *H. pylori* survives in the range of pH 4 – 6.5 and thus this bacterium requires special mechanisms to adapt to the extreme acidic environment. Urease is the main enzyme that can increase pH by catalyzing the hydrolysis of urea that generates ammonia and carbon dioxide.<sup>26</sup> Ammonia neutralizes the acidity inside the gastric lumen and forms a neutral microenvironment surrounding the bacterium (Scheme 1).<sup>41</sup>



**Scheme 1.** Hydrolysis reaction of urea catalyzed by urease.

In general, urease can be found inside the cytoplasm of all other bacteria and plants,<sup>42</sup> while the urease from *H. pylori* can also be present on the surface of the cell but only when it undergoes spontaneous autolysis.<sup>43</sup> The role of cytoplasmic urease is protecting the bacterium from acidity because it could increase the periplasmic pH and membrane potential in combination with UreI, a proton-gated urea channel<sup>44,45</sup> while the external urease is essential for survival of the acid-exposed organism. The crystal structure of nickel-containing urease from *H. pylori* (Fig. 4) shows that the protein is built from two different subunits -  $\alpha$  and  $\beta$  ( $\alpha$  - 61.7 kDa,  $\beta$  - 26.5 kDa) forming a heterotrimeric  $(\alpha\beta)_3$  assembly.<sup>46</sup>



**Figure 4.** Cartoon view of the crystal structure of urease from *H. pylori* (PDB ID: 1E97).<sup>46</sup>  $\alpha$  and  $\beta$  subunits are colored in yellow and pink, respectively, while the two nickel(II) ions are labelled as grey spheres. This figure was produced using CCP4MG.<sup>47</sup>

Adhesins and outer membrane proteins are an additional class of *H. pylori* virulence factors.<sup>26</sup> They are responsible for the adhesion of bacterial cells to the gastric epithelium. A list of part of the adhesion proteins involved in the infectious process is given in Table 1.

**Table 1.** Adhesins and their role in the pathogenesis of *H. pylori*

<b>Protein / gene cluster</b>	<b>Predicted role</b>	<b>Association with <i>H. pylori</i> related disease</b>
BabA	Binds to fucosylated Le <sup>b</sup> blood group antigen on cells <sup>48</sup>	<i>babA2</i> allele has been implicated in peptic ulcer disease and gastric cancer <sup>49</sup>
SabA	Mediates binding to sialic acid containing glycoconjugates (sialyl-Le <sup>x</sup> and sialyl-Le <sup>a</sup> antigens) and is involved in the activation of neutrophils <sup>50</sup>	–
SabB	Unknown binding specificity	Absence of SabB expression by the phase variation is involved in the development of duodenal ulcers <sup>51</sup>

In addition to the previously described virulence factors, several other factors are important for the successful colonization and survival of *H. pylori* in the gastric epithelial cells such as motility, detoxification, metal ion homeostasis and heat shock regulation. They will be in more detail discussed in the next paragraphs.

### 2.3. *H. pylori* flagellum

*H. pylori* motility requires proper structural organization of 2–6 unipolar flagella.<sup>52</sup> In order to survive in the hard stomach environment and to permanently colonize it, *H. pylori* has to move through the mucous layer and adhere to gastric epithelial cells, in particular during the initial phases of the infection<sup>53,54</sup> and in doing so it has to rely on flagella.<sup>55,56</sup> In contrast to many other Gram-negative bacteria, *Helicobacter* (and *Campylobacter*) species possess an unusual velocity in viscous media, possibly due to their helical shapes and to the presence of exclusively polar flagella.<sup>55</sup> The structural organization and control of flagella in Gram-negative bacteria has been thoroughly studied.<sup>55, 57–59</sup> The complex flagellar structure is composed of approximately 30 different proteins, but many additional proteins (like chaperons and pumps)

are necessary for flagella expression and assembly, resulting in a total of at least 45 proteins (listed in Table 2).

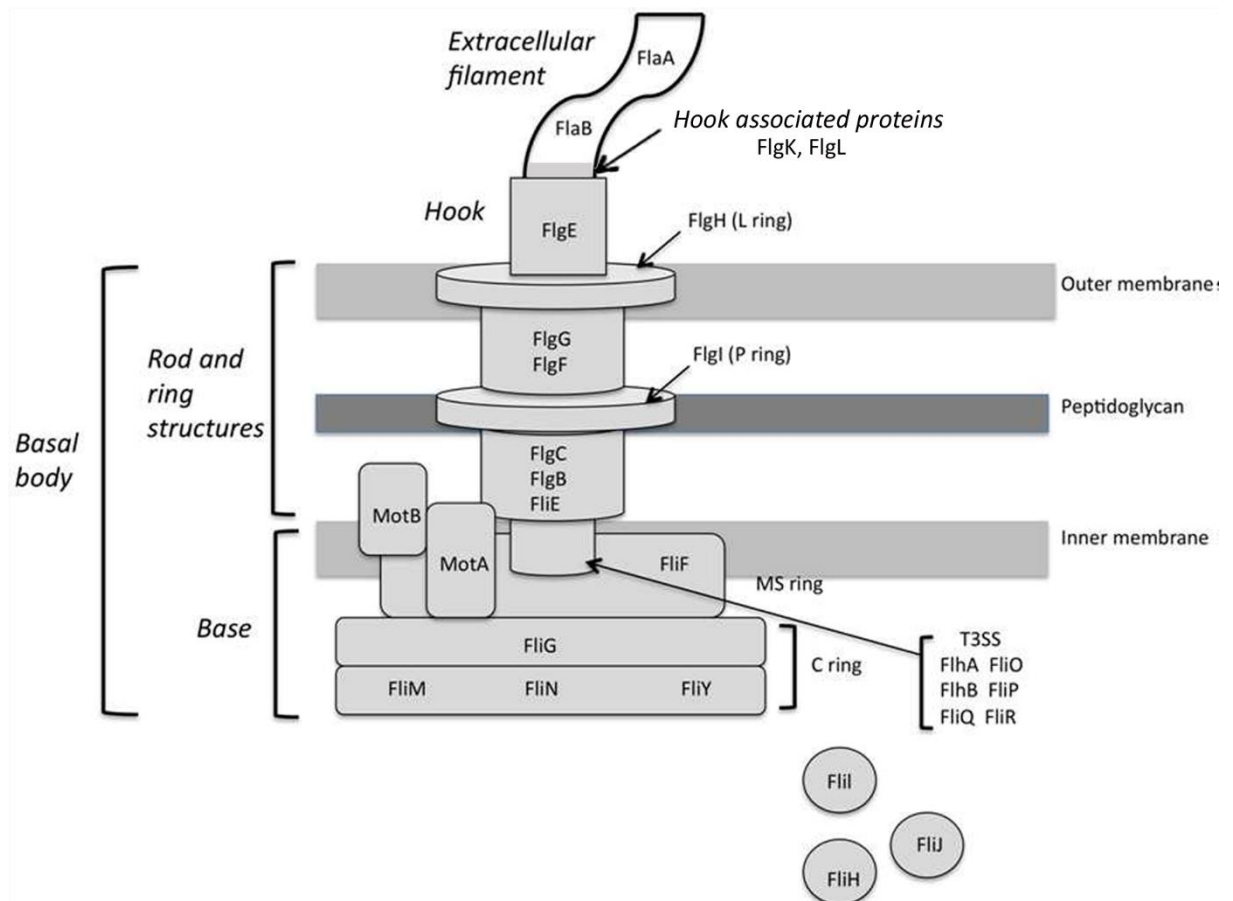
**Table 2.** *H. pylori* proteins involved in flagellum assembly (adapted from Pulić *et al.*, 2014).<sup>60</sup>

Location	Protein name	Name for strain 26695	Typology	Role	Complex	Number of AA	PDB-ID structure		
Export system	FliO	HP0583	Membrane protein	Protein associated with FliP		293			
	FliP	HP0685				Protein associated with FliO and MS-ring		153	
	FliQ	HP1419				Export component		Associated with FlhA and FlhB	88
	FliR	HP0173		Associated with the MS ring	225				
	FlhA	HP1041		Target for soluble export proteins	733	3MYD			
	FlhB	HP0770		Target for soluble proteins	358				
	FliH	HP0353	Soluble protein	Negative regulator of FliI	The complex FliHIJ forms an ATPase complex	258			
	FliI	HP1420				ATPase; export protein outside of C-ring		Interaction with FliH	434
	FliJ	HP0256				Putative protein		142	
	FlgJ	HP0245		Putative protein	105				
MS ring	FliF	HP0351	Transmembrane protein	Transmembrane component of rotor	Associated with FlhA	389			
C ring	FliG	HP0352	Cytoplasmic protein	Rotor switch protein	Associated with MotA	343	3USW; 3USY		
	FliM	HP1031				Involved in the rotor switch; binds the CheY-P protein	N-terminus interacts with FliF	354	4GC8, 4FQ0
	FliN	HP0584				Flagellar export component; role in switching protein		106	
	FliY	HP1030				Flagellar export component		283	
Stator	MotA	HP0815	Transmembrane protein	Forms a proton channel through the membrane		210			
	MotB	HP0816				Associated with MotA		257	3SO3, 3SOH (periplasmic region), 3IMP, 3CYP, 3CYQ (C-terminal domain)
P ring	FlgI	HP0246		Part of bushing		285			
L ring	FlgH	HP0325		Part of bushing		237			
Proximal rod	FliE	HP1557		Has a role in rod stability during the growth; part of the export gate; MS ring rod junction protein		109			
	FlgB	HP1559			140				
	FlgC	HP15558			161				
Distal rod	FlgF	HP1092				269			
	FlgG	HP1585				262			
Hook-filament	FlgK	HP1119	Soluble	HAP1		606			
	FlgL	HP0295			HAP3	828			

junction protein							
Filament capping	FliD	HP0752		HAP2			685
Hook cap	FlgD	HP0907	Soluble	Rod-modification protein; required for hook polymerization			301
Hook	FliK	HP0906		Hook length regulator			77
	FlgE	HP0870		Hook protein	Helped by FlgD in polymerization		718
Filament	FlaA	HP0601		Flagellin subunit	Polymerizes with FlaB		510
	FlaB	HP0115 HP1076		Flagellin subunit			514
	FliS	HP0753	Cytosolic protein specific for FlaA, FlaB		Interacts with FliS		120
Regulatory protein	FlgM	HP1122		Encodes the inhibitor of flagellum $\sigma^{28}$ -factor			77
	FliA	HP1032		Negative regulator of FlgM			255
	FlgR	HP0730		Activates transcription with $\sigma^{54}$ -factor			381
Chemotactic factor	CheY	HP1067		Chemotaxis	Interacts with FliM		124
Chaperon	FlgN	HP1457		Chaperon FlgK-FlgL			210
	FliT	HP0754		Chaperon FliD			79
	FlgA	HP1477	Periplasmic protein	Chaperon involved in P-ring assembly			218
Flagellar biosynthesis protein	FlhF	HP1035					459
Flagellar assembly factor	FliW1	HP1154					135
Paralysed flagellar protein	pFla	HP1274					801
Flagellar assembly factor	FliW2	HP1377					129

The flagellum is a rotatory nano-machine that can be described as composed of two portions, the extracellular filament and the hook-basal body (Fig. 5). The latter represents the flagellar motor that converts the chemical into mechanical energy.<sup>57</sup> The hook-basal body can be divided into three substructures: (1) the base, localized in the inner membrane and spanning to the cytoplasm; (2) the rod and ring structures, located in the periplasm; and (3) the hook, present on the surface. The assembly of the flagella and the stator requires proper interactions with the peptidoglycan layer through which the organelle has to pass for externalization.<sup>61</sup> Roure *et al.* (2012)<sup>61</sup> demonstrated that even though the flagella were correctly assembled, lack of the

appropriate lytic transglycosylase (MltD) resulted in incorrect localization of the flagellar motor protein *HpMotB* to the bacterial pole and with a loss of motility.



**Figure 5.** Structural organization of the *H. pylori* flagellum (adapted from Pulić *et al.*, 2014).<sup>60</sup>

Until now little is known about the structure and assembly of flagellar proteins from *H. pylori*. Moreover, only seven 3D structures (including flagellar chaperons) have been determined so far (Table 2).

The relevance of studying the molecular architecture of flagellar proteins is crucial for designing new therapeutic targets since flagella are components essential for bacterial colonization of the host. On the other side, the flagellum represents an ideal example of a molecular machine and the mechanism of conversion of chemical into mechanical energy. The bacteria have been able to devise it during billion years of evolution and it can eventually be applied for nano-technological purposes.



### 2.3.1. The hook

The hook, which is a tubular structure that connects the basal body to the extracellular filament, is made of about 120 copies of a single protein, called FlgE. Until now, only the crystal structure of FlgE from *Salmonella enterica subsp. enterica serovar* Typhimurium (PDB ID: 1WLG)<sup>62</sup> has been determined at 1.8 Å resolution. The structure corresponds to residues 71-369 out of a total of 402. The crystallized fragment of FlgE lacks both the N- and C-terminal regions, since the full-length protein formed filaments and thus failed to crystallize. The same behavior occurs during the crystallization of the full-length flagellin protein. The group of authors that solved the crystal structure of FlgE built a model of the hook by using electron cryomicroscopy and image analysis, together with the docked crystal structure of FlgE. According to the density map, the hook is composed of three domains: the outermost domain at the surface (7.5 nm), the middle domain (5-6 nm) and the inner core domain that forms a tube (1 nm thick; 3 nm axial lumen).

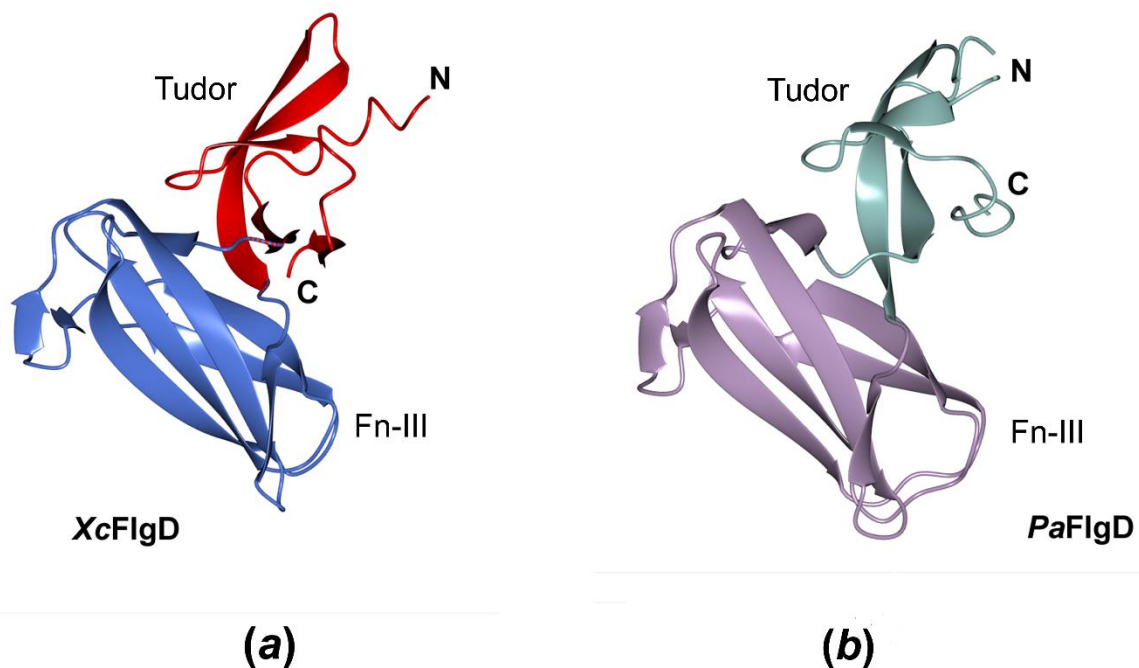
FlgE is a protein that tends to form filaments and, for this reason, it needs the presence of other proteins, FliK, FlhB and FlgD, in order to properly assemble the hook outside the external membrane of the bacterium.<sup>63</sup> FliK is supposed to be involved in measurement of the correct hook length<sup>64,65</sup> while FlhB (part of the type III secretion system, T3SS) is located in the inner membrane and helps the hook formation by interacting with the C-domain of FliK. When this interaction occurs, it induces a signal for termination of the export of proteins involved in the hook assembly and a signal for the export of proteins necessary for the filament formation.<sup>66</sup>

There are also two proteins between the hook and the filament, called *hook associated proteins* - FlgK and FlgL. These proteins form a very short hook-filament junction zone, important for adapting these two mechanically different structures. The hook is relatively flexible, while the filament works as a propeller and for this reason has a much more rigid structure. According to this fact, those two proteins should share structural characteristics similar to the hook and filament proteins.

### 2.3.2. FlgD protein

FlgD in *Salmonella typhimurium* is predicted to control the correct number of FlgE monomers that form the hook.<sup>67</sup> In this bacterium FlgD is absolutely needed for the assembly of the flagellar hook, but it has not been detected in the mature flagellum.<sup>68</sup> FlgD, like the majority of the flagellar extracellular components, is exported via a specific T3SS, located at the base of the flagellum. After assembly of the basal body in the cytoplasmic membrane and the C-ring in the cytoplasmic space, the flagellar export apparatus is built within the C-ring and gets ready to export the axial components, among which is FlgD, to construct the rod, hook, hook-filament and the long filament, according to this order.<sup>66</sup> FlgD is associated to the apparatus when the rod is completed, but it is discarded as soon as the hook is completed, i.e. when FlgK is added to the hook. During this process the interaction between FlgE and FlgD may take place.<sup>69</sup>

Two crystal structures of FlgD from other bacteria have been determined: *Pseudomonas aeruginosa* (PaFlgD, PDB ID: 3OSV)<sup>12</sup> and *Xanthomonas campestris* (XcFlgD, PDB ID: 3C12)<sup>14</sup>, Fig. 6. Both structures lack the N-terminal domain, which is predicted to be largely flexible, at least for the isolated proteins. A study of FlgD from *Escherichia coli*<sup>70</sup> indicated that the first 71 N-terminal residues represent a signal for the export of the protein through a T3SS into the flagellar channel. T3SS is an essential part of the flagellum apparatus that allows the proper export of the proteins necessary for the assembly of the flagellum itself. In both FlgD structures monomer is composed of two domains rich in  $\beta$  strands: Fn-III nad tudor domains (Fig. 6). Kuo *et al.* (2008)<sup>14</sup> performed DALI search against the PDB using the coordinates of the two separate domains of XcFlgD. They found that tudor domain superimposes well with the spinal muscular atrophy tudor domain and a number of chromo domain (a member of tudor family that can interact with specific methylated lysines in histones). Fn-III domain superimposes well with the Fn-III domain of cell adhesion and heparin-binding proteins.



**Figure 6.** Crystal structure of FlgD monomer from: (a) *X. campestris* (PDB ID: 3C12)<sup>14</sup>, (b) *P. aeruginosa* (PDB ID: 3OSV)<sup>12</sup>. This figure is produced using CCP4MG.<sup>47</sup>

#### 2.4. Detoxification and metal ion homeostasis in *H. pylori*

Regulation of metal ion homeostasis is of critical importance to all living organisms. Particularly, adaptation of *H. pylori* to the conditions in the gastric mucosa involves acquisition mechanisms that overcome a temporary lack of the metal ions. Metal ions have role in various biochemical reactions and both deficiency and excess of particular metal ion can lead to the cell death or growth delay. Metal ions are cofactors of enzymes that catalyze reactions like electron transport, energy metabolism, redox reactions and are involved in the regulation of the osmotic pressure of the cell.<sup>26</sup>

Nickel ions are essential for the proper structural organization and activity of the main pathogenicity factors, urease and hydrogenase. Moreover, several accessory proteins like HypA, HypB and UreG are needed for the maturation of the [Ni,Fe] hydrogenase and urease.<sup>71,72</sup> In the acidic microaerobic environment *H. pylori* utilizes FeoB protein for transport of the ferrous ion ( $\text{Fe}^{2+}$ ).<sup>73</sup> Ferric reductase is another type of enzyme that reduces ions  $\text{Fe}^{3+}$  to  $\text{Fe}^{2+}$

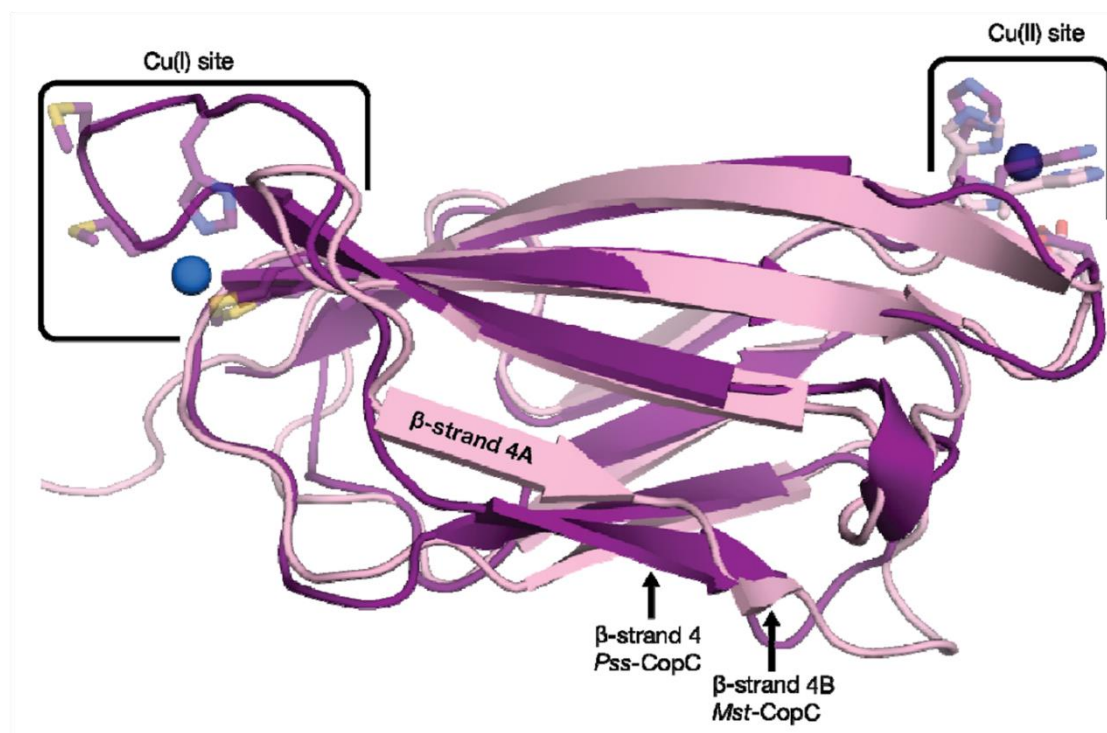
which are then transported by the FeoB system. Beside the mentioned iron dependent enzymes, iron storage proteins in *H. pylori* have been characterized – Pfr ferritin and HP-NAP bacterioferritin. Pfr ferritin is a 19 kDa protein known to be responsible for the regulation of the concentration of free iron ions below the toxic levels as well as an iron deposit that can release and reuse the stored iron under the reduced iron conditions.<sup>74</sup> Cobalt ions are also important for the activity of arginase enzyme which is involved in the nitrogen metabolism<sup>75</sup> and immune response of *H. pylori*.

#### 2.4.1. Copper ion homeostasis

Besides nickel, iron and cobalt, copper ion is also an important cofactor in the regulation of metal homeostasis in *H. pylori* and plays significant role in electron transport, oxidases, hydroxylases and can mediate the formation of the oxygen reactive species. In addition, several proteins are involved in the transport of copper ions and control of concentration of free copper ions in the cytoplasm below toxic values. Among them are P-type ATPase CopA,<sup>76</sup> HP1326 (CrDA), HP1327 (CrDB), HP1328 and HP1329.<sup>16</sup> An interesting study reported by Bereswill *et al.* (2000)<sup>77</sup> suggested that the ferric uptake regulator (Fur) protein participates not only in the regulation of ferritin-mediated iron storage but also in the copper repression of Pfr synthesis. They showed that iron starvation, as well as medium supplementation with nickel, zinc, copper and manganese at nontoxic concentrations, repressed synthesis of ferritin in the wild-type strain but not in the *H. pylori fur* mutant.

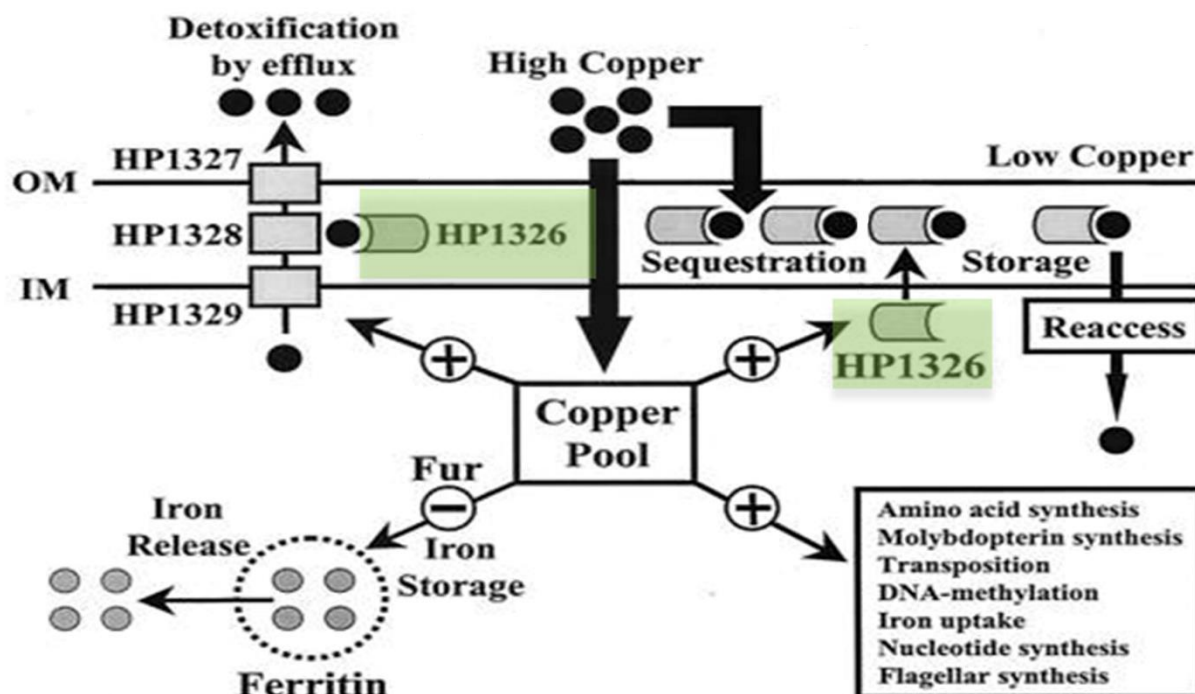
#### 2.4.2. CrdA protein

CrDA (HP1326) is a 13.8 kDa putative copper resistance determinant rich in methionines. In contrast to known copper chaperones, the protein sequence of CrDA lacks cysteine and histidine residues. Since CrDA is a secreted protein, it is predicted that it contains the first 20 amino acids as signal peptide. Protein sequences of CopC from *Pseudomonas syringae* and CrDA from *H. pylori* share a common MXXMPGMP amino acid motif. Crystal structures of the copper binding CopC protein from *Pseudomonas syringae* (PDB ID: 2C9Q),<sup>78</sup> and *Methylosinus trichosporium* (PDB ID: 5ICU),<sup>79</sup> revealed that this class of proteins can contain different binding arrangements: a canonical two binding site (Cu(II) and Cu(I)) or just a single Cu(II) binding site, Fig. 7.



**Figure 7.** Structure alignment of *MstCopC* (light pink) and *PssCopC* (purple). Copper ions are presented as blue spheres. This figure is taken from Lawton *et al.* (2016).<sup>79</sup>

At this moment, there are no structural data on the CrdA protein and the only available studies are RNA profiling.<sup>16</sup> According to the former study ORF HP1326 that encodes the *H. pylori* CrdA protein, showed a strong transcription activity upon increased copper supplementation while the nonheme iron containing ferritin (Pfr) synthesis was repressed (Fig. 8). Similar behaviour was observed when the neighboring genes HP1327 (CrdB) and HP1328 (CzcB) were inactivated. Other homologs of known copper regulation activity like *E. coli* CueR, *Pseudomonas* CopR/S and *Ralstonia* CzcRS are not present in the *H. pylori* genome. Taking all together, Waidner *et al.* (2002)<sup>16</sup> proposed a novel type of copper efflux pump called Czc system that is consisted of copper resistance determinants CrdA (HP1326), CrdB (HP1327), CzcB (HP1328) and CzcA (HP1329), Fig. 8, which *H. pylori* requires for keeping the concentration of free copper ions below toxic levels.

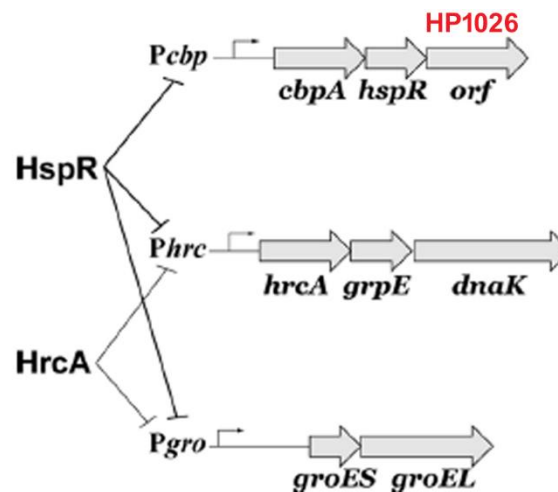


**Figure 8.** Schematic overview of *H. pylori* responses to increased copper concentrations. Induction (+) and repression (–) of gene expression mediated by the increased copper pool are indicated by arrows. The putative composition of the Czc system of *H. pylori* HP1326 to HP1329 was constructed using the *Ralstonia* Czc system as a model. Putative location of CrdA (HP1326) protein is indicated in green. This figure is adapted from Waidner *et al.* (2002).<sup>16</sup>

## 2.5. Heat shock proteins in *H. pylori*

Heat shock proteins (HSPs) are, together with the transcription regulatory factors HrcA and HspR, another important factors of *H. pylori* successful infection and adaptation to the extreme gastric environment.<sup>80,17</sup> These factors, besides their main role in protection of the bacterium from the environmental stresses, are also involved in the regulation of urease activity and adhesion to epithelial cells.<sup>81,43</sup> Generally, Gram-negative bacteria like *E. coli* show positive regulation of HSPs through a specialized RNA polymerase  $\sigma^{32}$  factor, which is absent in the genome of *H. pylori*. Opposite to the *E. coli* regulation, transcription control of HSPs in *H.*

*pylori* is negatively regulated by HrcA and HspR. The HSPs of *H. pylori* are organised in three multicistronic operons, transcriptionally controlled by three upstream promoters ( $P_{gro}$ ,  $P_{hrc}$ ,  $P_{cbp}$ ), Fig. 9. It has been demonstrated that under the normal growth conditions two regulators – HrcA and HspR are both needed to repress transcription from  $P_{hrc}$  and  $P_{gro}$  promoters in an independent<sup>80</sup> or cooperative manner<sup>82</sup> while HspR by itself controls the  $P_{cbp}$  promoter. The same dual heat shock regulation is presented in *Staphylococcus aureus* where CtsR and HrcA synergistically repress transcription of *dnaK* and *groESL* operons.<sup>83</sup> When *H. pylori* is exposed to heat shock the repression is blocked and HSP transcription is induced. Moreover, Roncarati *et al.* (2007)<sup>80</sup> supposed that controlled synthesis of HSPs such as GroES, GroEL and DnaK chaperones could affect the flagellar assembly or other cellular structures and contribute to the proper motility.



**Figure 9.** Structural organization of *H. pylori* heat shock operons and their transcriptional regulation controlled by HrcA and HspR regulators. *Hp1026* gene (highlighted in red) is part of one of the heat shock operons regulated by HspR. This figure is adapted from Roncarati *et al.* (2011).<sup>17</sup>

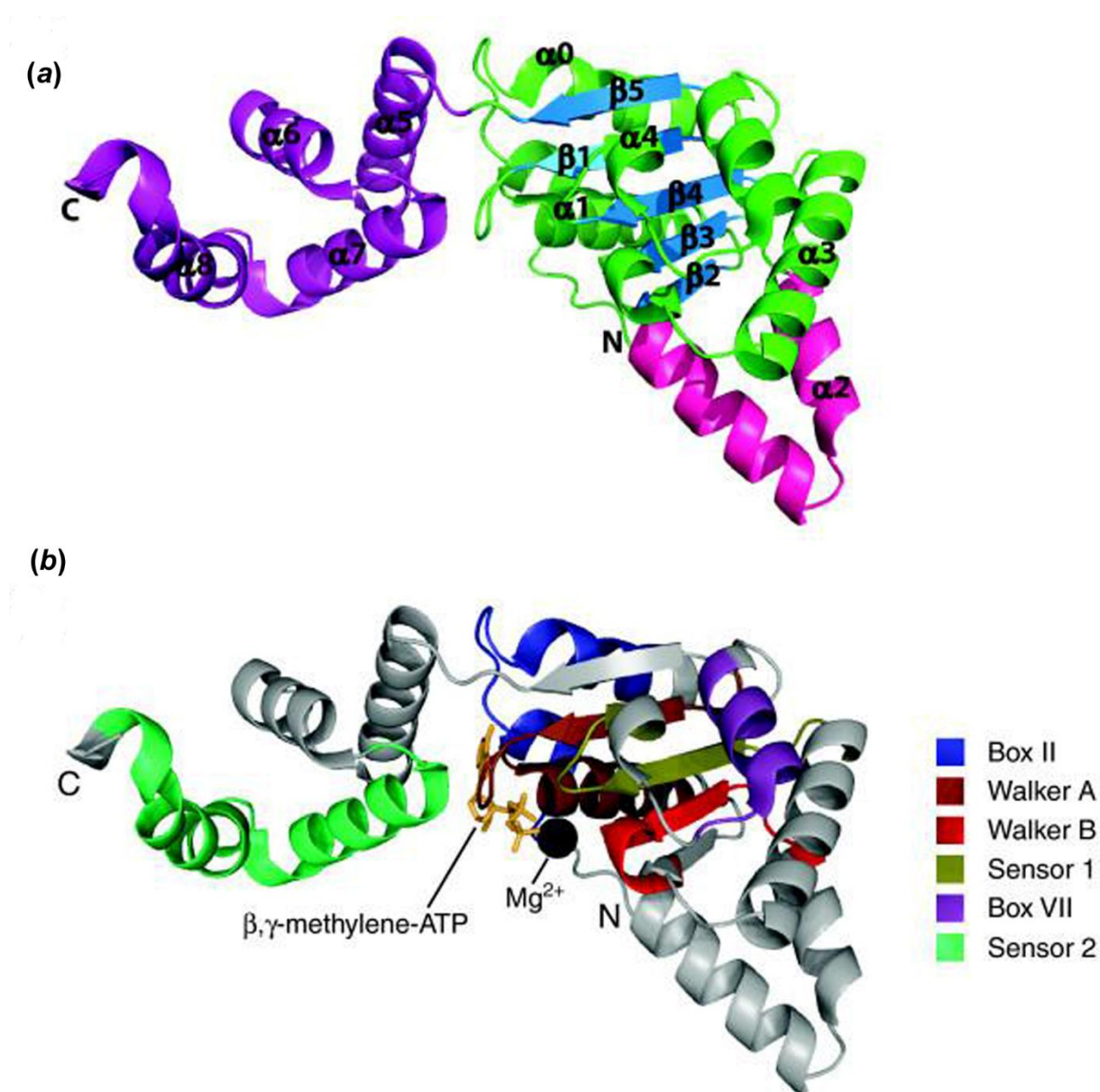
### 2.5.1. HP1026 protein

HP1026 is an uncharacterized protein with an interesting gene location under the control of one of the heat shock promoters ( $P_{cbp}$ ), and the HspR repressor. Two other genes, *cbpA* and *hspR*, that are located upstream of the *hp1026* gene (Fig. 9) are regulated by the same promoter and repressor. It is known that *cbpA* codes for a homologue of cochaperone curved DNA-

binding protein CbpA of *E. coli*<sup>84</sup> while *hspR* codes for the negative regulator of  $P_{gro}$ ,  $P_{cbp}$  and  $P_{hrc}$  promoters.<sup>85</sup>

HP1026 is a 43.8 kDa protein that is according to UniProt database predicted to be a conserved hypothetical helicase-like protein. The protein is classified to belong to the AAA<sup>+</sup> superfamily of ATPases. AAA<sup>+</sup> designation describes protein involvement in processes like facilitation of protein folding and unfolding, assembly or disassembly of proteins complexes, protein transport and degradation, function in replication, recombination, repair and transcription.<sup>86</sup> AAA<sup>+</sup> superfamily comprises proteins that possess the conserved structural central ATPase domain of about 250 amino acids called the AAA<sup>+</sup> module. The AAA<sup>+</sup> domain is often divided into two structural subdomains, an N-terminal P-loop NTPase alpha-beta-alpha ( $\alpha\beta\alpha$ ) subdomain that is connected to a smaller C-terminal alpha helical ( $\alpha$ ) subdomain. The  $\alpha\beta\alpha$  subdomain adopts a Rossman fold and contains several motifs involved in ATP binding and hydrolysis, including classical motifs Walker A and Walker B (Fig. 10).<sup>87</sup> The  $\alpha$  subdomain is much less conserved across AAA<sup>+</sup> proteins. AAA<sup>+</sup> proteins usually function as oligomeric rings, with a common hexameric arrangement that provides a hole through which DNA or RNA can be thread.





**Figure 10.** Crystal structure of the DnaA monomer belonging to the AAA<sup>+</sup> superfamily from *Aquifex aeolicus* (PDB ID: 2HCB)<sup>88</sup> (a) The  $\alpha$ -helices and random coils (colored green) and the  $\beta$ -strands (colored blue) of the core  $\alpha\beta\alpha$  nucleotide-binding domain, with the exception of the two equal-sized helical inserts, which are colored pink. The small  $\alpha$ -helical domain is colored purple. (b) Major motifs in the AAA<sup>+</sup> module are colored as indicated in the legend. The bound adenosine 5'-[ $\beta,\gamma$ -methylene]triphosphate ( $\beta,\gamma$ -methylene-ATP, a nonhydrolyzable ATP analog, orange sticks) and  $Mg^{2+}$  (black sphere) are also shown. This figure is adapted from Snider *et. al* (2008).<sup>87</sup>

*H. pylori* HP1026 shares an overall amino acid identity of 31 % and 26 % with *E. coli* MgsA (maintenance of genome stability protein A) and *Archaeoglobus fulgidus* RFC (replication factor C), respectively. MgsA and RFC are proteins involved in the process of DNA replication while according to the UniProt database, HP1026 is supposed to be a helicase like protein. *EcMgsA* is important in facilitating the recovery of stalled replication forks while *AfRFC* is involved in opening and closing the DNA polymerase processivity factors onto DNA. Although the amino-acid sequence homology is low between HP1026 and described proteins the conserved fold of the nucleotide-binding domain is present in them like in HP1026. Crystal structures of *EcMgsA* (PDB ID: 3PVS)<sup>89</sup> and *AfRFC* (PDB ID: 2CHQ)<sup>90</sup> revealed that both proteins possess an ATP binding site, as it is also predicted for the HP1026 protein. The *EcMgsA* monomer is composed of three domains: ATP binding and helical lid domains (conserved among AAA<sup>+</sup> proteins), and a tetramerization domain. The *AfRFC* protein is built of conserved AAA<sup>+</sup> domains 1 and 2, while domain 3 forms a five-helix bundle. As common for the AAA<sup>+</sup> proteins, *AfRFC* oligomerizes as a hexamer.

In general, *H. pylori* helicases are poorly structurally characterized. So far the only known crystal structures of *H. pylori* helicases are of proteins DnaB, a dodecameric replicative helicase (PDB ID: 4ZC0),<sup>91</sup> and of a helicase binding domain of the DnaG protein (PDB ID:4EHS).<sup>92</sup> This finding places structural investigations on the *H. pylori* helicase-like proteins in a challenging field of research.



## § 3. MATERIALS AND METHODS

### 3.1. Materials

#### 3.1.1. Equipment

**Table 3.** Devices and instruments (basic laboratory equipment is not listed).

<i>Instrument</i>	<i>Manufacturer</i>
ÄKTA Purifier	GE Healthcare
Avanti J-25 centrifuge	Beckman Coulter
Shaker	Thermo Scientific
Spectrophotometer Evolution 60	Thermo Scientific
One Shot Cell breakage system	Constant System Ltd., UK
Eppendorf centrifuge	Eppendorf
Vibra Cell Homogenizer	Bioblock Scientific
Ultrasonic bath	Elma
Electrophoresis Mini-PROTEAN Tetra System	BioRad
Heating/cooling dry block CH 100	Biosan
Nanodrop 1000	Thermo Scientific
Electrophoresis system 2117 Multiphor II	LKB
Oryx 8 robot	Douglas instruments
DLS, Zetasizer Nano ZS	Malvern Instruments Ltd
MALS, Astra 6.1	miniDAWN TREOS
Quadrupole–TOF spectrometer	Waters, Manchester, UK
Spectropolarimeter	Jasco Analytical Instruments
Microscope SZX 12	Olympus
UV fluorescence microscope MZ16F	Leica
T100 Thermal cycler	Bio-Rad
Real time PCR	Applied Biosystems
Molecular imaging system	Carestream Health, Inc.

## 3.1.2. Chromatography supplies

**Table 4.** Columns, media and molecular weight markers used for affinity and size-exclusion chromatography.

<i>Chromatography supplies</i>	<i>Manufacturer</i>
PD-10 desalting columns	GE Healthcare
His Trap Ni-NTA agarose	GE Healthcare
GSTrap	GE Healthcare
Strep-Tactin	IBA
Gel filtration molecular weight markers	Sigma Aldrich

## 3.1.3. Crystallographic supplies

**Table 5.** Crystallization screens.

<i>Screen</i>	<i>Manufacturer</i>
PEGs II, PACT SUITE, JCSG Core (Suites I–IV)	Qiagen
Structure Screen 1 & 2, PGA Screen, Solubility Tool Kit	Molecular Dimensions
Additive Screen	Hampton Research

**Table 6.** Crystallization plates.

<i>Type</i>	<i>Manufacturer</i>
MRC 96-well sitting drop crystallization plate	Molecular Dimensions
MRC MAXI 48-well sitting drop crystallization plate	Molecular Dimensions
24-well XRL plate	Molecular Dimensions

**Table 7.** Cryocrystallography equipment.

<i>Type</i>	<i>Manufacturer</i>
Mounted CryoLoops	Molecular Dimensions
EMBL/ESRF sample changer kit	Molecular Dimensions
Mounted Mesh LithoLoop	Molecular Dimensions

#### 3.1.4. Kits

**Table 8.** Kits used for preparation of DNA and for staining of the gels for visualizing proteins according to the manufacturers' protocols.

<i>Type</i>	<i>Manufacturer</i>
QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
ProteoSilver Silver Stain Kit	Sigma Aldrich

#### 3.1.5. Filters and membranes

**Table 9.** Ultra centrifugal filters used to concentrate dilute protein solutions, membranes used for filtering sterile aqueous solutions and transfer membrane used for western blotting.

<i>Type</i>	<i>Manufacturer</i>
Amicon Ultra centrifugal filter units with a molecular weight cut-off of 3 kDa, 10 kDa and 30 kDa, respectively.	Millipore
0.2 µm cellulose acetate membranes	Millipore
PVDF membrane, 0.2 µm	Millipore

#### 3.1.6. Chemicals and buffers

All buffers were prepared with redistilled water (MilliQ gradient water system) and then filtered using a Sartolab filter system with PES membrane (0.22 µm pore size). If not stated otherwise, all chemicals were obtained in analytical grade from the following companies: *Acros Organics*;

*AppliChem; Carl Roth; Fisher Scientific; Fluka; Jena Bioscience, Merck, Roche GmbH (Mannheim); Serva Electrophoresis; Sigma Aldrich.*

### 3.1.7. Enzymes

Restriction enzymes and polymerases for molecular biological applications as well as proteases for biochemical usage were obtained from *New England Biolabs (NEB)*, *Sigma Aldrich* and *ThermoFisher Scientific*. All enzymes were used in their appropriate buffer as specified by the manufacturer.

**Table 10.** Features of enzymes used for molecular and biochemical applications. Recognition sequences of restriction enzymes and proteases are in superscript.

<i>DNA modifying enzymes</i>	<i>Application</i>	<i>Manufacturer</i>
<i>Phusion</i> high-fidelity DNA polymerase	Amplification of DNA (proofreading)	ThermoFisher Scientific
<i>Taq</i> DNA polymerase	Amplification of DNA (no proofreading)	ThermoFisher Scientific
T4 DNA ligase	Ligation of DNA fragments	ThermoFisher Scientific
BamHI <sup>5' G GATCC 3'</sup>	Restriction of DNA	NEB
XhoI <sup>5' C TCGAG 3'</sup>	Restriction of DNA	NEB
DNase I	Digestion of DNA	Sigma Aldrich
Rnase A	Digestion of RNA	ThermoFisher Scientific
Trypsin <sup>K, R (on -COOH terminus)</sup>	Proteolysis	Sigma Aldrich
$\alpha$ - Chymotrypsin <sup>Y, W, F, L, M (on -COOH terminus)</sup>	Proteolysis	Sigma Aldrich
$\gamma$ - Thermolysin <sup>L, F (on -NH<sub>2</sub> terminus)</sup>	Proteolysis	Sigma Aldrich
Thrombin <sup>LVPR GS</sup>	Removal of His <sub>6</sub> -tag present at protein	Novagen

Prescission Protease <sup>LEVLFQ GP</sup>	expressed in vector
	pET 15b / Proteolysis
	Removal of GST-tag
	present at protein
	expressed in vector
	pGEX-6P-1
	Sigma Aldrich

### 3.1.8. Antibodies

**Table 11.** Features of primary and secondary antibodies.

Primary antibody	Secondary antibody	Application	Manufacturer
Monoclonal Anti-His <sub>6</sub> (mouse)	–	Detection of HHHHHH-epitope in proteins	Roche
–	Polyclonal Goat anti-Mouse Antibody, horseradish peroxidase (HRP) conjugate		General Electrics

### 3.1.9. Oligonucleotides

Oligonucleotides (Table 12) were synthesized by *Sigma-Aldrich*. The lyophilized and desalted oligonucleotides were dissolved in sterile water at a concentration of 100  $\mu\text{mol L}^{-1}$  and stored at  $-20\text{ }^{\circ}\text{C}$ .

**Table 12.** Oligonucleotides used as primers to amplify gene sequences.

Name	Sequence (5' to 3')	Application
GST_HP1326-for	GAGAGAGGGATCCATGCAAACCCTAAAAGCCAAC	Amplifying
GST_HP1326-rev	GAGAGAGCTCGAGTTATAAATCCAGGCTTGTTTTTAG	GST_HP1326



Strep_HP1326-for	GAAGGAGATATACATATGTGGAGCCACCCGCAGTTCGAAAA ACAAACCCTAAAAGCCAAC	Amplifying Strep_HP1326
Strep_HP1326-rev	GTGATGGTGGTGTGATGATGTTATAAATCCAGGCTTGTTTTAGC	
HP1326_His-for	GAAGGAGATATACATATGCAAACCCTAAAAGCCAACGAC	Amplifying
HP1326_His-rev	GTGATGGTGGTGTGATGATGTAAATCCAGGCTTGTTTTAGC	HP1326_His
HP0907_His-for	GGAGATATACATATGGCTATTGATTTAGCAGAAG	Amplifying
HP0907_His-rev	GTGATGGTGGTGTGATGATGTGCTGTCTCTTTAGG GG	HP0907_His

### 3.1.10. Vectors

**pET15b** (*Novagen*), Fig. 11 and Table 13, is a 5708 bp long vector that carries a N-terminal hexa His-tag sequence followed by a thrombin recognition site and MCS, including sites for XhoI, BamHI and NdeI. Expression of genes cloned into this vector is under the control of the *T7-lac* promoter. Additionally, the vector carries an ampicillin resistance gene (*ampR*) for selection.<sup>93</sup>

**pGEX-6P-1** (*GE Healthcare*), Fig. 11 and Table 13, is a 4984 bp long bacterial vector for expression of N-terminal GST-tagged recombinant proteins under the *tac* promoter. The vector contains a PreScission Protease recognition site for cleaving the protein from the GST-tag and an ampicillin resistance gene (*ampR*) for selection.<sup>94</sup>

**pETite C- His** (*Lucigen*), Fig. 11 and Table 13, is a 2235 bp long pre-processed vector that encodes the C-terminal hexa His-tag. The vector includes a kanamycin resistance gene (*kanR*) for selection and inducible expression under control of the *T7-lac* promoter.<sup>95</sup>

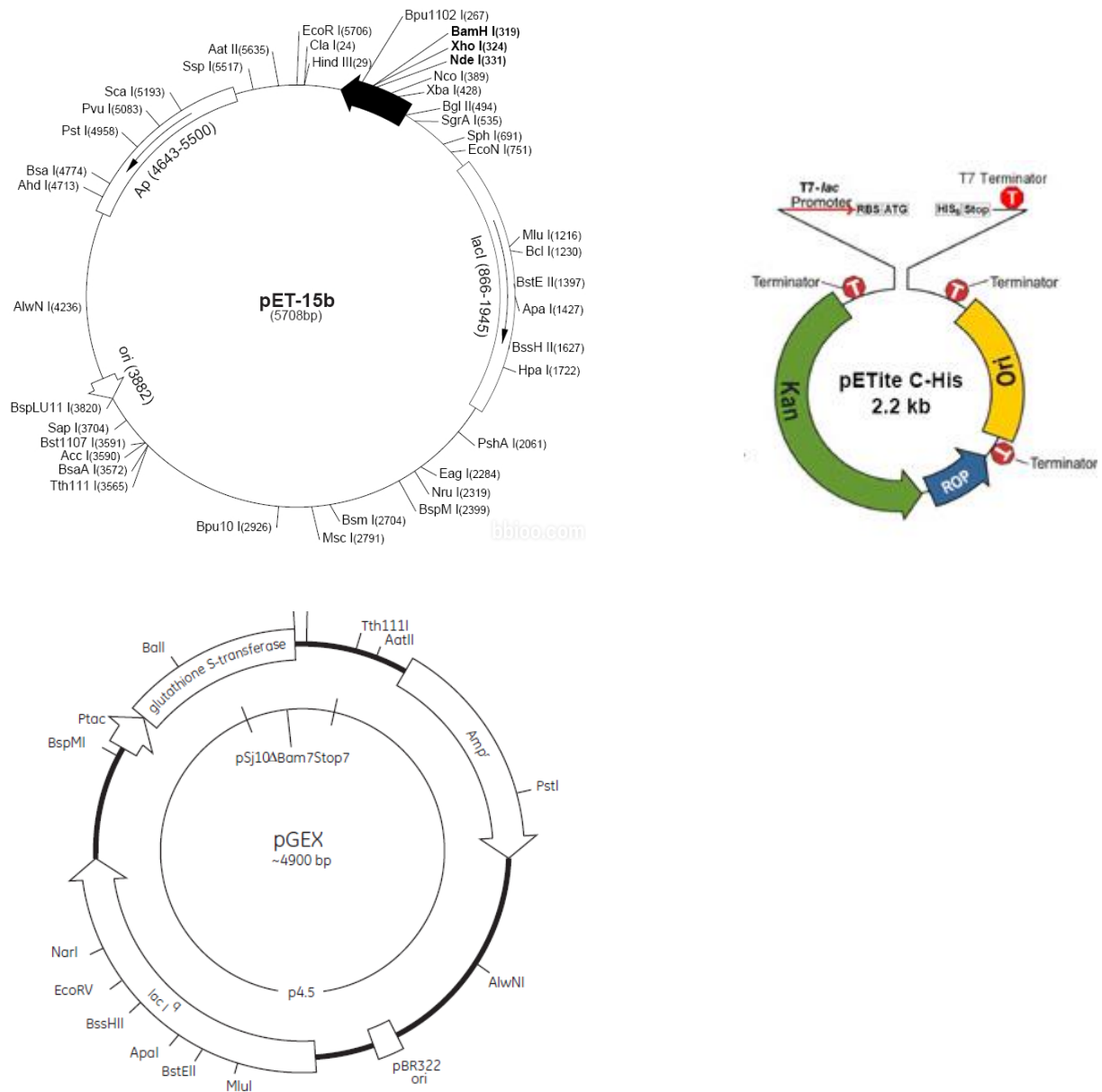


Figure 11. Schematic presentation of plasmid vector maps used for cloning.

**Table 13.** List of all designed constructs with starting and ending positions of the amino acid residues.

<i>Protein</i>	<i>Template</i>	<i>Vector</i>	<i>Start</i>	<i>End</i>
HP1326 (CrdA)	<i>H. pylori</i> G27	pETite C-His	23	125
		pGEX-6P-1	23	125
HP1026	<i>H. pylori</i> G27	pET15b	1	391
HP0907 (FlgD)	<i>H. pylori</i> G27	pETite C-His	1	316
	<i>H. pylori</i> 26695	pETite C-His	1	301

### 3.1.11. Bacterial strains: *Escherichia coli*

#### 3.1.11.1. Host strain for cloning

**One Shot TOP10** ( $F^-$  *mcrA*  $\Delta$ (*mrr-hsdRMS-mcrBC*)  $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$  *lacX74* *recA1* *araD139*  $\Delta$ (*araleu*)7697 *galU* *galK* *rpsL* (StrR) *endA1* *nupG*) (*Invitrogen*) are chemically competent cells that allow stable replication of high-copy number plasmids.<sup>96</sup>

**DH5 $\alpha$**  ( $F^-$   $\Phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZYA-argF*) U169 *recA1* *endA1* *hsdR17*( $r_k^-$ ,  $m_k^+$ ) *phoA* *supE44* *thi-1* *gyrA96* *relA1*  $\lambda^-$ ) (*Invitrogen*) are chemically competent cells that allow blue/white screening on plates containing either X-Gal or Bluo-Gal due to the  $\Phi$ 80*lacZ* $\Delta$ M15 genotype. Moreover, the *recA1* mutation helps to reduce the rate of recombination while propagating plasmid DNA, and the *endA1* mutation prevents unspecific digestion of DNA.<sup>96</sup>

#### 3.1.11.2. Strains for protein expression

**BL21 (DE3)** ( $F^-$  *ompT* *hsdS<sub>B</sub>* (*r<sub>B</sub>-m<sub>B</sub>-*) *gal* *dcm* (DE3)) (*Invitrogen*) are chemically competent cells deficient in both lon and ompT proteases. The designation (DE3) indicates that the host is a lysogen of  $\lambda$ DE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter. Such strains are suitable for production of protein from target genes cloned in pET vectors using IPTG as an inducer.<sup>97</sup>

**BL21pLysS (DE3)** ( $F^-$  *ompT* *hsdS<sub>B</sub>* (*r<sub>B</sub>-m<sub>B</sub>-*) *gal* *dcm* (DE3) pLysS (Cam<sup>R</sup>)) (*Invitrogen*) are chemically competent cells used to suppress basal expression of T7 RNA polymerase prior to induction and thus stabilize pET recombinants encoding target proteins that affect cell growth

and viability. The pLysS designation is given to hosts carrying a pET-compatible plasmid that encodes T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase.<sup>97</sup>

## 3.2. Microbiological methods

### 3.2.1. Media for cultivation of *E. coli* cells

Liquid cell cultures of *E. coli* were grown in LB (Luria Bertani) medium (0.5 % NaCl, 0.5 % yeast extract, 1 % tryptone, all *w/v*). Solid media were additionally supplied with 1.5 % (*w/v*) agar. All media were sterilized by autoclaving for 20 min at 120 °C. If necessary, appropriate antibiotics were added prior to use: for cells harboring pET15b and pGEX-6P-1 based vectors, ampicillin was added at a concentration of 100 µg mL<sup>-1</sup>. Cells containing the pETite C-His plasmid were selected using 30 µg mL<sup>-1</sup> kanamycin.

### 3.2.2. Transformation of chemically competent *E. coli* cells

Chemically competent cells (50 µL) were thawed on ice, mixed with 50–300 ng DNA and incubated on ice for 30 min. Afterwards the cells were heat shocked for 1 min at 42 °C and then placed on ice for further 2 min. After addition of 300 µL of LB medium, the cells were shaken at 37 °C for 45–60 min. The cell suspension was plated onto LB-agar plates supplied with appropriate antibiotics and incubated at 37 °C overnight (o/n).

## 3.3. Molecular biology methods

Experiments handling preparation and isolation of DNA were performed according to the instructions from the laboratory handbook J. Sambrook and D. Russell: *Molecular Cloning: A Laboratory Manual*, 3<sup>th</sup> edition, Cold Spring Harbor Laboratory Press (2001),<sup>98</sup> or according to the datasheet instructions of the specific chemical/kit that was used.

### 3.3.1. Amplification and isolation of plasmid DNA from *E. coli*

Isolation of plasmid DNA was performed with *Qiagen's QIAprep Spin Miniprep Kit* according to the manufacturer's manual. To amplify plasmid DNA, One shot TOP10 or DH5α cells were transformed with the appropriate vector (see 3.2.2.) and 5 mL o/n cultures were set up with a single colony.

### 3.3.2 Determination of DNA concentration

Concentration of DNA in solution was determined spectrophotometrically using *Nanodrop 1000* based on the absorbance at 260 nm.

### 3.3.3. Polymerase chain reaction (PCR)

PCR was used for DNA amplification. The proofreading high-fidelity DNA polymerase *Phusion* from *Pyrococcus* was used for amplification of DNA fragments for cloning strategies because of their low error rate. For colony-PCRs, the non-proofreading *Taq* DNA polymerase from *Thermus aquaticus* was used. Standard PCR reactions contained the components listed in Table 14. Annealing temperatures and elongation times were adjusted to primer and template composition and the size of the amplified fragment (Table 15).

**Table 14.** Composition of PCR reactions.

<i>Component</i>	<i>Cloning PCR (Phusion)</i>	<i>Cloning PCR (Taq)</i>
Template DNA (25 ng)	1 µL	1 µL
5' Primer (10 µmol L <sup>-1</sup> )	1 µL	0.25 µL
3' Primer (10 µmol L <sup>-1</sup> )	1 µL	
dNTPs (10 mmol L <sup>-1</sup> )	0.4 µL	0.65 µL
Phusion DNA polymerase (2U* / µL)	0.2 µL	–
5x Phusion HF buffer with MgCl <sub>2</sub>	4 µL	–
Taq polymerase (5U / µL)	–	0.25 µL
10x Taq reaction buffer MgCl <sub>2</sub> (25 mmol L <sup>-1</sup> )	–	2.5 µL 1.5 µL
Nuclease free H <sub>2</sub> O	12.4 µL	18.85 µL
Final volume of:	20 µL	25 µL

\* One unit is defined as the amount of enzyme that will incorporate 10 nmol of dNTP into acid insoluble material in 30 minutes at 74 °C.

**Table 15.** Thermal cycling profile for cloning and colony PCRs.

<i>Component</i>	<i>Cloning PCR (Phusion)</i>	<i>Cloning PCR (Taq)</i>
Initial denaturation	30 s / 98 °C	2 min / 94 °C
Denaturation	5-10 s / 98 °C	15 s / 94 °C
Annealing            35 cycles	10-30 s / primer specific temperature	15 s / primer specific temperature
Extension	15-30 s per kb / 72 °C	1 min per kb / 72 °C
Final extension	5-10 min / 72 °C	5-10 min / 72 °C

### 3.3.4. Agarose gel electrophoresis

DNA samples from PCR reactions were analyzed by agarose gel electrophoresis. Samples were loaded into wells of an agarose gel and subjected to an electric field, causing the negatively charged nucleic acids to move toward the positive electrode. Shorter DNA fragments traveled more rapidly, while the longest fragments remain closest to the origin of the gel, resulting in separation based on size. Supercoiled circular plasmid migrates more rapidly than the same plasmid when linearized whereas the nicked circular plasmid is the slowest in migration in comparison to the other two forms. Agarose gels (1 %, w/v) were prepared with TAE (40 mmol L<sup>-1</sup> Tris-base, 20 mmol L<sup>-1</sup> acetic acid, 1 mmol L<sup>-1</sup> EDTA pH = 8) buffer and with an addition of 50 µg L<sup>-1</sup> ethidium bromide. All samples were mixed with DNA loading dye (0.25 % (w/v) bromophenol blue, 0.25 % (w/v) xylene cyanol FF, 0.01 mol L<sup>-1</sup> EDTA pH = 8, 3 % (w/v) glycerol) prior to loading. Two different dyes are present for visual tracking of DNA migration during electrophoresis. The presence of glycerol ensures that the DNA in the ladder and sample forms a layer at the bottom of the well. The EDTA included in the solution binds divalent metal ions and inhibits metal-dependent nucleases. 1 kb DNA ladder (*NEB*) was applied for size estimation of DNA fragments. Electrophoresis was carried out at a constant voltage of 120 V. Afterwards DNA bands were visualized by UV transillumination at 254 nm and documented with the *MiniBIS Imaging systems (DNR, Bio-Imaging Systems Ltd.)*. Ethidium bromide is a fluorescent dye that intercalates between bases of nucleic acids and allows detection of DNA fragments in gels. Ethidium bromide possesses UV absorbance maxima at 300 and 360 nm. Additionally, it can absorb energy from nucleotides excited by absorbance of 260 nm radiation. Ethidium re-emits this energy as yellow/orange light centered at 590 nm.<sup>99</sup>

### 3.3.5. Cloning and sequencing

#### 3.3.5.1. Standard cloning: restriction and ligation of DNA

pGEX-6P-1 vector was linearized with restriction endonucleases BamHI and XhoI. The gene of interest (*HP1326*) was amplified (see 3.3.3) using primers (see 3.1.9.) that introduced an N-terminal BamHI recognition site and a C-terminal XhoI site to allow directed integration into the vector.

In the ligation reaction (Table 16) the vector was mixed with the gene of the insert and incubated together with T4-DNA ligase for 2 hours at 22 °C. The reaction was quenched by letting the reaction mixture at 65 °C for 10 min. After transformation of the cloned gene into DH5 $\alpha$  cells the colonies were tested for successful integration of the insert by colony PCR. DNA from positive clones was amplified and isolated according to section 3.3.1 and sequenced at the DNA-servis center *Macrogen Inc.* using universal T7 primers (T7 for: 5' TAATACGACTCACTATAG GG 3', T7 rev: 5' GCTAGTTATTGCTCAGCGG 3').

**Table 16.** Ligation reaction mixture.

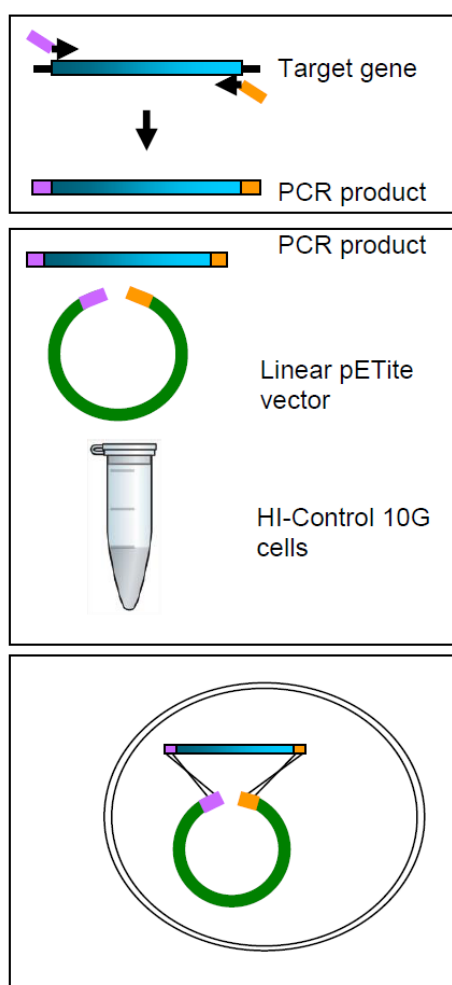
<i>Component</i>	<i>V / <math>\mu</math>L</i>
Digested pGEX (100 ng)	5
<i>HP1326</i> gene (44 ng)	1
T4 ligase (5 U* / $\mu$ L)	0.5
10 x T4 ligase buffer	1
Nuclease free H <sub>2</sub> O	2.5
Final volume:	10

\* One unit is defined as the amount of enzyme required to give 50 % ligation of HindIII fragments of  $\lambda$  DNA (5' DNA termini concentration of 0.12  $\mu$ mol L<sup>-1</sup>, 300-  $\mu$ g/mL) in a total reaction volume of 20  $\mu$ L in 30 minutes at 16 °C in 1X T4 DNA Ligase Reaction Buffer.

Cloning of *HP1026* gene in vector pET15b was done in a collaboration with Dr. A. Danielli and Prof. V. Scarlato from the Department of Biology, Bologna, Italy.

## 3.3.5.2. Enzyme free-cloning

1  $\mu$ L of pre-processed linearized pETite C-His vector (25 ng) was mixed with 3  $\mu$ L of the target PCR product (50 ng of *HP1326* / *HP0907*, see 3.1.10.) and transformed directly into 50  $\mu$ L of chemically competent One shot TOP10 cells (see 3.2.2.). The target gene had been previously amplified with primers that contained short homology to the ends of the pETite C-His vector allowing recombination within the host cell to fuse the blunt PCR product to the vector (Fig. 12) Sequencing of DNA from positive clones was performed like previously described in section 3.3.5.1.



**Figure 12.** Preparation of construct by enzyme free-cloning.<sup>100</sup>



### 3.4. Protein biochemical methods

Experiments handling preparation and isolation of proteins were performed according to the instructions from the laboratory handbook J. Sambrook and D. Russell: *Molecular Cloning: A Laboratory Manual*, 3<sup>th</sup> edition, Cold Spring Harbor Laboratory Press (2001),<sup>98</sup> or according to the datasheet instructions of the specific chemical/ kit that was used.

#### 3.4.1. Protein quantification

The concentration of pure protein solutions was determined spectrophotometrically by reading absorption at 280 nm due to the high absorbance of aromatic amino acid residues. Considering absorbance and the protein specific absorption coefficient that was calculated on the basis of the primary sequence using *ExPASy's ProtParam* software,<sup>101</sup> the protein concentration was determined by applying the Beer-Lambert law.

#### 3.4.2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The quality of the purified proteins samples, samples from expression studies, and protein crystals was verified by discontinuous electrophoresis under denaturing conditions (SDS-PAGE). SDS is an anionic detergent which denatures proteins and confers a negative charge to all protein molecules, since a polypeptide chain binds an amount of SDS proportional to its molecular mass. The final separation of proteins depends entirely on the differences in the relative molecular mass of the polypeptides, so that in the electric field the negatively charged molecules will migrate toward the positively charged anode. For the experiments in this study the used percentage (*w/v*) of stacking and separating acrylamide gels was 4 % and 16 %, respectively (Table 17). Prior to use, all samples were supplied with 4x loading buffer (Table 18) and incubated at 100 °C (to improve denaturation) for 2 min in the case of pure proteins and for 10 min for the samples from expression studies.

**Table 17.** Composition of the stacking and separating acrylamide gels.

<i>Component</i>	<i>Stacking gel</i>	<i>Separating gel</i>
Acrylamide/bis-acrylamide (29:1 w/v) 40 %, (v/v)	4 %	16 %
Buffer Tris-HCl, pH 6.8	0.125 mol L <sup>-1</sup>	–
Buffer Tris-HCl, pH 8.8	–	0.375 mol L <sup>-1</sup>
SDS, (w/v)	0.1 %	0.1 %
APS	0.7 µg mL <sup>-1</sup>	0.7 µg mL <sup>-1</sup>
TEMED, (w/v)	0.05 %	0.05 %

**Table 18.** Composition of the 4x loading buffer.

Tris-HCl pH = 6.8	62.5 mmol L <sup>-1</sup>
β-Mercaptoethanol	12.5 mmol L <sup>-1</sup>
Glycerol, (v/v)	6.25 %
SDS, (w/v)	1.25 %
Bromophenol Blue, (w/v)	0,002 %

Different types of protein markers were used depending on the application. Unstained *Bio-Rad's Precision Plus Protein Standard* was used when gels were stained using the *Coomassie Brilliant Blue R-250* solution (Table 19) or by using the commercial *ProteoSilver Silver Stain Kit* for silver staining. When western blotting was performed after electrophoresis, pre-stained *Bio-Rad's Precision Plus Protein Standard* was used. SDS-PAGE was carried out at constant current (30-50 mA) and maximum voltage (600 V) within 1–2 h. Composition of the buffer used during electrophoresis is presented in Table 20. Gels stained by the Coomassie solution were de-stained with hot water.

**Table 19.** Composition of the *Coomassie Brilliant Blue R-250* staining solution.

Coomassie Brilliant Blue R-250	2.5 g L <sup>-1</sup>
Acetic acid, (v/v)	10 %
Ethanol, (v/v)	45 %

**Table 20.** Composition of the electrophoresis buffer.

Buffer Tris-HCl, pH 8.3	3.03 g L <sup>-1</sup>
Glycin	14.4 g L <sup>-1</sup>
SDS, (w/v)	0.1 %

Silver staining is a very sensitive method to detect very low amounts of protein in a SDS-PAGE gel with a detection limit of 0.1 ng.<sup>102</sup> In this study the silver staining method was used to verify samples of protein crystals. After electrophoresis, the gel was incubated in fixation solution (50 % (v/v) ethanol, 10 % (v/v) acetic acid) for 20 min, and then washed with ethanol (30 %, v/v) for 10 min, and then with ultrapure water for additional 10 min. Afterwards the gel was treated with a sensitizer solution (1 %, v/v ProteoSilver Sensitizer) for another 10 min, and then washed twice for 10 min with ultrapure water. Silver solution (1 %, v/v ProteoSilver Silver Solution) was applied for 10 min followed with washing for 1 min with ultrapure water. The staining process (the reduction of silver ions) was initiated by addition of the developing solution (5 %, v/v ProteoSilver Developer 1, 0.1 %, v/v ProteoSilver Developer 2) for 5 min, and was terminated by addition of ProteoSilver Stop solution for 5 min, and then washed with ultrapure water for 15 min. All used commercial solutions (ProteoSilver Silver Solution, ProteoSilver Sensitizer, ProteoSilver Developer 1, ProteoSilver Developer 2, ProteoSilver Stop Solution) are part of the ProteoSilver Plus Silver Stain Kit (*Sigma Aldrich*).

#### 3.4.3. Western blotting

The western blotting method includes transfer of proteins from a gel to a membrane and their detection on the surface of the membrane by antibodies. In this work it was used to detect presence of His-tagged proteins. After SDS-PAGE, the separated protein samples were transferred from the acrylamide gels to the polyvinylidene fluoride (PVDF) membrane by the semi-dry western blotting method using the instrument *2117 Multiphor II (LKB)*. Before transferring, the gels were incubated with shaking in the transfer buffer (Table 21) for 15 min. All membranes and blotting papers were wetted in the transfer buffer, and the blot sandwich was assembled as follows: three sheets of paper, the transfer membrane, the gel and finally three more sheets of blotting paper were placed inside a semi-dry electro-blotting system with the membrane facing the anode. Protein transfer was performed at a constant current of 1 mA/cm<sup>2</sup> for 1.5 hours. After blotting, the protein transfer was confirmed by staining with the *Ponceau S*

(0.1 % (w/v), 5 % (v/v) acetic acid) solution. Afterwards transfer membranes were further analyzed by immunodetection. At first, membranes were washed with ultrapure water and incubated with 1 % (w/v) casein-TBS solution (Table 21) to saturate free binding sites of the membrane. After 60 min of incubation, membranes were washed three times with TBS buffer; each step lasting 10 min. Primary antibody (see 3.1.9.) was diluted 1,000 times in a 1 % casein-TBS solution and was applied to the membrane o/n at 4 °C. Membranes were washed again three times for 10 min with the TBS buffer. The second antibody (see 3.1.9.) was diluted 10,000 times also in the 1 % casein-TBS solution and applied to the membrane for 60 min. Then, washing with TBS buffer was repeated as before. The bound secondary antibodies were detected via the linked enzyme horseradish peroxidase using *Immun-Star WesternC chemiluminescent kit (Bio-Rad Laboratories)*. The secondary antibody contains horseradish peroxidase, which is an enzyme that catalyzes luminol oxidation in the presence of hydrogen peroxide. The reaction forms an excited state of the substrate that emits light as it decays to the ground state. Luminescent signals derived from the oxidation of luminol were detected and quantified with the luminescent image analyzer *Molecular Imaginig system (Carestream Health, Inc.)*.

**Table 21.** List and composition of buffers used in western blotting.

<i>Transfer buffer</i>	14.4 g L <sup>-1</sup> glycine, 3 g L <sup>-1</sup> Tris-HCl pH = 7.5, 20 % (v/v) ethanol
<i>TBS buffer</i>	25 mmol L <sup>-1</sup> Tris-HCl pH = 7.5, 150 mmol L <sup>-1</sup> NaCl, 0.2 % (v/v) <i>Tween</i>

#### 3.4.4. Proteolytic cleavage

##### 3.4.4.1. Removal of the hexa His-tag from HP1026 with thrombin

Purified His-tagged HP1026 protein was incubated with thrombin (0.1 U / μL) in a protein-protease ratio 1:100, and buffer containing 20 mmol L<sup>-1</sup> Tris-HCl pH 7.5, 150 mmol L<sup>-1</sup> NaCl. The reaction mixture was incubated o/n at 4 °C and gently shaken. The cleaved and uncleaved HP1026 were separated by running the whole mixture through nickel nitrilotriacetic acid (Ni-NTA) column (see 3.6.1.2.) since the cleaved protein does not bind to the resin. Collected flowthrough was then passed through a benzamidine sepharose column (*GE Healthcare*) for

binding of residual thrombin and thus preventing secondary activity from this source. The yield of cleavage was monitored by SDS-PAGE and western blotting.

#### *3.4.4.2. Removal of the GST-tag from GST\_HP1326 with PreScission protease*

Purified GST\_HP1326 protein was treated with PreScission protease to cleave the N-terminal GST tag from the fusion protein. The protein:protease ratio was adjusted to 1:200 in a buffer containing 50 mmol L<sup>-1</sup> Tris-base pH 7.5, 50 mmol L<sup>-1</sup> NaCl, 1 mmol L<sup>-1</sup> EDTA, 1 mmol L<sup>-1</sup> dithiothreitol (DTT). The sample was incubated at 4 °C on a shaker o/n. PreScission protease as well as uncleaved HP1326 contain the GST tag, so they were both removed from the cleaved sample by passing the sample through a glutathione agarose (*GE Healthcare*) column according to the manufacturer protocol (see 3.6.1.3.).

#### *3.4.4.3. Limited proteolysis of the FlgD (HP0907) protein*

Limited proteolysis of FlgD was also investigated. Protein regions exposed on the surface, either present in loops or in unstructured regions, are usually exposed to proteolytic cleavage, while the globular parts and stable secondary structure elements are not accessible to proteases. Each of the four samples of FlgD ( $\gamma = 3 \text{ mg mL}^{-1}$ ) in 20 mmol L<sup>-1</sup> Tris-base pH 7.5, 150 mmol L<sup>-1</sup> NaCl was digested with selected proteases (trypsin,  $\alpha$  chymotrypsin, thermolysin or thrombin) in 1:1000 and 1:500 ratios at 25 °C and incubated for an hour. The yield of digestion was monitored by SDS-PAGE.

### 3.5. Expression of recombinant proteins in *E. coli*

#### 3.5.1. Expression of *Crda* (HP1236), *HP1026* and *FlgD* (HP0907)

*E. coli* BL21(DE3) cells were chemically transformed with a desired plasmid (see Table 13 in 3.1.10.). The used vectors contain either T7 (pET15b, pETite) or *tac* (pGEX-6P-1) promoters that are inducible by isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) and are located upstream from the gene of interest. BL21(DE3) is the T7 host strain used for expression of cloned genes from the bacteriophage T7 promoter. The genome of bacteria harbors the T7 bacteriophage RNA polymerase gene under the control of the IPTG inducible *lacUV5* promoter. The *lac* repressor protein is responsible for maintaining inducible control over the *lacUV5* promoter as well as the T7-*lac* promoter on the vector.<sup>100</sup> In case of protein toxicity BL21pLysS (DE3) cells were used for expression of toxic proteins since these cells provide tighter control of protein expression. BL21pLysS (DE3) cells contain pLysS that constitutively expresses low levels of T7 lysozyme, which reduces basal expression of recombinant genes by inhibiting basal levels of T7 RNA polymerase.

LB medium was inoculated with a single colony from a freshly streaked plate and treated with the appropriate antibiotic for the plasmid and host strain and was shaken (180 rpm) at 37 °C o/n. Fresh LB medium containing the appropriate antibiotic was inoculated with the overnight culture and incubated with shaking at 28 °C or 16 °C. In this work expression performed at different temperatures was tested to establish the best expression condition for different constructs. When the OD<sub>600</sub> of 0.6 was reached, the protein expression was induced by adding 0.5 mM IPTG. After induction of 4 h and o/n performed at 28 °C and 16 °C, respectively, cells were harvested by centrifugation (5000 rpm, 15 min, 4 °C) and flash-frozen in liquid nitrogen and stored at -80 °C. Afterwards, cells were thawed on ice, re-suspended in a protein specific buffer (Table 22), and disrupted with the One Shot Cell breakage system (*Constant System Ltd.*, UK; 1.36 kbar). By standard, the cell suspension was passed through the system twice. Buffers used in cell disruption contain phenylmethylsulfonyl fluoride (PMSF) as a serine protease inhibitor. The lysate was later clarified by centrifugation (18 000 rpm, 30 min, 4 °C) and the soluble fractions were used in further steps of purification.

**Table 22.** Composition of buffers used in cell disruption for a specific protein.

<i>CrdA</i>	<i>GST_CrdA, HP1026, FlgD</i>
20 mmol L <sup>-1</sup> Tris-base pH 7.5	20 mmol L <sup>-1</sup> Tris-base pH 7.5
200 mmol L <sup>-1</sup> NaCl	150 mmol L <sup>-1</sup> NaCl
0.05 % Triton X-100	2 mmol L <sup>-1</sup> PMSF
2 mmol L <sup>-1</sup> (PMSF)	

### 3.5.2. Expression of selenomethionine- *HP1026* and *FlgD* (*HP0907*)

Expression of recombinant selenomethionine derivatized proteins was performed by using the non-methionine auxotrophic *E. coli* BL21 strain (DE3). The bacteria were grown in minimal medium M9 (Table 23) supplemented with vitamin B<sub>1</sub>, amino acids (L-leucine, L-isoleucine, L-valine, L-phenylalanine, L-lysine, L-threonine) and L-seleno-methionine<sup>103</sup> at 28 °C as described in section 3.5.1. All amino acids were purchased from *Sigma Aldrich*.

**Table 23.** Composition of the M9 medium.

Na <sub>2</sub> HPO <sub>4</sub>	6.78 g L <sup>-1</sup>
KH <sub>2</sub> PO <sub>4</sub>	3 g L <sup>-1</sup>
NaCl	0.5 g L <sup>-1</sup>
NH <sub>4</sub> Cl	5 g L <sup>-1</sup>
MgCl <sub>2</sub>	1 mmol L <sup>-1</sup>
CaCl <sub>2</sub>	0.1 mmol L <sup>-1</sup>
Trace elements* (v/v)	0.01 %
D+ Glucose (w/v)	2 %

\* Trace elements: 0.1 g L<sup>-1</sup> MnSO<sub>4</sub>, 0.5 g L<sup>-1</sup> FeCl<sub>3</sub>, 8 mg L<sup>-1</sup> CuCl<sub>2</sub>, 0.05 g L<sup>-1</sup> ZnO, 10 mg L<sup>-1</sup> Co(NO<sub>3</sub>)<sub>2</sub>, 10 mg L<sup>-1</sup> (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>, 5 g L<sup>-1</sup> EDTA

### 3.6. Purification of CrdA, HP1026 and FlgD

Purity of samples during all steps of purification was monitored by the SDS-PAGE.

#### 3.6.1. Affinity chromatography

##### 3.6.1.1. Affinity chromatography using a Strep-Tactin

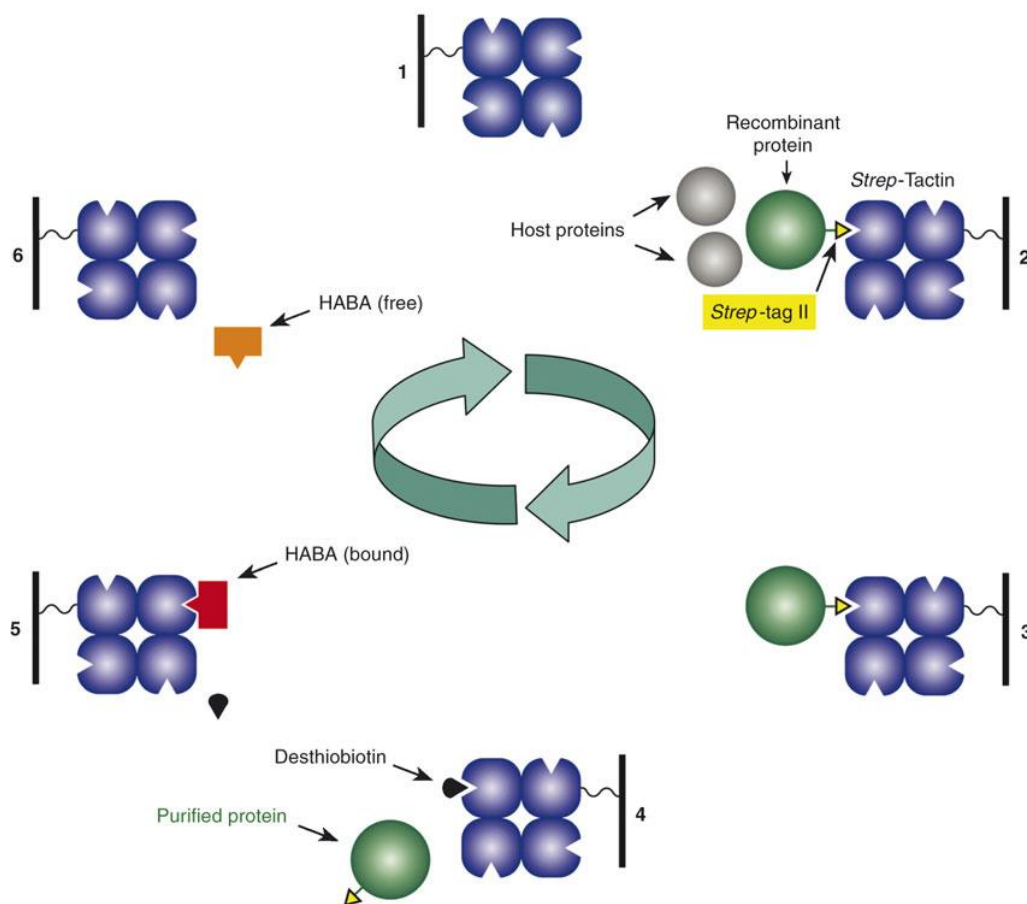
Strep-tagged CrdA was purified by affinity chromatography using a strep-tactin resin. The Strep-tag (NH<sub>2</sub>-WSHPQFEK-COOH) purification system is based on the highly selective interaction between the Strep-tag fused to the protein and a specially engineered streptavidin called strep-tactin. The tagged protein binds to immobilized strep-tactin during affinity purification. Protein is eluted with a buffer containing desthiobiotin, which is the natural ligand of streptavidin (Fig. 13).<sup>104</sup>

The strep-tactin resin (1 mL) was equilibrated with buffer A (Table 24), and the crude lysate was mixed with resin at 4 °C o/n. Unbounded proteins were removed from the resin by washing the resin with 30 mL of buffer A. The Strep\_CrdA protein was eluted from the resin by addition of 20 mL of buffer E (Table 24). Regeneration of the resin was performed by addition of yellow HABA (2-[4'-hydroxy-benzeneazo] benzoic acid) solution. HABA displaced desthiobiotin and turned the resin red due to the binding of HABA as a red colored hydrazine isomer to strep-tactin and prior to the next purification it was removed from the resin by washing it with buffer A.

**Table 24.** Composition of buffers used in affinity chromatography of StrepCrdA.

<i>Buffer A</i>	<i>Buffer E</i>
20 mmol L <sup>-1</sup> Tris-base pH 7.5	20 mmol L <sup>-1</sup> Tris-base pH 7.5
200 mmol L <sup>-1</sup> NaCl	200 mmol L <sup>-1</sup> NaCl
	5 mmol L <sup>-1</sup> desthiobiotin

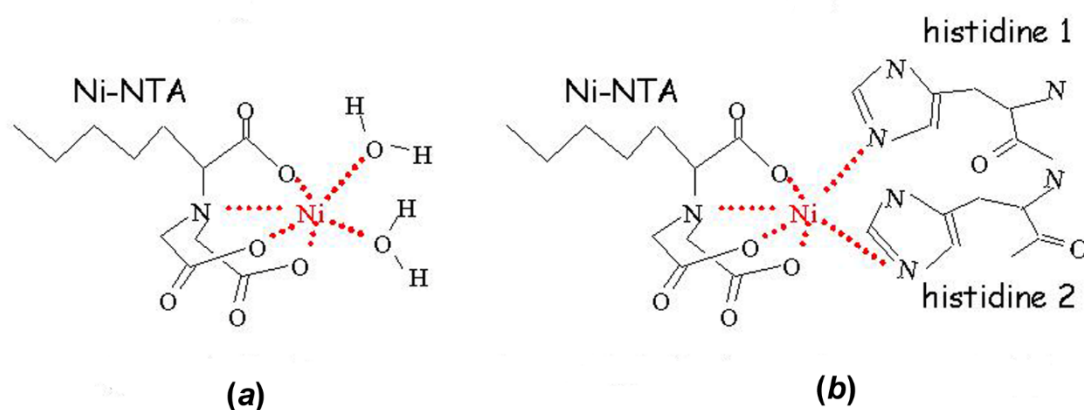




**Figure 13.** Schematic presentation of affinity chromatography with strep-tactin.<sup>104</sup>

#### 3.6.1.2. Affinity chromatography with Ni-NTA agarose

Purification of polyhistidine affinity-tagged proteins can be facilitated by utilization of the nickel-nitrilotriacetic resin. The nitrilotriacetic acid chelator linked to agarose binds His-tagged proteins (5-10 mg His-tag protein / Ni-NTA resin mL) since the nitrogen atoms of the imidazole ring inside the His-tag have affinity toward the nickel ion and occupy two out of six nickel coordination sites (Fig. 14).



**Figure 14.** Ni-NTA chelating (a) two water molecules, or (b) two histidines.<sup>105</sup>

In this study the Ni-NTA affinity purification of proteins (CrDA, HP1026, FlgD) containing the His<sub>6</sub>-tag was performed using the 1 mL columns. The column was firstly equilibrated with protein specific buffer A (Table 25) and then the soluble part of lysate was added to the column. To obtain a high purity sample, the column was connected to an ÄKTA Purifier system, and a washing step of 50 mL of buffer A supplemented with 20 mmol L<sup>-1</sup> imidazole was performed to eliminate unspecifically bound proteins. Finally, elution of the protein was performed in buffer A with a stepwise gradient of imidazole from 20 to 500 mmol L<sup>-1</sup>.

**Table 25.** Composition of buffer A used in Ni-NTA affinity chromatography of CrDA\_His, FlgD\_His and His\_HP1026.

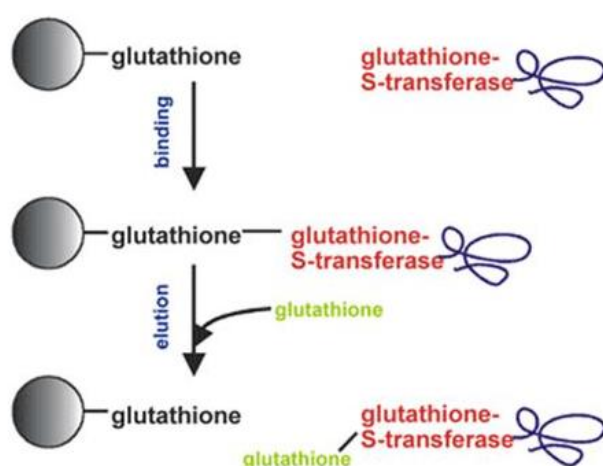
<i>CrDA_His</i>	<i>FlgD_His, His_HP1026</i>
20 mmol L <sup>-1</sup> Tris-base pH 7.5	20 mmol L <sup>-1</sup> Tris-base pH 7.5
200 mmol L <sup>-1</sup> NaCl	150 mmol L <sup>-1</sup> NaCl

### 3.6.1.3. Affinity chromatography with glutathione sepharose

GST-fused proteins were purified using a glutathione sepharose resin. Glutathione S-transferase (GST) is an enzyme found in most organisms and can make the tagged protein more soluble. The purification method is based on the high affinity of GST for glutathione (binding capacity is > 5 mg GST/mL chromatography medium). When applied to the affinity medium, GST-

tagged proteins bind to the glutathione ligand, and impurities are removed by washing with binding buffer. Tagged proteins are then eluted from the chromatography medium using a buffer containing reduced glutathione (Fig. 15).

In this work GST\_CrdA was purified using a glutathione sepharose 1 mL column that was firstly washed with buffer A (Table 26), followed by the incubation of resin with a GST-tagged protein for 1 h at 4 °C. Afterwards, the column was washed with 10 mL of buffer A and protein elution was performed using the buffer E (Table 26).



**Figure 15.** Principle of GST-tag protein purification.<sup>106</sup>

**Table 26.** Composition of buffers used in affinity chromatography of GST\_CrdA.

<i>Buffer A</i>	<i>Buffer E</i>
140 mmol L <sup>-1</sup> NaCl	10 mmol L <sup>-1</sup> reduced glutathione
2.7 mmol L <sup>-1</sup> KCl	50 mmol L <sup>-1</sup> Tris-HCl pH = 8.0
10 mmol L <sup>-1</sup> Na <sub>2</sub> HPO <sub>4</sub>	
1.8 mmol L <sup>-1</sup> KH <sub>2</sub> PO <sub>4</sub> pH = 7.5	

### 3.6.2. Size-exclusion chromatography (SEC)

Analytical size-exclusion chromatography was used to separate proteins according to the different size and to determine homogeneity of a protein sample. In this study the used columns were *Superdex75 10/300* and *Superdex200 10/300* for the separation of proteins in the range  $M_r$

= 3,000-70,000 and  $M_r = 10,000-600,000$ , respectively (*GE Healthcare*). Prior to use, the column was equilibrated at a flow rate of 0.5-1 mL/min with 1.5–2 column volume (CV) of the appropriate buffer (Table 27). Fractions containing the pure protein were pooled and concentrated for crystallization purpose or other structural characterization assays.

To estimate the oligomeric state of protein, calibration SEC was carried out with protein molecular mass standards: vitamin B12 (1.35 kDa), myoglobin (17 kDa), ovalbumin (44 kDa) and  $\gamma$  globulin (158 kDa), and the void volume was determined by using Blue Dextran (*GE Healthcare*).

**Table 27.** Composition of buffers used for SEC.

<i>Crda</i>	<i>FlgD</i>	<i>HP1026</i>
20 mmol L <sup>-1</sup> Tris-base pH 7.5 200 mmol L <sup>-1</sup> NaCl	20 mmol L <sup>-1</sup> Tris-base pH 7.5 150 mmol L <sup>-1</sup> NaCl	20 mmol L <sup>-1</sup> Tris-base pH 7.5 150 mmol L <sup>-1</sup> NaCl
20 mmol L <sup>-1</sup> Tris-base pH 7.5 200 mmol L <sup>-1</sup> NaCl 0.5 mmol L <sup>-1</sup> Cu(OOCCH <sub>3</sub> ) <sub>2</sub>		20 mmol L <sup>-1</sup> Tris-base pH 7.5 120 mmol L <sup>-1</sup> NaCl 30 mmol L <sup>-1</sup> KCl 1 mmol L <sup>-1</sup> MgCl <sub>2</sub>

### 3.6.3. Ultrafiltration

Ultrafiltration was used to concentrate dilute protein solutions or for buffer exchange. Amicon Ultra centrifugal filters were used with a molecular weight cut-off of 3 kDa (for *Crda*), 10 kDa (for *GST\_Crda*) and 30 kDa (for *HP1026*, *FlgD*). This system uses anisotropic semi-permeable membranes to separate macromolecular species and solvents primarily on the basis of size. The filtrate (flow-through) is cleared of macromolecules which are significantly larger than the membrane pores. Desired protein solution was transformed to the adequate filter and centrifugation at 4000 rpm at 4 °C was performed. Molecules smaller than the size of membrane pores were passed through the membrane while the concentration of protein solution increased.

## 3.7. Refolding of the *Crda* protein solubilized from inclusion bodies

After the cell disruption (4 g of fresh cells), lysate-containing CrdA was centrifuged at 18,000 rpm for 30 min at 4 °C to form a solid pellet. The pellet was solubilized by addition of 15 mL of denaturation buffer (6 mol L<sup>-1</sup> guanidinium chloride, 25 mmol L<sup>-1</sup> Hepes pH 7.9, 0.1 mmol L<sup>-1</sup> DTT, 0.2 mmol L<sup>-1</sup> PMSF) and shaken at 20 °C o/n. The sample was then centrifuged at 22,000 rpm and 4 °C for 30 min. The pellet was resuspended in 15 mL of ultrapure water while the supernatant was collected and renaturation by dilution of CrdA was performed. The denatured proteins were diluted hundred to thousand fold with a refolding buffer that did not contain denaturants. In the dilution method, the protein concentrations are also decreased, thus preventing intermolecular aggregation of protein during the procedure.<sup>107</sup> A set of 7 buffers was prepared to test for optimal refolding conditions (Table 28). Refolding buffers R1 to R7 (1.5 mL) were transferred to 2 mL reaction tubes and 100 µL of denatured supernatant were added by slow dilution and incubated for 10 min at room temperature. Afterwards, samples were centrifuged for 15 min at 14,000 rpm at 4 °C and the supernatant was purified by affinity chromatography (see 3.6.1.1. and 3.6.1.2.). Samples from the elution, washing fractions and pellet fraction after denaturation (15 µl) were supplied with 4x loading buffer and analyzed by SDS-PAGE according to 3.4.2. The success of refolding treatment was evaluated by circular dichroism and 1D-NMR spectroscopy.

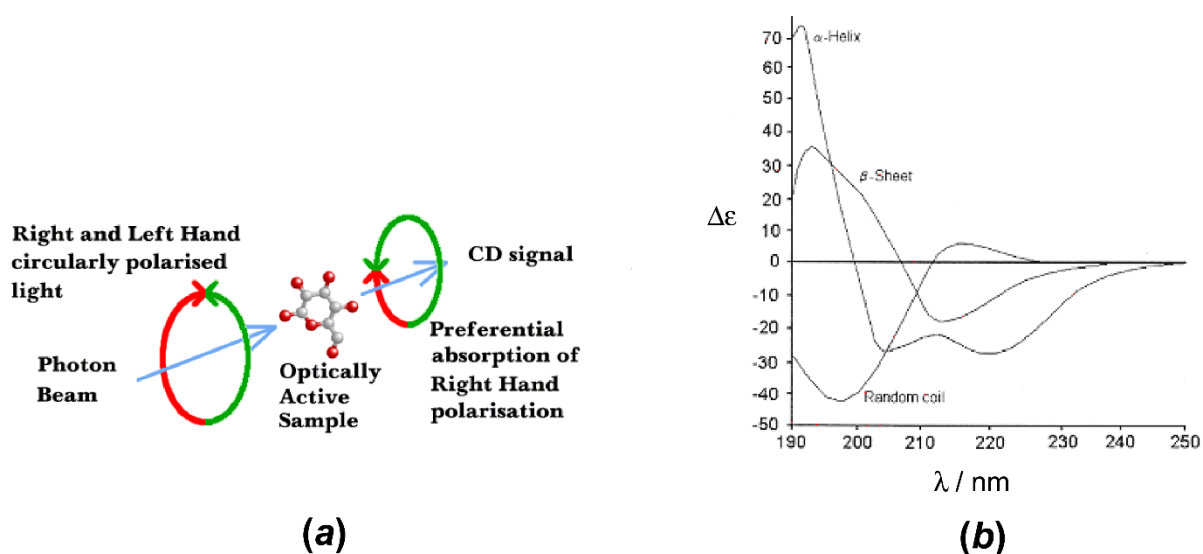
**Table 28.** Composition of re-naturation buffers.

Composition	R1	R2	R3	R4	R5	R6	R7
MES / mmol L <sup>-1</sup> pH 6	50	50	50	–	–	–	–
TRIS / mmol L <sup>-1</sup> pH 8	–	–	–	50	50	50	50
NaCl / mmol L <sup>-1</sup>	9.6	200	240	9.6	9.6	240	240
KCl / mmol L <sup>-1</sup>	50	50	10	0.4	0.4	10	10
MgCl <sub>2</sub> / mmol L <sup>-1</sup>	2	2	–	2	2	–	2
CaCl <sub>2</sub> / mmol L <sup>-1</sup>	2	2	–	2	2	–	2
PEG 3350, w/v	0.05 %	–	0.05 %	–	–	0.05 %	0.05 %
Triton X-100, v/v	–	0.5 %	0.5 %	0.5 %	–	–	0.5 %

### 3.8. Protein characterization

#### 3.8.1. Circular dichroism

Circular dichroism (CD) refers to the differential absorption of the left and right circularly polarized components of circularly polarized light (Fig. 16a). The CD instrument displays the dichroism at a given wavelength of radiation expressed as either the *difference in absorbance* of the two components or as the *ellipticity* in degrees. In biological systems the observed ellipticities are of the order of 10 millidegrees, so a careful attention should be paid for setting the experimental conditions to detect a meaningful information.<sup>108</sup> The chiral arrangement of peptide bonds in different forms of regular secondary structure found in proteins exhibits distinct far-UV CD spectra (Fig. 16b).



**Figure 16.** (a) Principle of CD spectroscopy, (b) Far-UV CD spectra associated with various types of secondary structure components.<sup>109</sup>

In this study, CD in the far-UV region (195–260 nm) was used for secondary structure analysis of proteins and for confirmation of the folding state after the refolding treatment. Far-UV CD spectra were recorded with a *Jasco* spectro-polarimeter in a quartz cuvette with a path length of 0.5 mm and a volume of 70  $\mu\text{L}$ . All measurements were performed with water diluted protein samples ( $c = 0.1 - 0.5 \text{ mg mL}^{-1}$ ) at room temperature. All spectra were corrected for buffer absorption prior to analysis. Spectra were recorded with a step resolution of 1 nm and a bandwidth of 2 nm. The scan speed was set to 50 nm/min and 4 spectra were accumulated. Afterwards, using software CDNN<sup>110</sup> for quantitative secondary structure estimation the observed data expressed as  $\theta$  (in millidegrees) were converted to  $\Delta\epsilon$  by the equation:

$$\Delta\epsilon = \theta \times (\text{MWR} / \gamma \times d) \times 100$$

$\theta$  - observed ellipticity in degrees

MWR- the mean residue molar weight of the protein (molecular mass in Daltons / number of residues)

$\gamma$  - protein concentration in  $\text{mg mL}^{-1}$

$d$  - the optical path length in cm.

CDNN is program which uses CD spectra of a set of proteins for which secondary and tertiary structures are known to be comparable to the CD spectra of the protein being studied.

### 3.8.2. *One-dimensional NMR spectroscopy (1D NMR)*

1D-NMR was used as an additional technique for rough evaluation of the presence of secondary structure elements of the refolded sample of CrdA. This measurement was performed in collaboration with Prof. Massimo Bellanda (University of Padua, Department of Chemistry, Padua).

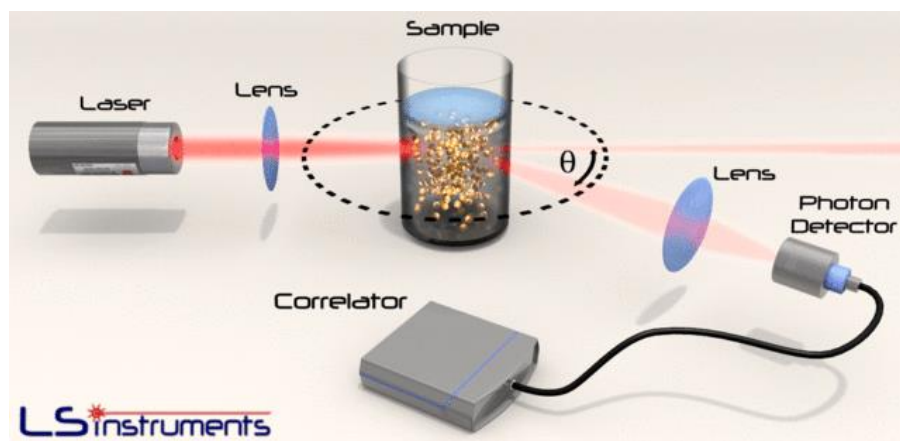
### 3.8.3. *Light scattering*

Dynamic and static light scattering techniques (DLS, SLS) are used in this work to investigate the size distribution, stability and aggregation state of protein molecules in solution. The main difference between DLS and SLS is that the DLS instrument measures at a fixed angle (90° to the incident laser light) and can only determine the mean particle size in a limited size range, while the multi-angle SLS instruments can determine the full particle size distribution.

#### 3.8.3.1. *Dynamic light scattering (DLS)*

The basic principle in a dynamic light scattering experiment (also known as photon correlation spectroscopy or quasi-elastic light scattering) is that the sample is illuminated by a laser beam and the fluctuations of the scattered light are detected at a known scattering angle  $\theta$  by a fast photon detector (Fig. 17). Molecules in solution obey Brownian motion, which means that small molecules move faster than large ones.





**Figure 17.** Schematic representation of the optical setup of a DLS system.<sup>111</sup>

The obtained information during the DLS analysis is about the time scale of the movement of the scatterers. The translational diffusion coefficient  $D_t$  is derived from these data using an auto-correlation function. Since the hydrodynamic radius  $R_h$  is related to the  $D_T$  of the particles  $R_h$  can be calculated using the Stokes-Einstein relation:

$$D_t = k_B T / 6 \pi \eta R_h,$$

where  $k_B$  is Boltzmann's constant ( $1.381 \times 10^{-23} \text{ J K}^{-1}$ ) and  $\eta$  the dynamic viscosity of the solvent.<sup>112</sup>

In this study, DLS measurements of proteins (*Zetasizer Nano ZS, Malvern Instruments Ltd*) were performed to examine the dispersity of samples at concentration of  $5\text{--}10 \text{ mg mL}^{-1}$  at 298 K. Prior to use samples were centrifuged for 10 min at 10,000 rpm, and 40  $\mu\text{L}$  of supernatant was transferred into the quartz ZEN 2112 cuvette.

### 3.8.3.2. Static light scattering (SLS)

SLS is used to determine the weight average molecular weight  $M_w$  of a protein. Since the measurement of the scattering intensity is performed at different angles, it is possible to evaluate the root mean square radius of the macromolecules, also called the radius of gyration  $R_g$ .

In order to examine the oligomeric state of the FlgD protein, the analytical size-exclusion chromatography of the diluted protein ( $1 \text{ mg mL}^{-1}$ ) in combination with multi angle light

scattering (MALS, *Astra 6.1*, *miniDAWN TREOS*) was carried out by using a Superdex 200 10/300 GL (*GE Healthcare*) column equilibrated with buffer containing 20 mmol L<sup>-1</sup> Tris-base pH 7.5, 150 mmol L<sup>-1</sup> NaCl. This experiment was performed at the Division of Biochemistry, Faculty of Science, University of Zagreb.

#### 3.8.4. Thermofluor assay

The influence of different additives, buffers and affinity tag on the protein stability was tested by fluorescence-based thermal stability assay. In this procedure a fluorescence dye interacts with exposed hydrophobic regions of the protein because of their partial or full unfolding upon heat denaturation. The obtained result is a sigmoidal curve that allows calculation of the melting temperature ( $T_m^i$ ).<sup>113</sup>

The thermofluor assay was performed in case of the HP1026 protein using a real-time PCR instrument (*Applied Biosystems*). Each well of a 96-wells plate was filled with 44 µL of protein (1.7 µmol L<sup>-1</sup>) solution, 2 µL of tested additive / buffer (Table 29 and 30) and 4 µL of Sypro orange dye 62× (*Invitrogen*), and sealed with an optic view cover seal. The plate was then centrifuged at 4 °C at 1000 rpm for 30 seconds immediately before the start of the assay to remove possible air bubbles. The measurement was set from 16 to 96 °C with initial stepwise increments of 1 °C. The wavelengths for excitation and emission were 490 and 575 nm, respectively.

**Table 29.** Composition of the Additive screen (Hampton research) used in the thermofluor assay.

1. (A1) 0.1 mol L <sup>-1</sup> Barium chloride dihydrate
2. (A2) 0.1 mol L <sup>-1</sup> Cadmium chloride hydrate
3. (A3) 0.1 mol L <sup>-1</sup> Calcium chloride dihydrate
4. (A4) 0.1 mol L <sup>-1</sup> Cobalt(II) chloride hexahydrate
5. (A5) 0.1 mol L <sup>-1</sup> Copper(II) chloride dihydrate
6. (A6) 0.1 mol L <sup>-1</sup> Magnesium chloride hexahydrate
7. (A7) 0.1 mol L <sup>-1</sup> Manganese(II) chloride tetrahydrate

<sup>i</sup>  $T_m$  is a temperature where the protein is 50 % unfolded.

8. (A8) 0.1 mol L <sup>-1</sup> Strontium chloride hexahydrate
9. (A9) 0.1 mol L <sup>-1</sup> Yttrium(III) chloride hexahydrate
10. (A10) 0.1 mol L <sup>-1</sup> Zinc chloride
11. (A11) 0.1 mol L <sup>-1</sup> Iron(III) chloride hexahydrate
12. (A12) 0.1 mol L <sup>-1</sup> Nickel(II) chloride hexahydrate
13. (B1) 0.1 mol L <sup>-1</sup> Chromium(III) chloride hexahydrate
14. (B2) 0.1 mol L <sup>-1</sup> Praseodymium(III) acetate hydrate
15. (B3) 1.0 mol L <sup>-1</sup> Ammonium sulfate
16. (B4) 1.0 mol L <sup>-1</sup> Potassium chloride
17. (B5) 1.0 mol L <sup>-1</sup> Lithium chloride
18. (B6) 2.0 mol L <sup>-1</sup> Sodium chloride
19. (B7) 0.5 mol L <sup>-1</sup> Sodium fluoride
20. (B8) 1.0 mol L <sup>-1</sup> Sodium iodide
21. (B9) 2.0 mol L <sup>-1</sup> Sodium thiocyanate
22. (B10) 1.0 mol L <sup>-1</sup> Potassium sodium tartrate tetrahydrate
23. (B11) 1.0 mol L <sup>-1</sup> Sodium citrate tribasic dihydrate
24. (B12) 1.0 mol L <sup>-1</sup> Cesium chloride
25. (C1) 1.0 mol L <sup>-1</sup> Sodium malonate pH 7.0
26. (C2) 0.1 mol L <sup>-1</sup> L-Proline
27. (C3) 0.1 mol L <sup>-1</sup> Phenol
28. (C4) 30% v/v Dimethyl sulfoxide
29. (C5) 0.1 mol L <sup>-1</sup> Sodium bromide
30. (C6) 30% w/v 6-Aminohexanoic acid
31. (C7) 30% w/v 1,5-Diaminopentane dihydrochloride
32. (C8) 30% w/v 1,6-Diaminohexane
33. (C9) 30% w/v 1,8-Diaminooctane
34. (C10) 1.0 mol L <sup>-1</sup> Glycine
35. (C11) 0.3 mol L <sup>-1</sup> Glycyl-glycyl-glycine
36. (C12) 0.1 mol L <sup>-1</sup> Taurine
37. (D1) 0.1 mol L <sup>-1</sup> Betaine hydrochloride
38. (D2) 0.1 mol L <sup>-1</sup> Spermidine
39. (D3) 0.1 mol L <sup>-1</sup> Spermine tetrahydrochloride
40. (D4) 0.1 mol L <sup>-1</sup> Hexamine cobalt(III) chloride
41. (D5) 0.1 mol L <sup>-1</sup> Sarcosine
42. (D6) 0.1 mol L <sup>-1</sup> Trimethylamine hydrochloride
43. (D7) 1.0 mol L <sup>-1</sup> Guanidine hydrochloride
44. (D8) 0.1 mol L <sup>-1</sup> Urea
45. (D9) 0.1 mol L <sup>-1</sup> β-Nicotinamide adenine dinucleotide hydrate
46. (D10) 0.1 mol L <sup>-1</sup> Adenosine-5'-triphosphate disodium salt hydrate
47. (D11) 0.1 mol L <sup>-1</sup> TCEP hydrochloride
48. (D12) 0.01 mol L <sup>-1</sup> GSH (L-Glutathione reduced), 0.01 mol L <sup>-1</sup> GSSG (L-Glutathione oxidized)
49. (E1) 0.1 mol L <sup>-1</sup> Ethylenediaminetetraacetic acid disodium salt dihydrate
50. (E2) 5% w/v Polyvinylpyrrolidone K15
51. (E3) 30% w/v Dextran sulfate sodium salt
52. (E4) 40% v/v Pentaerythritol ethoxylate (3/4 EO/OH)
53. (E5) 10% w/v Polyethylene glycol 3,350
54. (E6) 30% w/v D-(+)-Glucose monohydrate
55. (E7) 30% w/v Sucrose
56. (E8) 30% w/v Xylitol
57. (E9) 30% w/v D -Sorbitol

58. (E10) 12% w/v myo-Inositol
59. (E11) 30% w/v D -(+)-Trehalose dihydrate
60. (E12) 30% w/v D -(+)-Galactose
61. (F1) 30% v/v Ethylene glycol
62. (F2) 30% v/v Glycerol
63. (F3) 3.0 mol L <sup>-1</sup> NDSB-195
64. (F4) 2.0 mol L <sup>-1</sup> NDSB-201
65. (F5) 2.0 mol L <sup>-1</sup> NDSB-211
66. (F6) 2.0 mol L <sup>-1</sup> NDSB-221
67. (F7) 1.0 mol L <sup>-1</sup> NDSB-256
68. (F8) 0.15 mmol L <sup>-1</sup> CYMAL®-7
69. (F9) 20% w/v Benzamidine hydrochloride
70. (F10) 5% w/v n-Dodecyl-N,N-dimethylamine-N-oxide
71. (F11) 5% w/v n-Octyl-β- D -glucoside
72. (F12) 5% w/v n-Dodecyl-β- D -maltoside
73. (G1) 30% w/v Trimethylamine N-oxide dihydrate
74. (G2) 30% w/v 1,6-Hexanediol
75. (G3) 30% v/v (+/-)-2-Methyl-2,4-pentanediol
76. (G4) 50% v/v Polyethylene glycol 400
77. (G5) 50% v/v Jeffamine M-600 pH 7.0
78. (G6) 40% v/v 2,5-Hexanediol
79. (G7) 40% v/v (±)-1,3-Butanediol
80. (G8) 40% v/v Polypropylene glycol P 400
81. (G9) 30% v/v 1,4-Dioxane
82. (G10) 30% v/v Ethanol
83. (G11) 30% v/v 2-Propanol
84. (G12) 30% v/v Methanol
85. (H1) 10% v/v 1,2-Butanediol
86. (H2) 40% v/v tert-Butanol
87. (H3) 40% v/v 1,3-Propanediol
88. (H4) 40% v/v Acetonitrile
89. (H5) 40% v/v Formamide
90. (H6) 40% v/v 1-Propanol
91. (H7) 5% v/v Ethyl acetate
92. (H8) 40% v/v Acetone
93. (H9) 0.25% v/v Dichloromethane
94. (H10) 7% v/v 1-Butanol
95. (H11) 40% v/v 2,2,2-Trifluoroethanol
96. (H12) 40% v/v 1,1,1,3,3,3-Hexafluoro-2-propanol

**Table 30.** List of additives (not present in Additive screen) and buffers used in the thermofluor assay.

500 $\mu\text{mol L}^{-1}$ Adenosine-5'-( $\gamma$ -thio)-triphosphate tetralithium salt
500 $\mu\text{mol L}^{-1}$ ADP Adenosine-5'-diphosphate sodium salt
500 $\mu\text{mol L}^{-1}$ GTP Guanosine-5'-triphosphate sodium salt hydrate
500 $\mu\text{mol L}^{-1}$ GDP Guanosine -5'-diphosphate sodium salt
1 $\text{mol L}^{-1}$ MES pH 6.5
1 $\text{mol L}^{-1}$ HEPES pH 7
1 $\text{mol L}^{-1}$ TRIS pH 7.5
1 $\text{mol L}^{-1}$ TRIS pH 8
1 $\text{mol L}^{-1}$ TRIS pH 8.5
1 $\text{mol L}^{-1}$ TRIS pH 8.8
1 $\text{mol L}^{-1}$ TRIS pH 9.5

### 3.8.5. Mass spectrometry

Mass spectrometry was used to monitor the level of degradation of FlgD\_26695 in solution and also the crystallized FlgD\_26695. The sample from the crystal of the tetragonal form of FlgD\_26695 was prepared by dissolving the crystal in the SDS-PAGE loading buffer. This sample together with full-length FlgD\_26695 was checked by SDS-PAGE, and the bands obtained from the crystallized sample and full-length FlgD\_26695 were isolated and in gel digested with trypsin. The fractions of the extracted peptides were dried out, dissolved in 50 % acetonitrile, supplemented with 0.1 % formic acid and directly injected in the nano-ESI source. Mass measurements were performed with a quadrupole-TOF spectrometer (Waters, Manchester, UK) (capillary voltage: 2800 – 3000 V; cone voltage: 45 V; scan time: 1 s; interscan: 0.1 s). Analyses of the spectra were performed by using the MASSLYNX software (Micromass, Wythenshew, UK). The mass of the FlgD\_G27 monomer was determined by mass analysis of the peaks isolated by reverse phase chromatography (C4-column, RP-HPLC). All mass experiments were performed in collaboration with Prof. P. Polverino de Laureto from the Department of Pharmaceutical Sciences, University of Padua.

### 3.8.6. ATPase activity assay by reversed-phase high-performance liquid chromatography (RP-HPLC)

ATPase activity of the HP1026 protein was tested by utilizing RP-HPLC with a detection limit of 0.05 nmol of ADP hydrolyzed from ATP. After termination of the ATPase reaction the product of enzymatic hydrolysis ADP and substrate ATP were separated by RP-HPLC. The amount of ADP was quantified from the chromatographic peak by measuring the absorbance of ADP at 259 nm.<sup>114</sup>

To examine the activity of purified enzyme a master mix reaction (260  $\mu\text{L}$ ) was prepared at a concentration of 30  $\mu\text{mol L}^{-1}$  HP1026 in buffer containing 20  $\text{mmol L}^{-1}$  Tris-base pH 7.5, 150  $\text{mmol L}^{-1}$  NaCl and 10  $\text{mmol L}^{-1}$   $\text{MgCl}_2$  and incubated for 1 min at 30 °C. Afterwards ATP was added in a concentration of 1  $\text{mmol L}^{-1}$ . The reaction mixture was performed at 30 °C and 40- $\mu\text{L}$  aliquots were taken in a time-dependent manner. The ATP hydrolysis reaction was quenched by incubation of the reaction mixture at 100 °C for 2 min and centrifuged at 14,000 rpm for 15 min. A 20- $\mu\text{L}$  aliquot of the clear supernatant was applied to a RP-HPLC column (*Synergi MAX-RP 80A*, 150  $\times$  4.6 mm, 4  $\mu\text{m}$ ; *Phenomenex*) equilibrated with the buffer of 50  $\text{mmol L}^{-1}$   $\text{Na}_2\text{HPO}_4$  pH 6, 10  $\text{mmol L}^{-1}$  tetra-*n*-butylammonium bromide and 4 % (v/v) acetonitrile at a flow rate of 0.8 mL/min. The tetra-*n*-butylammonium bromide was added to the buffer for separation of ADP and ATP that are hydrophilic molecules. The absorbance of ADP was monitored at 254.4 nm using a UV detector. Prior to running the reaction mixture, runs of pure ADP and ATP solutions (in water and in a buffer used in the activity assay) at a concentration of 1  $\text{mmol L}^{-1}$  was performed. The same experiment was performed only by changing the concentration of substrate ATP in an incubation time of 15 min at 30 °C. The obtained results were evaluated using a nonlinear regression (applying Michaelis-Menten model, performed using GraphPad Prism)<sup>115</sup> for the calculation of kinetic parameters ( $K_m^{\text{ii}}$ ,  $v_{\text{max}}^{\text{iii}}$ ).

---

<sup>ii</sup>  $K_m$  is the substrate concentration needed to achieve a half-maximum enzyme velocity.

<sup>iii</sup>  $V_{\text{max}}$  is the velocity of the enzyme extrapolated to very high concentrations of substrate.

### 3.8.7. Ultraviolet-visible (UV-VIS) titration

UV-VIS titration was performed to investigate the affinity of the CrdA protein toward copper(II) ions. Before performing the titration experiment CrdA was incubated with 10 mmol L<sup>-1</sup> EDTA to remove any possibly bound metal ions, and then the buffer was exchanged to the buffer containing 30 mmol L<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub> pH 7 and 50 mmol L<sup>-1</sup> NaCl. Copper titrations were performed on an Agilent 8453 *UV-vis Spectroscopy System* at room temperature in the wavelength range 200–800 nm. 1 mL of CrdA (150 μmol L<sup>-1</sup>) was titrated with CuSO<sub>4</sub> using a stock solution at a concentration of 10 mmol L<sup>-1</sup>. All spectra were blanked against the used buffer. Reaction was incubated for 5 min after every new addition of titrant before measuring the absorbance.

### 3.8.8. Crystallographic studies

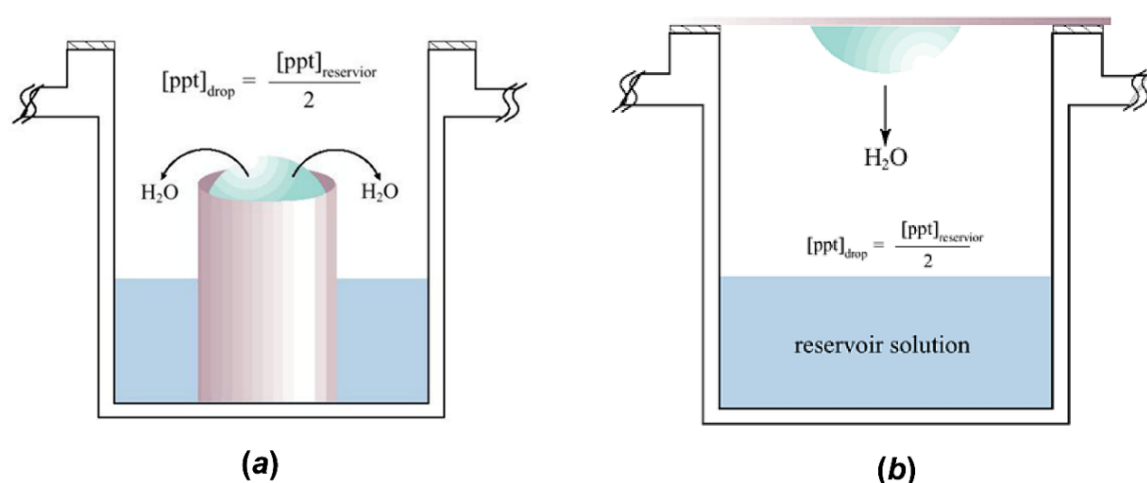
X-ray diffraction on a single crystal was used for the crystal structure determination.

#### 3.8.8.1. Crystallization

*The Solubility Tool Kit (Molecular Dimension)* was used to get a more detailed information on the solubility profile of proteins and to determine the suitable concentration of proteins for crystallization trials. The strategy comprises testing solubility of proteins by changing the concentration and type of salt at constant pH values of 4.5 and 9.0. This allows initial solubility screening in salt solutions, since most proteins will carry a net positive and net negative charge at pH 4.5 and pH 9.0, respectively. Another approach includes testing the ability for crystal growth at a pH close to the pI in addition of organic precipitants, such as polyethylene glycols (PEGs) and MPD (2-methyl-2,4-pentanediol) at different ratios since ionic strength is poor at this pH value because the net charge of the protein is near zero.

Initial crystallization trials were carried out with different 96-well screens from *Qiagen* or *Molecular Dimensions*. All initial screens were prepared by an Oryx 8 robot (*Douglas Instruments*) using the sitting drop vapor diffusion technique (Fig. 18a) in the MRC 96-well sitting drop plate (*Molecular Dimensions*). In this technique, water evaporates from the protein

containing droplet since the precipitant concentration is lower than in the reservoir. The system will equilibrate creating a supersaturated condition in the drop therefore increasing protein and precipitant concentrations and the chance of crystal growth.<sup>116</sup> Each drop was prepared by mixing equal volumes of protein and precipitant solutions (500 nL), and equilibrated against 75  $\mu\text{L}$  of the mother liquid solution. Plates were sealed with *Crystal Clear Sealing Tape* from *Hampton Research* and stored at a temperature of choice (20 °C and / or 4 °C).



**Figure 18.** Diffusion experiment setup for: (a) sitting drop, (b) hanging drop.<sup>116</sup>

After the hit conditions containing crystals were obtained, they were optimized by varying the drop size, volume ratio between the precipitant and protein, and ratio of components of the precipitant solution. For the optimization purpose, the hanging drop vapor diffusion technique was also used in a 24-well XRL plates (Fig. 18b). Additionally, the influence of 96 different additives (*Additive Screen*) on the initial screening solution in a ratio 1:9 was explored. In case of the HP1026, protein microseeding was applied as well. Crystals obtained from a sitting drop experiment were transferred to a 1.5 mL reaction tube prepared with 50  $\mu\text{L}$  of precipitant solution and a PTFE bead. The crystals were sonicated twice for one minute. This stock solution of seeds was used for further sitting drop trials in a set up with 1  $\mu\text{L}$  drops and the final ratio of seeding solution of 10 % (v/v).

Before the diffraction experiment, potential protein crystals were visualized by using the UV fluorescence microscope for differentiation of protein from salt crystals. Prior to data collection, crystals were directly cooled in liquid nitrogen or, if needed, frozen with a cryoprotectant.



Diffraction quality of crystals was tested at the ID14-4 beamline of the *European Synchrotron Radiation Facility* (ESRF, Grenoble, France) or at the PXIII beamline of the *Swiss Synchrotron Light Source* (SLS, Villigen, Switzerland).

### 3.8.8.2. Data processing and structure determination

Diffraction data of the native monoclinic form of FlgD (FlgD\_m) were measured at the ID14-4 beamline of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Crystals of both native and seleno-methionine tetragonal form of FlgD (FlgD\_t) were measured at the PXIII beamline of the Swiss Synchrotron Light Source (SLS, Villigen, Switzerland). In case of SeMet\_FlgD\_t a fluorescence emission scan at the selenium edge was performed to optimize the energy for anomalous measurements. Diffraction data of FlgD\_m and FlgD\_t were indexed and integrated with software iMOSFLM<sup>117</sup> and XDS<sup>118</sup>, respectively. All datasets were merged and scaled with Scala<sup>119</sup>, contained in the CCP4 crystallographic package.<sup>120</sup> The tetragonal structure was solved by MAD data using the Shelx software.<sup>121</sup> Initial phases were submitted to the *Buccaneer* software<sup>122</sup> for density modifications and model building. The obtained model was used for refinement against the native data set at higher resolution with Phenix,<sup>123</sup> and checked and manually adjusted with the graphic software Coot.<sup>124</sup> The monoclinic form was solved by Molrep<sup>125</sup> starting from the refined tetragonal structure as the template. During refinement of FlgD\_m non-crystallographic symmetry restraints were imposed under the REFMAC5 program.<sup>126</sup>

### 3.8.9. Bioinformatics methods and software tools

Sequence similarity searches were performed with *BLAST*<sup>127</sup> algorithm accessing databases *UniProtKB*<sup>128</sup> and *NCBI's* non-redundant protein sequence database. For multiple alignments of protein sequences *ClustalW2*<sup>129</sup> or *Clustal Omega*<sup>130</sup> were used. Translation of nucleotide sequences was performed with *Translate*. For computing physicochemical properties of a protein, the program *ProtParam* was used. Potential cleavage sites for proteases were identified with *PeptideCutter*. All last three programs are part of *ExPASy's*<sup>101</sup> proteomics tools. For the prediction of protein secondary structure elements, *PSIPRED*<sup>131</sup> software was used. Prediction of signal peptides was done by using the *LipoP*<sup>132</sup> software. Molecular graphics images of protein structures deposited at the protein database *RCSB PDB* ([www.pdb.org](http://www.pdb.org))<sup>133</sup> were

produced with the *CCP4MG*<sup>47</sup> or *Pymol*<sup>134</sup> software. PISA<sup>135</sup> software was used to calculate crystal packing contacts.



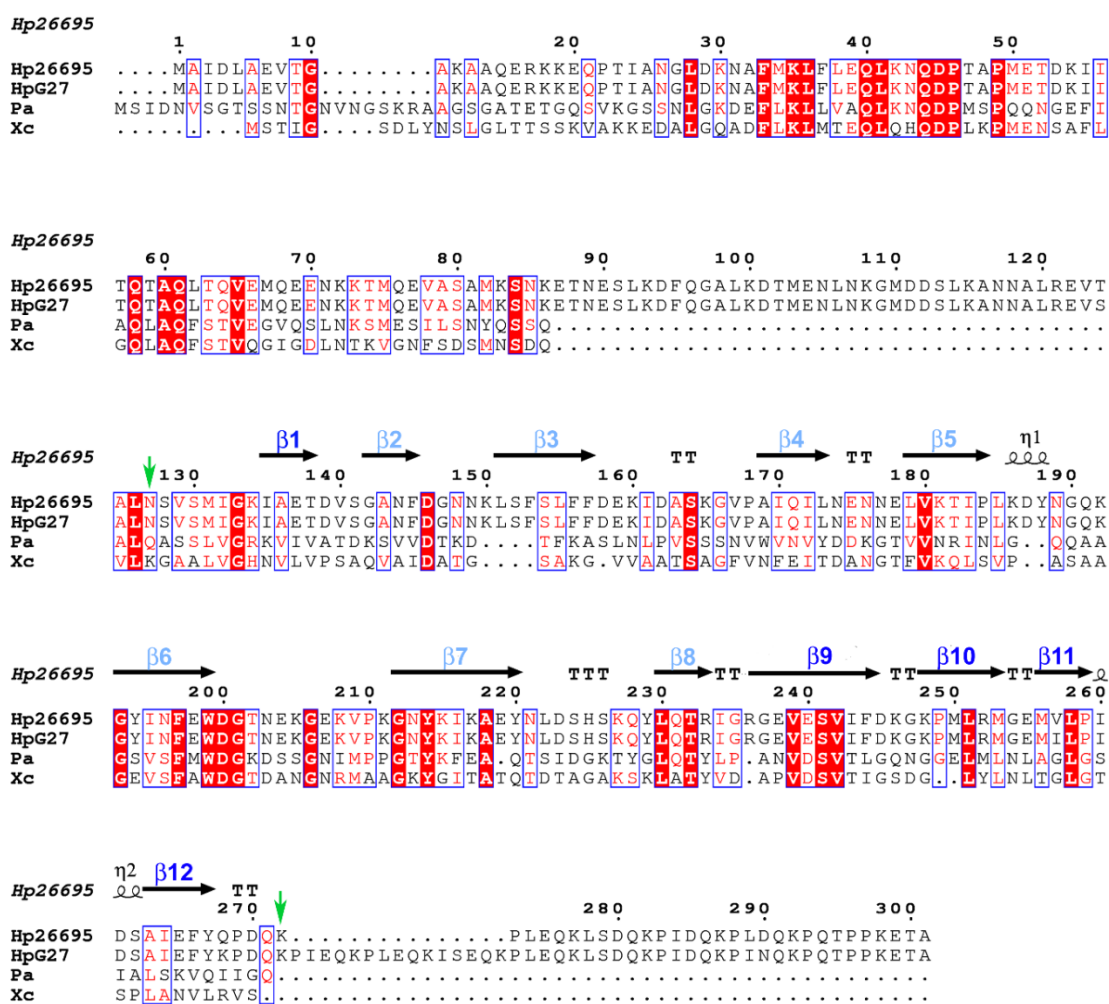
## § 4. RESULTS AND DISCUSSION

### 4.1. FlgD protein from *H. pylori* (HpFlgD)

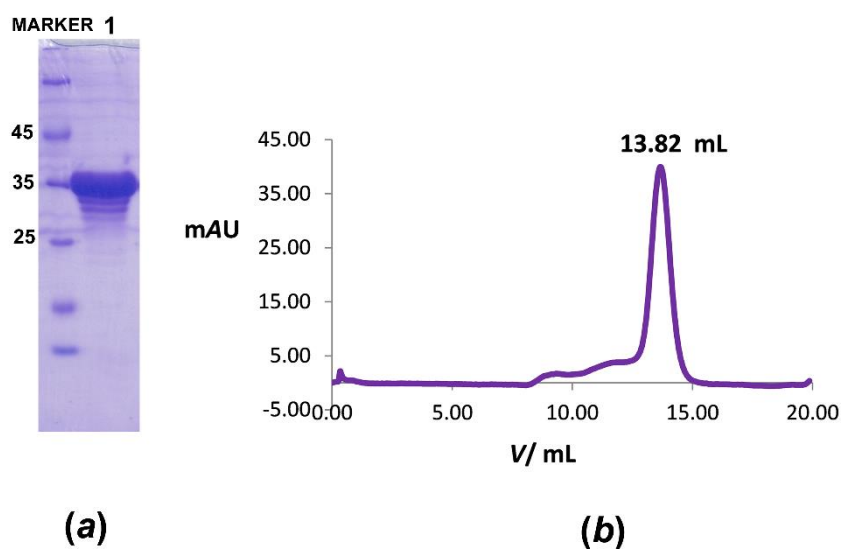
#### 4.1.1. Cloning, expression and purification

The HP0907 gene was cloned from two different *H. pylori* strains, 26695 and G27. Cloning from different strains was performed since the *HpFlgD* sequence analysis by the *PSIPRED* bioinformatic tool (<http://bioinf.cs.ucl.ac.uk/psipred/>) predicted 15 extra disordered amino acids (Pro273 to Lys287) of the protein from strain G27 in respect to strain 26695 (Fig. 19). This approach was used to gain a more successful crystallization outcome.

Recombinant *HpFlgD* was successfully expressed and purified to homogeneity by Ni-NTA chromatography and size-exclusion chromatography (Fig. 20). Purified native *HpFlgD*\_26695 and *HpFlgD*\_G27 were concentrated to 17.5 mg mL<sup>-1</sup> and 30 mg mL<sup>-1</sup>, respectively, and used in the crystallization experiment.



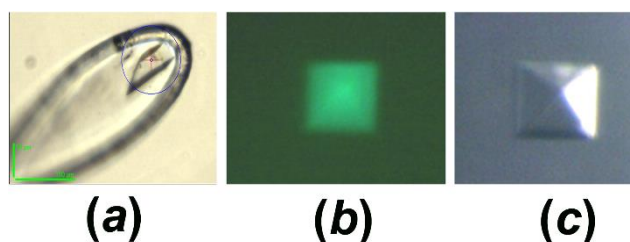
**Figure 19.** Sequence alignment of *HpFlgD\_26695* (HP, O25565\_HELPHY), *HpFlgD\_G27* (HP, B5Z7R3\_HELPG), *PaFlgD* (PA, Q9I4Q0\_PSEAE) and *XcFlgD* (XC, Q8P9B5\_XANCP) using CLUSTAL W2.<sup>129</sup> The alignment figure was generated using ESPript 3.<sup>136</sup> The numbering system is based on the *HpFlgD\_26695* sequence. Residues conserved in all four sequences are labelled in white and highlighted in red, while similar residues are labelled in red. Labels of the secondary structure elements ( $\beta$  sheets) found in the crystal structure of *HpFlgD\_t* are presented in light blue (Fn-III domain) and dark blue (tudor domain) while the starting and ending residue found in the crystal structure of *HpFlgD\_t* is marked with a green arrow.



**Figure 20.** (a) SDS-PAGE; Lane 1 – full-length *HpFlgD\_26695*. (b) Analytical gel filtration chromatogram of *HpFlgD\_26695* (purple curve; molecular mass of the monomer ~34.5 kDa) which elutes as a single peak at the volume of 13.82 mL corresponding to the molecular mass of 77.19 kDa.

#### 4.1.2. Crystallization and data collection

Good quality crystals were grown by the sitting drop vapor diffusion technique. 1  $\mu\text{l}$  drops were prepared by mixing the protein and precipitant solutions in a ratio 1:1, and equilibrated against 75  $\mu\text{l}$  of the mother liquor in the reservoir. Two different crystal forms were obtained after approximately 2 months (Fig. 21). Monoclinic crystals of *HpFlgD\_G27* (*HpFlgD\_m*) were grown from the screening solution 4 of the PACT suite (Qiagen) (0.1 mol L<sup>-1</sup> SPG pH 7.0, 25 % PEG 1500) at 20 °C, while tetragonal crystals (*HpFlgD\_t*) of the construct *HpFlgD\_26695* were obtained from the screening solution 79 of the same screening kit (0.2 mol L<sup>-1</sup> sodium acetate, 0.1 mol L<sup>-1</sup> Bis-tris propane pH 7.5, 20 % PEG 3350) at 4 °C. In order to solve the phase problem, seleno-methionine derivatized *HpFlgD\_26695* was crystallized (concentration of 15 mg mL<sup>-1</sup>) at 277 K from solution 4 of the PACT suite. Monoclinic crystals were directly cooled in liquid nitrogen, whilst the tetragonal crystals were cryoprotected by using the precipitant solution supplemented with 20 % of ethylene glycol.



**Figure 21.** (a) Monoclinic crystal of native *HpFlgD\_G27* and (b, c) tetragonal crystal of native *HpFlgD\_26695*. Picture (b) was captured under the microscope using a fluorescence excitation filter (CWL / BW = 500 / 50 nm).

Diffraction data of native *HpFlgD\_m* were measured at the ID14-4 beamline of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and they belong to the *P2* space group (Table 31). Crystals of both native and seleno-methionine *HpFlgD\_t* belong to the tetragonal space group *I422* (Table 31). Diffraction data of this second crystal form were measured at the PXIII beamline of the Swiss Synchrotron Light Source (SLS, Villigen, Switzerland). A fluorescence emission scan at the selenium edge was performed to optimize the energy for anomalous measurements. Diffraction data of *HpFlgD\_m* and *HpFlgD\_t* were indexed and integrated with software *iMOSFLM*<sup>117</sup> and *XDS*,<sup>118</sup> respectively. All datasets were merged and scaled with *Scala*<sup>119</sup>, contained in the CCP4 crystallographic package.<sup>120</sup> Molecular replacement was attempted by using different models that were built upon the template structures of the orthologues from other species. This method failed due to significant differences in the overall structure. The tetragonal structure was solved by MAD data using the *Shelx* software.<sup>121</sup> Initial phases were submitted to the *Buccaneer* software<sup>122</sup> for density modifications and model building. The obtained model was used for refinement against the native data set at higher resolution (2.17 Å) with *Phenix*<sup>123</sup>, and checked as well as manually adjusted with the graphic software *Coot*<sup>124</sup>. The monoclinic form was solved by *Molrep*<sup>125</sup> starting from the refined tetragonal structure as the template. During refinement of *HpFlgD\_m* non-crystallographic symmetry restraints were imposed under the *REFMAC5* program.<sup>126</sup> The final statistics on data collection and refinement are summarized in Tables 31 and 32.

**Table 31.** Statistics of data collection and processing. Values for the outer shell are given in parentheses. 1800 frames of  $0.1^\circ$  were collected. One crystal was used to collect diffraction data for MAD anomalous phasing.

		Native <i>HpFlgD_m</i>	MAD data SeMet <i>HpFlgD_t</i>			Native <i>HpFlgD_t</i>
			Peak	Inflection	Remote	
Crystal system		Monoclinic		Tetragonal		Tetragonal
Space group		<i>P2</i>		<i>I422</i>		<i>I422</i>
Wavelength / Å		0.95372	0.97939	0.97957	0.97171	0.99988
Unit cell parameters / Å, °		<i>a</i> = 77.42 <i>b</i> = 34.22 <i>c</i> = 131.64 $\beta$ = 99.6	<i>a</i> = 76.42 <i>c</i> = 144.98	<i>a</i> = 76.49 <i>c</i> = 145.05	<i>a</i> = 76.47 <i>c</i> = 145.03	<i>a</i> = 76.42 <i>c</i> = 145.43
Z	unit cell	8	16	16	16	16
	asymmetric unit	4	1	1	1	1
Resolution range / Å		54.09 – 2.75 (2.88 – 2.75)	43.33 – 2.78 (2.9 – 2.78)	43.36 – 2.74 (2.89 – 2.74)	43.35 – 2.73 (2.88 – 2.73)	43.37 – 2.17 (2.28 – 2.17)
$R_{\text{merge}}$		0.077 (0.551)	0.099 (1.376)	0.112 (2.154)	0.100 (1.590)	0.059 (0.666)
$R_{\text{pim}}$		0.047 (0.345)	0.047 (0.586)	0.047 (0.955)	0.045 (0.715)	0.025 (0.332)
$\langle I/\sigma(I) \rangle$		14.4 (2.8)	15.1 (1.4)	14.5 (0.8)	15.5 (1.0)	20.5 (2.1)
CC half		0.991 (0.742)	0.998 (0.658)	0.998 (0.489)	0.998 (0.465)	0.999 (0.746)
Completeness / %		99.9 (99.5)	96.4 (78.6)	95.6 (72.7)	94.9 (68.2)	98.5 (90.0)
Anomalous Completeness / %			95.3 (72.3)	94.0 (62.2)	93.3 (58.6)	
Redundancy		3.6 (3.5)	6.9 (6.0)	6.9 (5.3)	6.9 (5.1)	6.9 (5.5)
Unique reflections		18375 (1808)	5489 (632)	5675 (596)	5704 (573)	11714 (1019)

**Table 32.** Refinement statistics.

	Native <i>HpFlgD_m</i>	Native <i>HpFlgD_t</i>
$R_{\text{work}} / R_{\text{free}}$	0.221/0.275	0.197/0.234
No. protein / solvent atoms	4628/44	1156/74
Mean <i>B</i> (Å <sup>2</sup> )	62.1	50.5
<b>R.m.s.d. from ideal value</b>		
Bond length (Å)	0.011	0.008
Bond angles (°)	1.49	1.11
<b>Geometry</b>		
Ramachandran favored (%)	97.0	95.14
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	1.8	0
Overall score	1.32	1.70

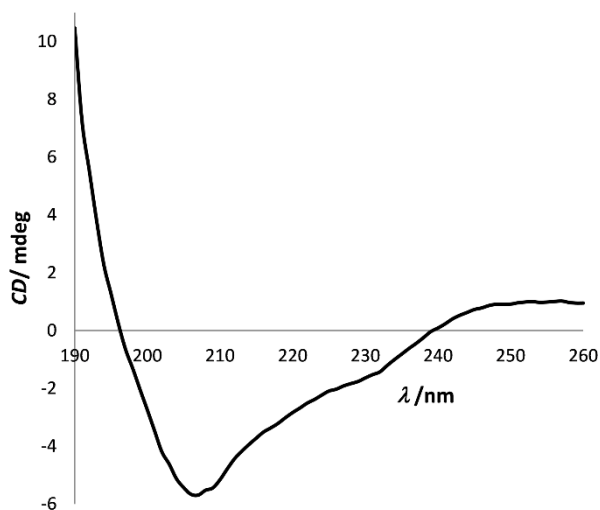
#### 4.1.3. Size of *HpFlgD* in solution

Recombinant His tagged *HpFlgD*\_26695 was eluted as single peak, at a volume of 13.82 mL (Fig. 20b). The molecular mass calculated from the analytical size-exclusion chromatography analysis was about 77.2 kDa for the His tagged *HpFlgD* that was in agreement with the size of protein estimated by MALS measurement (MW *HpFlgD*\_26695\_His<sub>6</sub> = 78.7 kDa). Considering that the *HpFlgD*\_26695 monomer is a His tag fusion protein (2.7 kDa), its molecular mass is



approx. 34.5 kDa and therefore, the *HpFlgD* oligomerization state is likely to be dimeric in the solution.

CD analysis was used to confirm the folded state of the recombinant samples as well as to evaluate the overall secondary structure content. According to the CD analysis, full-length *HpFlgD\_G27* is mainly composed of  $\beta$ -sheets and random coils. The presence of  $\alpha$ -helices was also detected but with a minor contribution (Fig. 22 and Table 33).

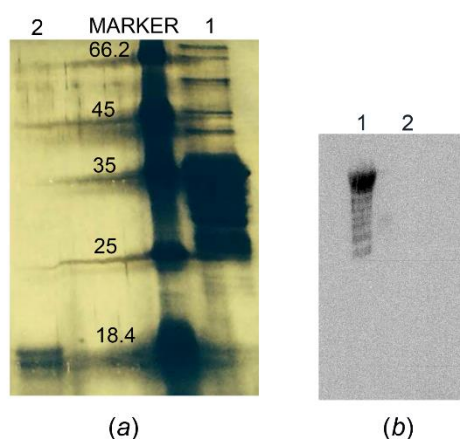


**Figure 22.** CD spectrum of the full-length *HpFlgD\_G27* in the far UV region (190–260 nm) presented as a CD signal in milidegrees.

**Table 33.** CD data of the full-length *HpFlgD\_G27* analyzed by the secondary structure analysis software, CDNN.<sup>110</sup> Deconvoluted results are shown as contributions of the various components to the protein secondary structure.

SECONDARY STRUCTURE ELEMENT	%
Helix	12.8
Antiparallel $\beta$ sheet	25.2
Parallel $\beta$ sheet	5.0
$\beta$ turn	22.8
Random coil	24.8

Results of the western blotting (Fig. 23) confirmed presence of a His tag at the C-terminus in the full-length *HpFlgD* while the crystallized fragment is lacking the signal for the anti His antibodies.

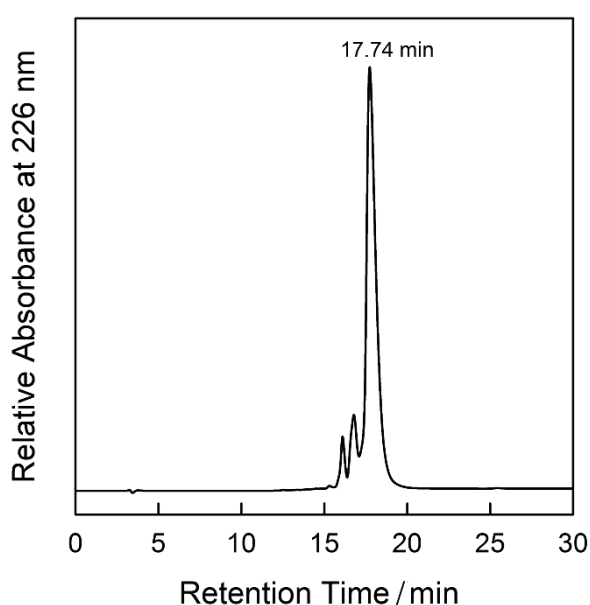


**Figure 23.** (a) SDS-PAGE; (b) Western blot against His tag at the C-terminal end. Lane 1 – full-length *HpFlgD\_26695* (top band) with initial degradation products (lower bands), Lane 2 – dissolved crystal of *HpFlgD\_t*.

#### 4.1.4. Size of *HpFlgD* in the crystal

Crystallization of the full-length proteins of both strains, *HpFlgD\_G27* and *HpFlgD\_26695*, was attempted as was demonstrated by SDS – PAGE electrophoresis that confirmed that full-

length protein was present in the sample used to set up crystallization experiments (Fig. 20a). Presence of the full-length protein was also confirmed by mass spectrometry. *HpFlgD\_G27* was prior mass analysis purified by RP-HPLC and major peak was isolated (Fig. 24). Mass spectrometry analysis confirmed that distinctive peak corresponds to 36178 Da. However, crystals of both forms (tetragonal, *HpFlgD\_t*, and monoclinic, *HpFlgD\_m*) contain a truncated form of *HpFlgD*. The proteolytic cleavage in *HpFlgD* proceeds much slower than in *XcFlgD* and *PaFlgD* where the crystals were obtained in 14 and 3 days, respectively, while in our case 2 months were needed.<sup>12,14</sup>



**Figure 24.** HPLC chromatogram of the full-length *HpFlgD\_G27*. The major species present in the solution corresponds to the size of 36,178 Da.

In both crystal structures the monomers comprise residues Asn127 – Lys272, which are well defined from the electron density. It can be seen from the sequence that amino acid residues are missing in both N- and C-terminal regions (Fig. 19). The first 126 amino acid residues cannot be seen in the electron density map, as well as the last 29 amino acid residues in *HpFlgD\_t* (strain 26695) and the last 44 residues in the case of *HpFlgD\_m* (strain G27). It has been argued that the N-terminal region is susceptible of proteolytic cleavage due to some degree of flexibility.<sup>14</sup> To determine whether these terminal parts were cleaved or are present in the crystal as disordered regions, some tetragonal crystals were dissolved and checked by SDS –

PAGE electrophoresis (Fig. 23a). The gel shows presence of two fragments corresponding to a size of about 17 kDa. These bands were extracted for the nano-ESI mass spectra analysis, which confirmed presence of four peptides that started and ended at different residues within the range of Glu122–Lys246 (Fig. 25). Despite the fact that a protease inhibitor cocktail was used during all of the steps of purification, after some time purified *HpFlgD* contained degradation products (Fig. 23b). It is hard to state whether the cleavage is a consequence of the presence of a small amount of *E. coli* protease that could cleave flexible N- and C-ends of *HpFlgD*, or because of the protein instability that causes spontaneous degradation over time.

```

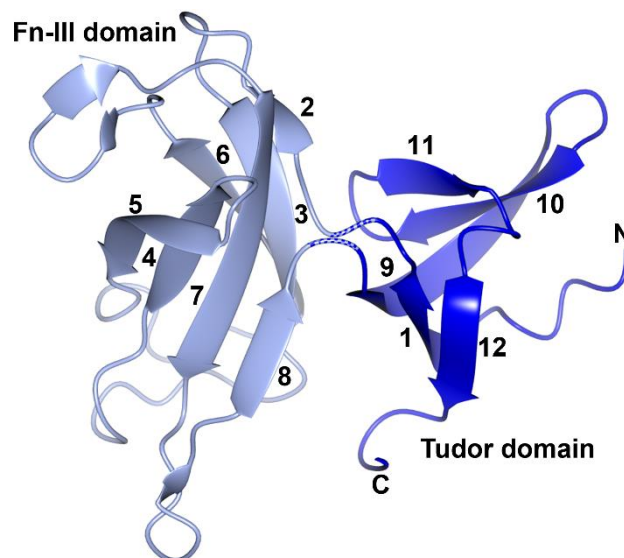
HP26695  1  MAIDLAEVTGAKAAQERKKEOPTIANGLDKNAFMKLFLEQLKNQDPTAPMETDKIITQTA  60
HP26695  61  QLTQVEMQEENKKTMQEVASAMKSNKETNESLKDFQGALKDTMENLNKGMDDSLKANNAL  120
HP26695 121  REVTAINSVSMIGKIAETDVSGANFDGNNKLSFSLFFDEKIDASKGVPAIQILNENNELV  180
HP26695 181  KTIPLKDYNGQKGYINFEWDTNEKGEKVPKGNYKIKAEYNLDSHSKQYLQTRIGRGEVE  240
HP26695 241  SVIFDKGKPMLRMGEMVLPIDSAIEFYQPDQKPLEQKLSDQKPIDQKPLDQKQTPPKETA  301

```

**Figure 25.** Results of the mass spectrometry: the peptides found in full-length *HpFlgD*\_26695 are bolded in red, while the peptides found in the tetragonal crystal of *HpFlgD*\_26695 are indicated by blue dotted lines.

#### 4.1.5. Structure of the *FlgD* monomer

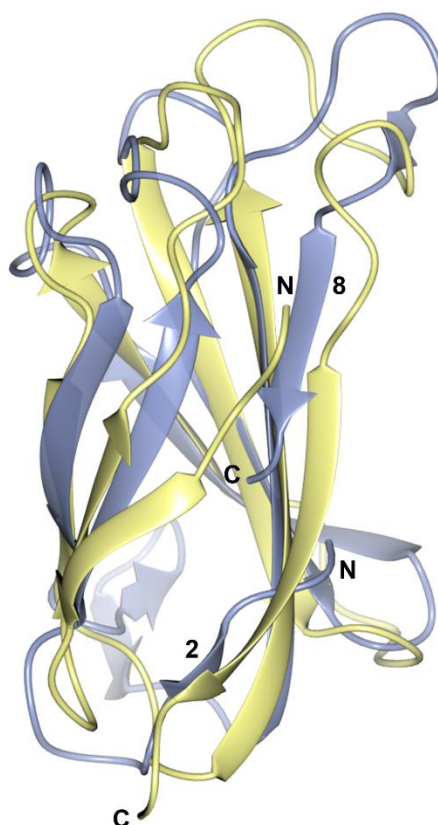
The two crystal forms are virtually identical, since the two strains, G27 (*HpFlgD*\_m) and 26695 (*HpFlgD*\_t), differ in only 2 amino acids in the crystallized fragments (residues 257 and 268, Fig. 19). The root mean square deviation (r.m.s.d.) between 146 equivalent C $\alpha$  atoms is 0.38 Å. *HpFlgD*, similarly to the other two *FlgD* structures previously determined,<sup>14,12</sup> is composed of two separate domains (Fig. 26). Domain II is continuous along the polypeptide chain, spanning from residue 141 to 234, whilst domain I includes 14 residues at the N-terminus (from 127 to 140) and all the remaining portion at the C-terminus, from 235 to 272. Lys272 is the last residue visible in the structure of both crystal forms.



**Figure 26.** Cartoon view of the *HpFlgD\_t* monomer. Secondary structure elements of the tudor and Fn-III domains are colored in dark blue and light blue, respectively. This figure was produced using CCP4MG.<sup>47</sup>

Domain I can be classified as belonging to the family of tudor domains, consisting of a strongly-bent five-strand  $\beta$ -barrel. Domain II can be defined as a fibronectin type III (Fn-III)-like domain. In fibronectin the module consists of two layers of  $\beta$ -sheets, one with three antiparallel strands and the other with four antiparallel strands. In *HpFlgD*, the Fn-III domain forms a  $\beta$ -sandwich also made by seven antiparallel  $\beta$ -strands, however, one layer is composed of only two antiparallel strands while the other has five antiparallel strands. This is the consequence of the two main differences:  $\beta$ 8-strand of *HpFlgD* does not overlap well with the last  $\beta$ -strand of the C-terminal domain in fibronectin; there is no overlap between the first  $\beta$ -strand of the Fn-III domain in *HpFlgD* ( $\beta$ 2-strand) and fibronectin, instead there is an overlap of the  $\beta$ 2-strand of *HpFlgD* with the last  $\beta$ -strand of the C-terminal domain in fibronectin. The superposition of  $C^\alpha$  atoms of domain II of *HpFlgD* with the fibronectin domain of 1FNA (Fig. 27),<sup>137</sup> the prototype of the family, gives a r.m.s.d. of 2.51 Å for 61 aligned residues. The Fn-III domain is also found as one of the five domains in the streptococcal cell wall of C5a peptidase (PDB ID 1XF1), a highly specific protease and adhesin/invasin.<sup>138</sup> Fibronectin, a multiple domains protein, can bind through a particular RGD motif in Fn-III domain to the specific cell membrane receptors, collagen, fibrin, and heparin,<sup>139</sup> and is a possible substrate of adherence of microorganisms to the eukaryotic cells.<sup>140</sup> Even though the flagellin component

D does not possess the common RGD motif, a similar tripeptide sequence (WDG) is found in the Fn-III domain of *HpFlgD*, as well as in the other two FlgD orthologues (*XcFlgD* and *PaFlgD*, Fig. 19). Similar architecture of the Fn-III domain in human proteins and proteins from pathogenic bacteria could be relevant for the recognition of the human host cells (by the extracellular matrix proteins or integrins) by pathogens for adhering to the human gastric epithelial cells to set up a persistent infection. Moreover, sharing the same domain could imply the ability of *HpFlgD* to be involved not only in the control of the proper structural organization of the flagellar hook, but also directly in the *H. pylori* pathogenesis.

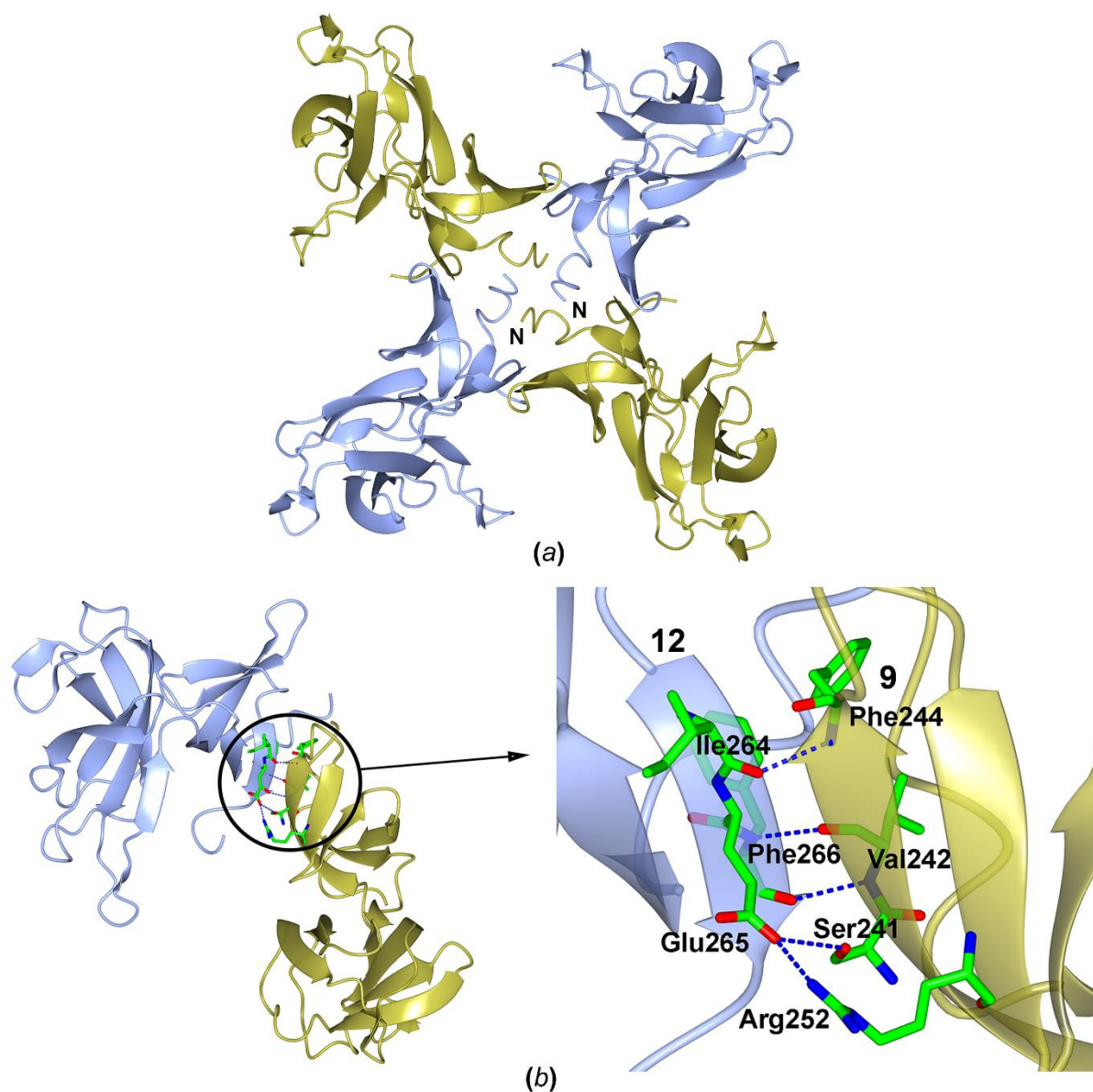


**Figure 27.** Superposition of the Fn-III domain in fibronectin (yellow) (PDB entry ID 1FNA) to the same domain in *HpFlgD\_t* (light blue). The r.m.s.d. for the superposition of 61 aligned C $\alpha$  atoms of fibronectin on *HpFlgD\_t* is 2.51 Å. This figure was produced using CCP4MG.<sup>47</sup>

Differences of the inter-domain orientation in the crystal structures of *HpFlgD* and other FlgD homologues are described in the second paragraph below.

#### 4.1.6. Quaternary structure and interfaces

The quaternary structure in both crystal forms is a tetramer. Four monomers are present in the asymmetric unit of *HpFlgD\_m*, corresponding to a molar volume ( $V_M$ ) of  $2.61 \text{ \AA}^3 \text{ Da}^{-1}$  and an approximate solvent content of 53 %. Monomers A and B, together with their pair related by a two-fold symmetry axis, form one tetramer (Fig. 28a). Another two monomers (C and D), from the asymmetric unit, with their pair related by a two-fold symmetry axis form other tetramer. In the structure of *HpFlgD\_t* there is one monomer in the asymmetric unit, with a  $V_M$  of  $3.23 \text{ \AA}^3 \text{ Da}^{-1}$  and an approximate solvent content of 62 %. The tetramer is generated by the four-fold symmetry axis. The analysis of the interfaces by software PISA<sup>135</sup> shows that the tetramer observed in the crystal corresponds to the molecule of truncated protein in solution. The surface area buried after tetramer formation corresponds to  $3898 \text{ \AA}^2$ , about 12.9 % of the total surface. Dimerization takes place through pairing of  $\beta$ -strand 9 and  $\beta$ -strand 12 of the two tudor domains from neighbouring monomers (Fig. 28b) giving rise to a continuous  $\beta$ -sheet between them. This is the largest interface area in the crystal structures of both *HpFlgD\_t* and *HpFlgD\_m*. This situation repeats four times, creating a network of hydrogen bonds and salt bridges between the four monomers (Table 34).



**Figure 28.** (a) *HpFlgD\_m* tetramer (monomer A is shown in blue and B in gold color). (b) Monomers A and B in the structure of *HpFlgD\_m* interconnected by hydrogen bonds (blue dashed bonds) with a close-up of the hydrogen bonds. This figure was produced using CCP4MG.<sup>47</sup>

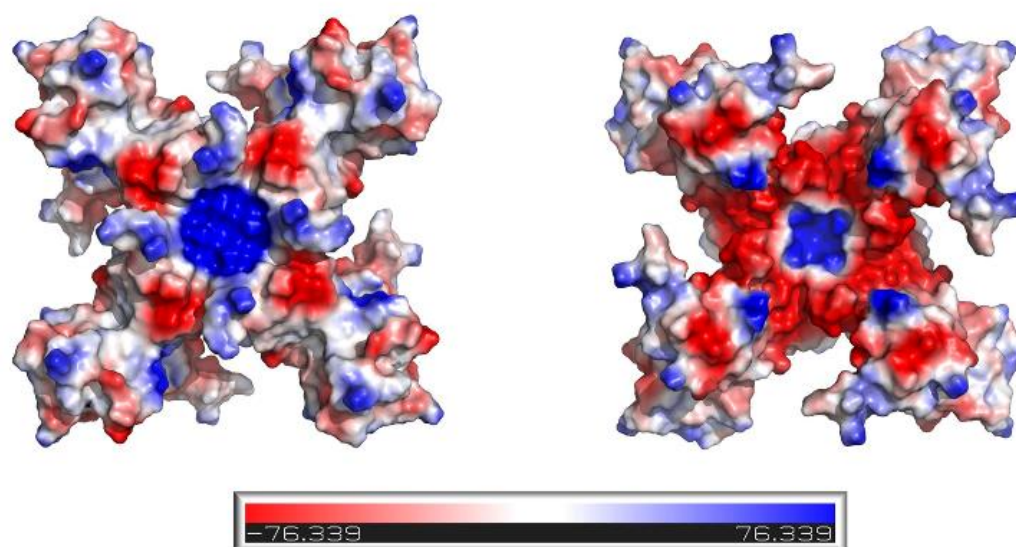


**Table 34.** Hydrogen bonds (Å) between monomers that are relevant for their assembly into a tetramer.

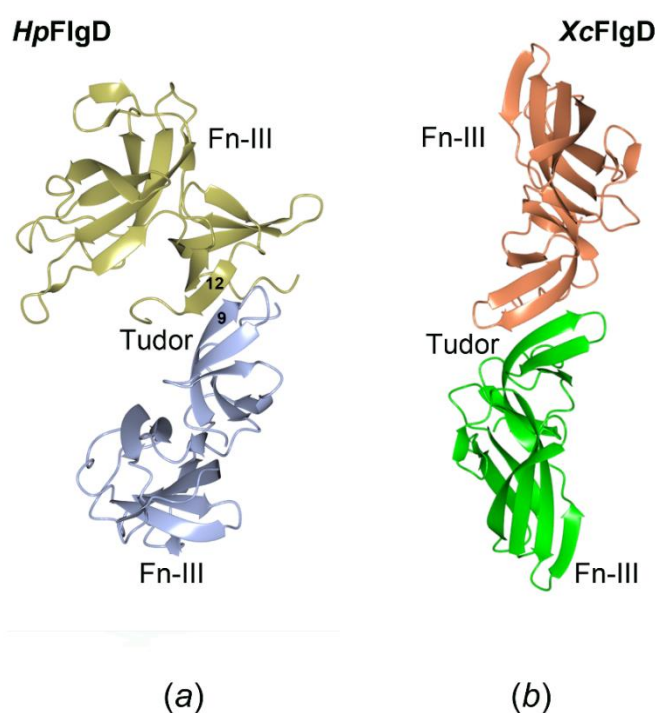
		<i>HpFlgD_t</i>		<i>HpFlgD_m</i>		
Monomer1	Monomer2[Symmetry code]	A···A[y+1, x, z]	B···A[-x, y, -z]	A···B	D···C[-x, y, -z]	C···D[x-1, y, z-1]
Ile264 [O]	Phe244 [N]	2.71	2.92	2.75	3.07	2.83
Phe266 [N]	Val242 [O]	2.88	2.82	2.80	2.94	2.90
Phe266 [O]	Val242 [N]	3.01	2.87	2.92	2.89	3.03
Glu265 [OE2]	Ser241 [OG]	2.93	3.36	3.20	3.15	3.17
*Glu265 [OE2]	Arg 252[ NH2]	3.43	2.79	2.61	2.72	2.85

\*Denotes the salt bridge.

In addition, the four N-termini of the chain present in the crystal pack together, filling the hole generated by tetramerization and giving rise to a compact structure that assumes the shape of a cap, with positive charges at the center of the convex surface, and positive charges at the concave center surrounded by negative charges on the surface (Fig. 29). An interface involving neighboring tudor domains is found also in other two homologous structures, labeled Type III in both of them, Fig. 30 (notation according to Kuo *et al.*, 2008).<sup>14</sup> However, there is a significant difference between Type III interface and the main interface in *HpFlgD*. Contrary to pairing of  $\beta$ -9 and  $\beta$ -12 strands in *HpFlgD*, there is pairing of two neighboring  $\beta$ -9 strands in Type III (this numeration is according to *HpFlgD* to enable comparison). Type III interface is not favorable for tetramer formation since it forms dimers (the monomers are related by a two-fold symmetry axis), contrary to the interface in *HpFlgD* where  $\beta$ -9 to  $\beta$ -12 pairing enables further pairing of the same kind and assembly into a tetramer. Interestingly, conserved residues are present only in the  $\beta$ -9 strand in all three FlgD protein structures (Fig. 19).

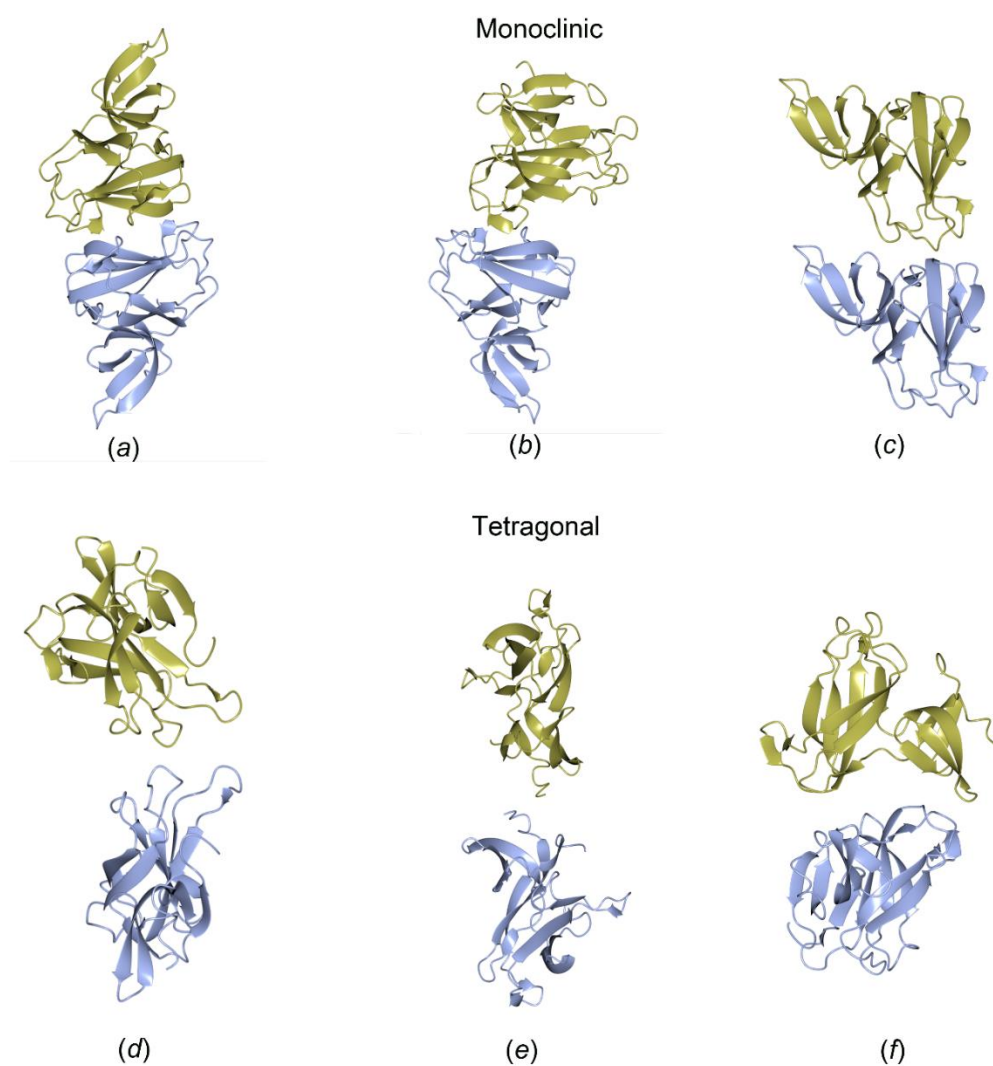


**Figure 29.** Qualitative electrostatic surface (represented as  $k_b T e^{-1}$ ) of the tetramer, seen along the convex (on the left) and concave (on the right) surface. This figure was prepared with software PyMol.<sup>134</sup>



**Figure 30.** Type of interface between the tudor domains present in the crystal structure of *HpFlgD* (a) and *XcFlgD* (b). This figure was produced using CCP4MG.<sup>47</sup>

Due to the shorter length of the  $\beta$ -12 strand in *XcFlgD* and *PaFlgD* than in *HpFlgD* a much smaller interface of the type  $\beta$ -9 to  $\beta$ -12 could only be formed, possibly not favourable for tetramer formation. Except for the main interface which is involved in the tetramer formation in *HpFlgD* the other three interfaces have smaller interface areas (Fig. 31, Table 35). None of these interfaces correspond to those in *XcFlgD* and *PaFlgD*, where interfaces of the Type I, II and III (notation according to (Kuo *et al.*, 2008)<sup>14</sup> are present.



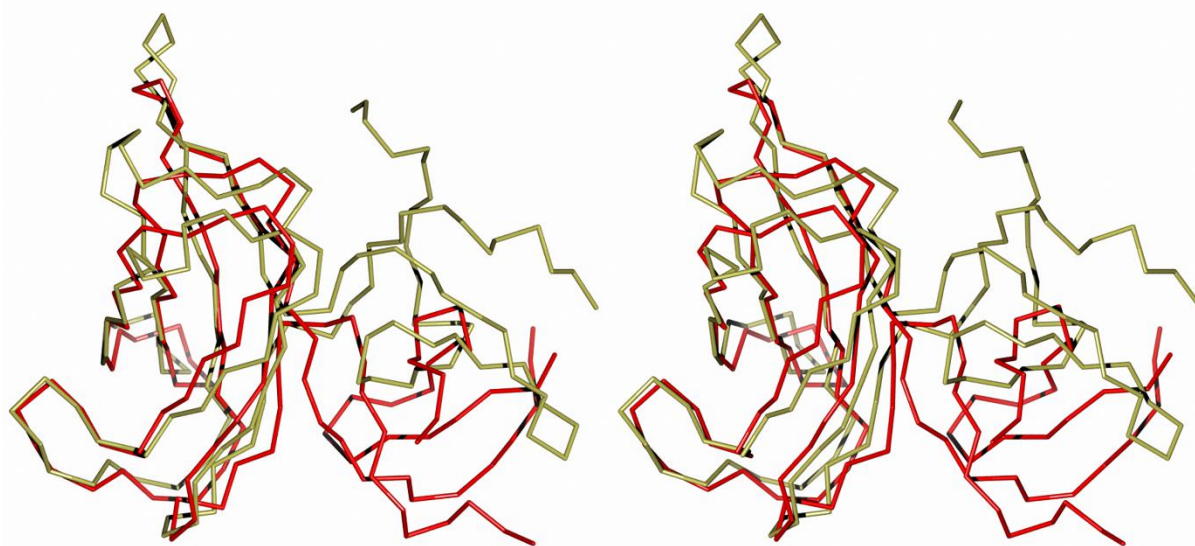
**Figure 31.** Different types of interfaces found between the molecules in the monoclinic crystal structure of *HpFlgD* (a, b, c) and in the tetragonal crystal structure of *HpFlgD* (d, e, f). This figure was produced using CCP4MG.<sup>47</sup>

**Table 35.** Different types of interfaces between the molecules in the crystal structure of *HpFlgD*.  $N_{HB}$  and  $N_{SB}$  refer to the number of hydrogen bonds and salt bridges, respectively. The interfaces labelled a–f with a superscript\* are shown in Fig. 31, while the interface labelled as t\* refers to the interface responsible for the tetramerization (as shown in Figs. 28b and 30).

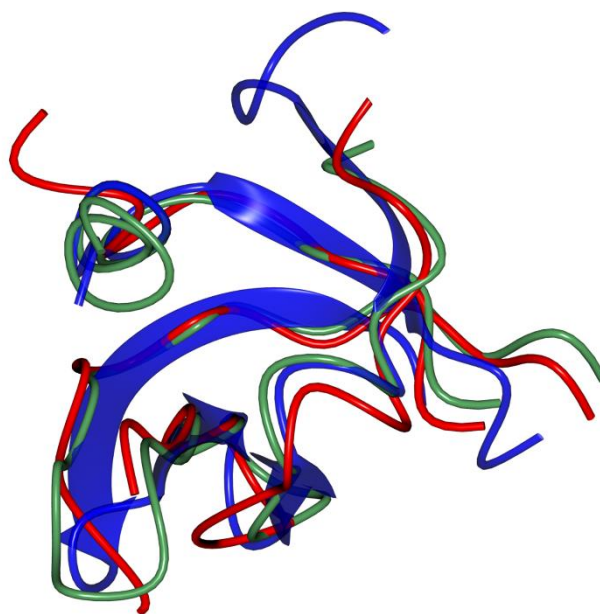
Crystal system	Interface type	Monomer1···Monomer2[Symmetry code]	Interface area / Å <sup>2</sup>	$N_{HB}$	$N_{SB}$
<b>Monoclinic</b>					
	t*	B···A[x, y, z]	521.9	7	3
	t	D···C[-x, y, -z]	492.5	6	2
	t	B···A[-x, y, -z]	492.1	7	2
	t	C···D[x-1, y, z-1]	494.8	8	3
	a*	C···B[x, y, z]	302.3	–	–
	b*	D···B[-x+1, y, -z]	297.3	1	-
	c*	D···D[x, y-1, z]	280.1	1	5
	c	A···A[x, y-1, z]	268.2	1	5
	c	B···B[x, y-1, z]	216.4	3	5
	c	C···C[x, y-1, z]	194.3	3	5
<b>Tetragonal</b>					
	t	A···A[-y+1, x, z]	478.8	12	2
	d*	A···A[x, -y+1, -z]	299.9	4	4
	e*	A···A[-x, -y+1, z]	201.5	2	–
	f*	A···A[y-1/2, x+1/2, -z + 1/2]	119.2	4	4

#### 4.1.7. Comparison with other structures

Identity of *HpFlgD* with *PaFlgD* and *XcFlgD* is 24 % and 20 %, respectively (Fig. 19). Despite this low sequence identity the three hook scaffolding components share common folds of each domain. However, the overall structure of each protein shows significant differences. In fact, Fn-III domain superimposes quite well with that of the orthologue FlgD from the other species (r.m.s.d.s are 2.17 Å for 93 residues in *XcFlgD* and 1.73 Å for 91 residues in *PaFlgD*). On the contrary, by superimposing *HpFlgD\_t* and *PaFlgD* the larger Fn-III domains superimpose well, however it can be observed that the tudor domain in *HpFlgD* has a totally different orientation with respect to the tudor domain in *PaFlgD* (Fig. 32), and also with respect to this domain in *XcFlgD* which has the same orientation as in *PaFlgD*. The tudor domain by itself can be superposed quite well in all three homologues (Figure 33). Different relative orientation of domains in *HpFlgD* than in others could imply the presence of two distinct orientation modes with a switch between them happening as a part of the mechanism during its activation.

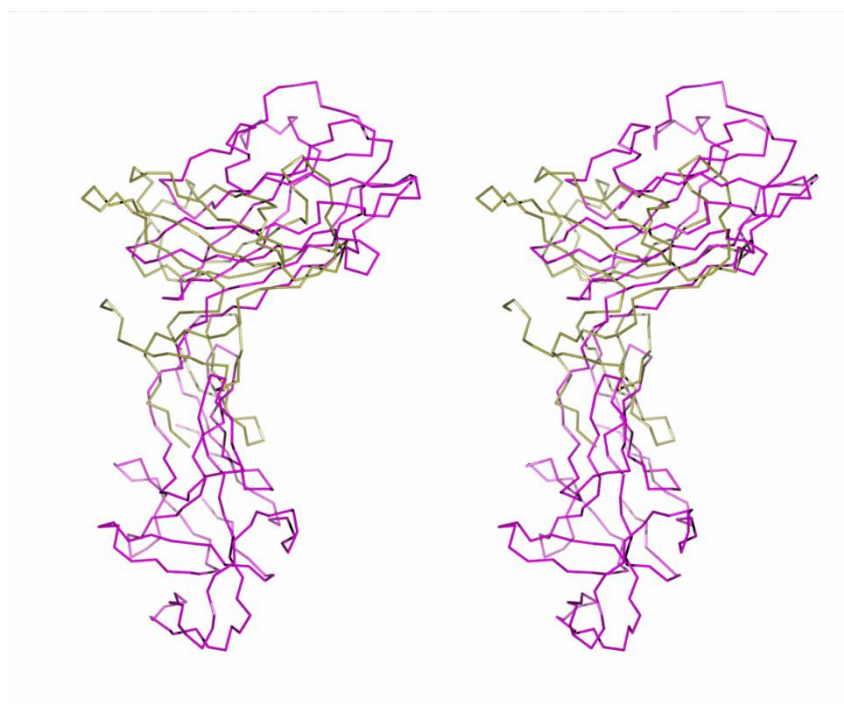


**Figure 32.** Stereoview of the superposed C $\alpha$  chain trace of *PaFlgD\_A* (red, PDB entry 3OSV;<sup>12</sup> on *HpFlgD\_t* (gold). The r.m.s.d. for the superposition of 91 aligned C $\alpha$  atoms of *PaFlgD\_A* on *HpFlgD\_t* is 1.73 Å. This figure was produced using CCP4MG.<sup>47</sup>



**Figure 33.** Superposition of the tudor domain in *HpFlgD\_t* (dark blue), *XcFlgD* (green) and *PaFlgD\_A* (red). The r.m.s.d.s for the superposition of 47 aligned C $\alpha$  atoms of *XcFlgD* on *HpFlgD\_t* and 43 aligned C $\alpha$  atoms of *PaFlgD\_A* on *HpFlgD\_t* are 2.09 Å and 1.55 Å, respectively. This figure was produced using CCP4MG.<sup>47</sup>

FlgD is the hook cap protein, a necessary component in the formation of the filament of the flagellum.<sup>141</sup> To build the hook, FlgD must be assembled at the end of the rod structure,<sup>68</sup> before another flagellin, FlgE, is secreted and placed underneath it. The crystal structure of FlgE (or at least of a proteolytic fragment) from *Salmonella enterica subsp. enterica* serovar Typhimurium has been determined (PDB ID 1WLG),<sup>62</sup> and a low resolution image of the organization of the hook was obtained by electron microscopy: 120 copies of the flagellin are arranged in a helical structure, with about 12 molecules per turn. Interestingly, *Hp*FlgD shares an architecture that roughly resembles that of *Hp*FlgE, the protein that constitutes the hook of the flagellum. Superposition of *Hp*FlgD to a model of *Hp*FlgE built by homology by the software Phyre<sup>2</sup> is shown in Fig. 34. The r.m.s.d for the superposition of C<sup>α</sup> atoms is 3.44 Å for 68 aligned residues, despite a sequence identity of less than 14 %. The most relevant aspect is that both *Hp*FlgD and *Hp*FlgE are composed of two domains, each of them with a similar architecture, despite both domains in *Hp*FlgE being larger than the corresponding domains in *Hp*FlgD. It can be speculated that this similarity could play some role in the mechanism of the hook assembly



**Figure 34.** Stereoview of the superposed C<sup>α</sup> chain trace of *Hp*FlgD<sub>t</sub> (gold) and the modeled *Hp*FlgE (purple). The r.m.s.d. for the superposition of 68 aligned C<sup>α</sup> atoms of modeled *Hp*FlgE on *Hp*FlgD<sub>t</sub> is 3.44 Å. This figure was produced using CCP4MG.<sup>47</sup>

#### 4.1.8. Proposed physiological role of truncated *HpFlgD*

Proteome analysis identified *HpFlgD* in the *H. pylori* culture supernatant within the pool of the most represented secreted components.<sup>142,143</sup> However, its function as well as processing has not been characterized so far. The N-terminal region of *HpFlgD* is susceptible to proteolytic cleavage, as confirmed by crystallized fragments (Fig. 23a). Analogous proteolytic events gave rise to crystals composed of only the C-terminal fragment in all FlgD structures characterized till now.<sup>12,14</sup>

The N-terminal region (pfam03963), despite the fact that no typical features of a canonical export signal are present, is well conserved among the flagellar hook capping proteins family (cl15992). The *HpFlgD* N-terminal sequence is predicted to include an initial flexible segment (corresponding to residues 1 to 27) followed by two predicted helical motifs and a further disordered segment (93-116; Psi-pred server results). The presence of some helical content in *HpFlgD* is confirmed by the CD data of the entire protein (Fig. 22 and Table 33). The N-terminal region is not only flexible and more prone to degradation, but has been demonstrated to be essential for the secretion in both *Salmonella enterica* and *Escherichia coli* variants. Indeed, the export signal of *EcFlgD* has been demonstrated to be comprised of the first N-terminal 71 amino acids,<sup>70</sup> while the first 100 amino acids of *SeFlgD* were proven to guarantee a secretion level equivalent to the full-length protein once expressed in fusion with a reporter gene.<sup>144</sup> Such secretion signals are surprisingly longer than the canonical ones and do not show the typical Sec machinery consensus features.

Another intriguing aspect of *HpFlgD* is that the residues at the C-terminus, after Lys272, are not visible in both of the crystal structures determined in this work. Indeed, all the sequences of FlgD homologues in any *Helicobacter* species deposited in the UniProt data bank present a peculiar and disordered fragment of variable length at the C-terminus. It is composed of repeats of five amino acid motifs according to the consensus sequence XQK(X)<sub>2</sub> where the first position X is mainly present as D and E (rarely as N). The two last positions (X)<sub>2</sub> can vary according to PI and PL or less often as LS and PQ. The conserved and repeated motifs vary only in terms of the number of repeats, while they are all followed by the sequence TPPKETA that represents, in the majority of cases, the very C-terminal fragment. Such a peculiar sequence of repeats is not present in the FlgD homologues of other bacteria and cannot be easily reconciled to a well-known motif or signal. Nevertheless, the C-terminus of FliC protein, the flagellin subunit

defining the filament, is bound in an extended form by the FliS chaperone,<sup>145</sup> preventing its premature polymerization. Such an interaction is a common strategy for flagellar export chaperones, to bind their substrates and control flagella biosynthesis. While no data has been acquired until now on the physiological role played by *HpFlgD* C-terminus, such a peculiar and highly conserved sequence suggests it could represent an interaction or recognition motif. A BLAST search throughout a non-redundant database of deposited sequences has shown that similar patterns can be identified in other bacterial or plasmodia proteins with quite different functions and in different positions in the primary structure, supporting the idea that it could represent a common mode of recognition of interaction partners, rather than a signal sequence.

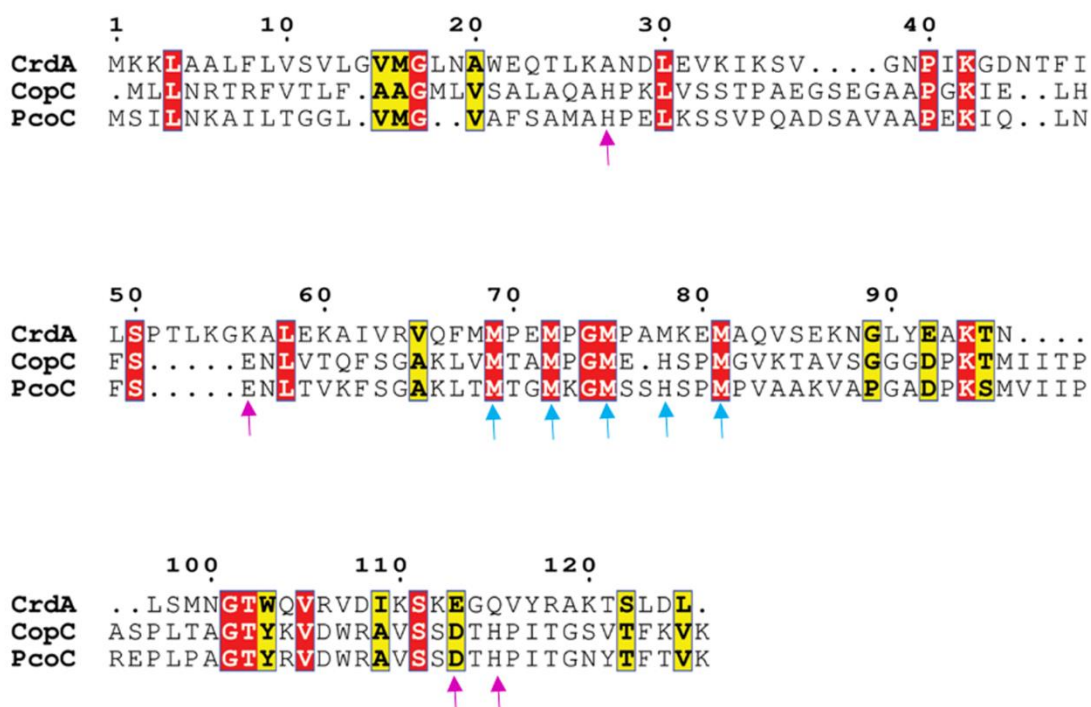
A relevant difference between *HpFlgD* and the other two FlgD homologues, whose structures are known, is represented by the quaternary organization. The entire *HpFlgD* is a dimer in solution, but it becomes a tetramer after cleavage of the N-terminus. This difference in the oligomerization state is likely due to the long flexible N-terminal portion, which is present in the purified protein, but is lost before crystallization. The latter apparently prevents tetramerization, since the N-termini in the form found in the crystal are in the center of the tetramer, in close proximity of the four-fold axis (Figs. 28a and 29). It is hard to say if the aggregation state of *HpFlgD* has some physiological meaning or if it represents an artifact of crystallization. However, previous considerations suggest that this cleavage has a functional significance, since the protein has to be exported into the outside space in order to interact with the hook, before the flagellum is fully formed. The flexible N-terminal part is also absent in the crystal structures of both *PaFlgD* and *XcFlgD*, strongly supporting the hypothesis that this part plays a role in the physiological mechanism of flagellar assembly. The N-terminal fragment could act as a signal for the export and the crystallized protein could correspond to the mature protein.



## 4.2. CrdA protein from *H. pylori*

### 4.2.1. Cloning, expression and purification of *StrepCrdA*, *CrdAHis* and *GSTCrdA*

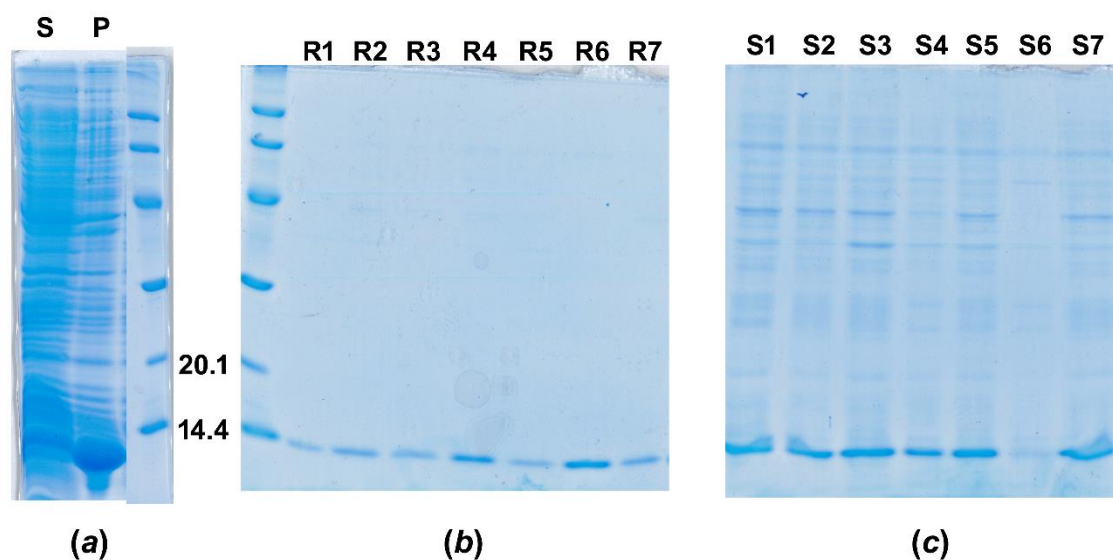
Cloning without the first 22 amino acids was performed since the CrdA sequence analysis by the *LipoP* bioinformatic tool (<http://www.cbs.dtu.dk/services/LipoP/>) predicted a cleavage site for an export signal and a lipidation site (aa 22–23), Fig. 35. To determine whether the CrdA sequence possesses conserved regions for the binding of copper ions, the sequence was compared with the sequences of the CopC and PcoC. These proteins belong to the CopC family and are periplasmic proteins involved in process that results in a change in state or activity of a cell or an organism as a result of a copper ion stimulus. The alignment showed (Fig 35) that CrdA sequence contains conserved amino acids, mainly methionines, that could be involved in the copper binding.



**Figure 35.** Sequence alignment of CrdA (HP, O25884), CopC (PS, P12376) and PcoC (EC, Q47454) using CLUSTAL OMEGA.<sup>130</sup> The alignment figure was generated using ESPrpt 3.0.<sup>136</sup> The numbering system is based on the CrdA sequence. Residues conserved in all three sequences are labelled in white and highlighted in red, while similar residues are labelled in

black and highlighted in yellow. Possible ligands for the Cu(II) and Cu(I) sites in CopC and PcoC are marked with magenta and turquoise arrows, respectively.<sup>78</sup>

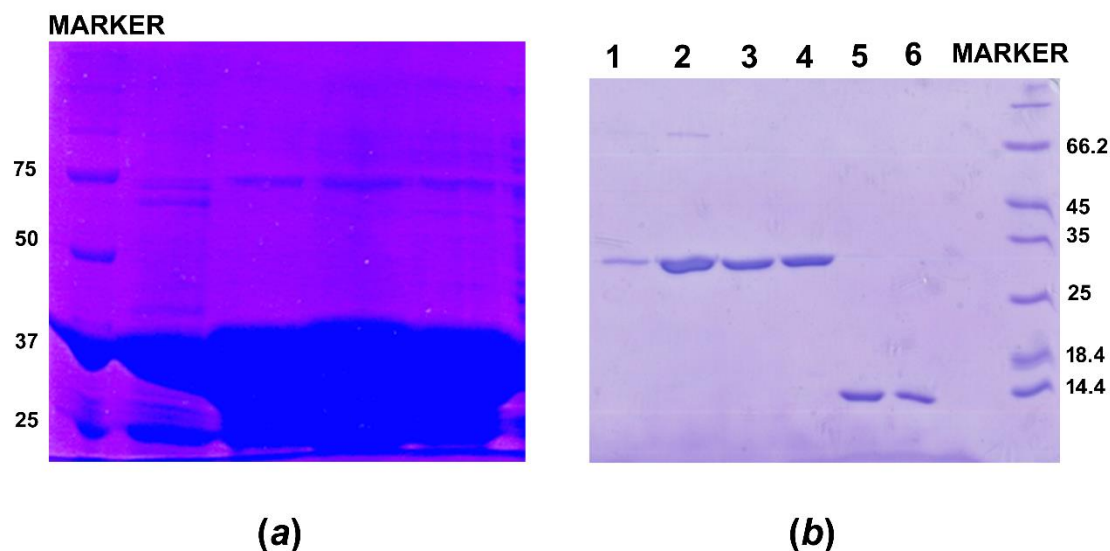
In the case of the CrdAHis-tag and Strep-tagCrdA protein, high expression levels were obtained but with a poor solubility (Fig. 36a), so refolding methods were implemented with success. The insoluble portion of CrdA was denatured and then renatured by slow dilution. The best renaturation outcome was obtained using the buffer R6 (Fig. 38b) since after the affinity chromatography purification the flowthrough of refolded CrdA in buffer R6 contained the lowest amount of protein (Fig. 36c).



**Figure 36.** SDS-PAGE (a) StrepCrdA, Lane S – soluble fraction, Lane P – insoluble fraction; (b) Lanes R1 – R7 - refolded CrdAHis in different renaturation buffers (R1 – R7, see Table 28 in section 3.7); (c) Lanes S1 – S7 – flow-through of CrdAHis in different renaturation buffers after the Ni-NTA affinity chromatography.

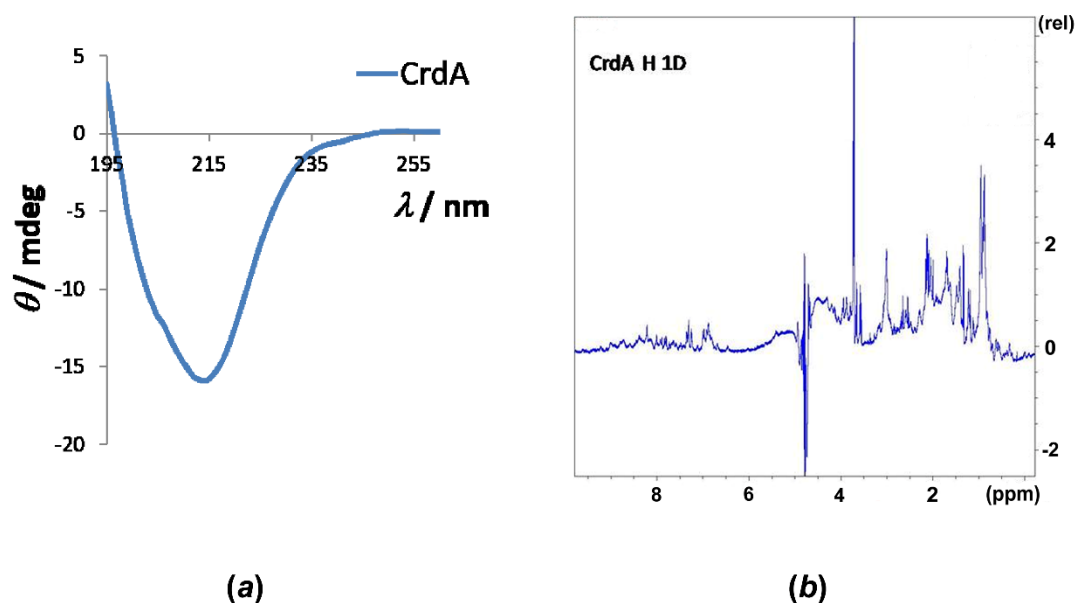
The best condition for expressing the GSTCrdA protein was obtained when *E. coli* cells were grown at 16 °C (Fig. 37a). As can be seen in Fig. 37b, SEC (*Superdex 75 10/300 column*) of cleaved GSTCrdA was successfully applied for the separation of the CrdA protein from the

GST-tag protein. This was possible since the GST protein is present as a dimer in the solution while CrdA as a monomer (see section 4.2.2.).



**Figure 37.** SDS-PAGE (a) GSTCrda (Molecular mass of the monomer ~ 38 kDa); (b) cleaved GSTCrda purified by SEC with lanes 1 – 4 and 5 – 6 corresponding to the GST (Molecular mass of the monomer ~ 26 kDa) and CrdA (Molecular mass of the monomer ~ 12.4 kDa) fractions, respectively.

The proper folding of renatured CrdA was confirmed by CD and 1D-NMR spectroscopy. The far-UV circular dichroism spectrum (Fig. 38a) shows a minimum at 214 nm and a peak of positive ellipticity for the values below 197 nm that are characteristic of a properly folded protein rich in  $\beta$  sheets (Table 36).<sup>108</sup> The 1D NMR spectrum of refolded CrdA indicates presence of secondary structure elements (Fig. 38b) since the peaks in the spectrum are sharp and narrow and cover a large range of chemical shifts. and the spectrum also contains peaks in the negative ppm range. In an unfolded protein, a given proton does not have one distinct environment, so the peak is an average of all the different environments in which that proton is found. Also, the environments of different protons aren't as distinct in an unfolded protein, so the peaks don't cover a large range of chemical shifts. In a folded protein, most protons are more constrained in a distinct conformation, and the environment of each proton is very dependent on its position in the tertiary structure of the protein.



**Figure 38.** Refolded CrdA (a) CD spectrum of the StrepCrdA in the far UV region (195–260 nm) presented as ellipticity in millidegrees; (b) 1D NMR spectrum of the CrdAHis.

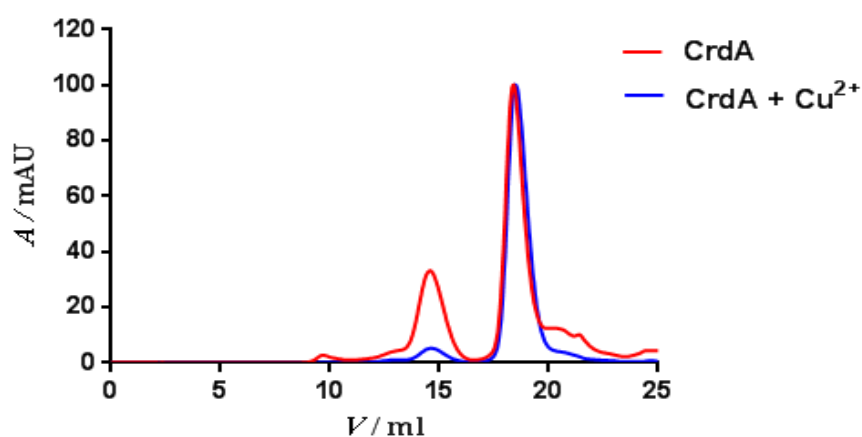
**Table 36.** CD data of the StrepCrdA analysed by the secondary structure analysis software, CDNN.<sup>110</sup> Deconvoluted results are shown as contributions of the various components to the protein secondary structure.

SECONDARY STRUCTURE ELEMENT	%
Helix	9.0
Antiparallel $\beta$ sheet	35.6
Parallel $\beta$ sheet	5.2
$\beta$ turn	19.5
Random coil	30.7

#### 4.2.2. Size of CrdA in solution

Incubation of CrdA with copper(II) ions or EDTA, followed by SEC, was carried out. Recombinant His tagged CrdA treated with copper was eluted as a single peak at a volume of 18.5 mL, while in case of CrdA treated with EDTA two different species were eluted: one at 14 mL and the other one at 18.5 mL (Fig. 39). The calculated molecular mass from the SEC experiments of the CrdA species in the presence of copper ions was about 14.8 kDa that was in

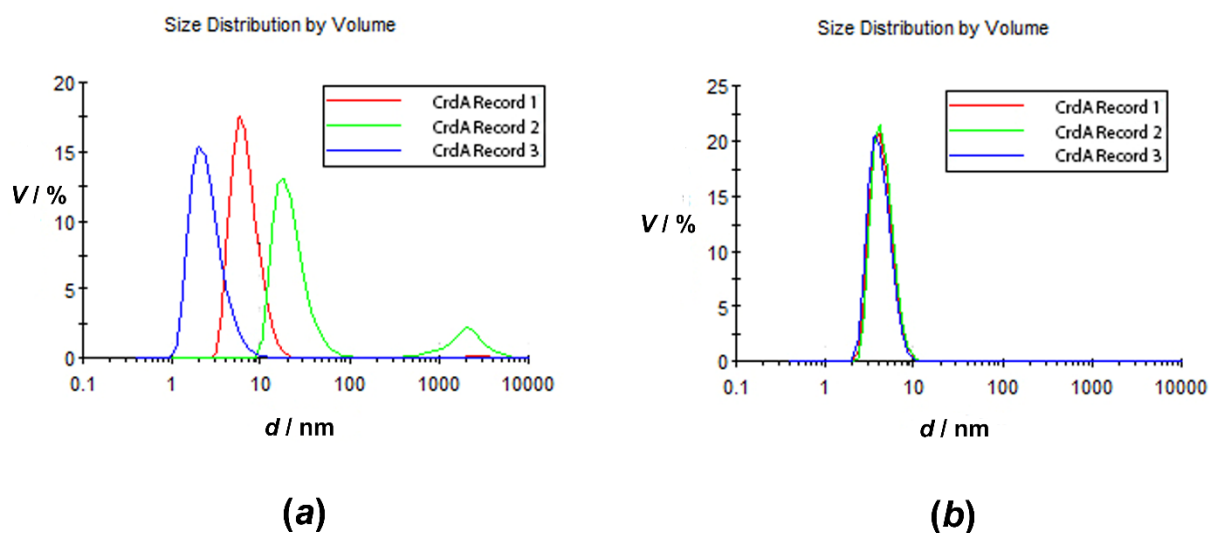
close agreement with the predicted molecular mass that is 12.4 kDa. The observed molecular mass indicates that the CrdA protein treated with copper is present as a monomeric species in the solution, which is consistent with the behaviour of the copper binding *PsCopC* protein in the solution.<sup>78</sup> In case of EDTA treatment, the CrdA protein eluted in two different oligomeric states - monomeric (14.8 kDa) and dimeric (31 kDa). Treatment with a metal chelating agent (EDTA) changed the behavior of CrdA, indicating that an equilibrium between the two different oligomeric species is present.



**Figure 39.** Analytical gel filtration chromatogram of CrdA treated with Cu(II)+ ions (blue curve; Molecular mass of the monomer ~ 12.4 kDa) which elutes as a single peak at the volume of 18.5 mL corresponding to the molecular mass of 14.8 kDa. CrdA chelated with EDTA (red curve) is composed of two different species that elute at 14 mL and 18.5 mL and correspond to the molecular mass of 31 kDa and 14.8 kDa, respectively.

The monodispersity of the CrdA protein with and without addition of detergent was evaluated at a protein concentration of 2 mg mL<sup>-1</sup> and at 298 K by dynamic light scattering. As it can be seen in Fig. 40 the dispersity of the pure CrdA sample and of the CrdA sample treated with nOG detergent ( $c = 3.8 \text{ mmol L}^{-1}$ ) differs significantly. The solution of the CrdA protein without detergent was heterogeneous and contained species in different oligomeric states and with a high level of aggregation. The polydispersity index for the CrdA sample and CrdA plus nOG was 0.941 and 0.335, respectively, implying that only the sample treated with detergent could be used for the crystallization purpose. The influence of other detergents like LDAO ( $c$

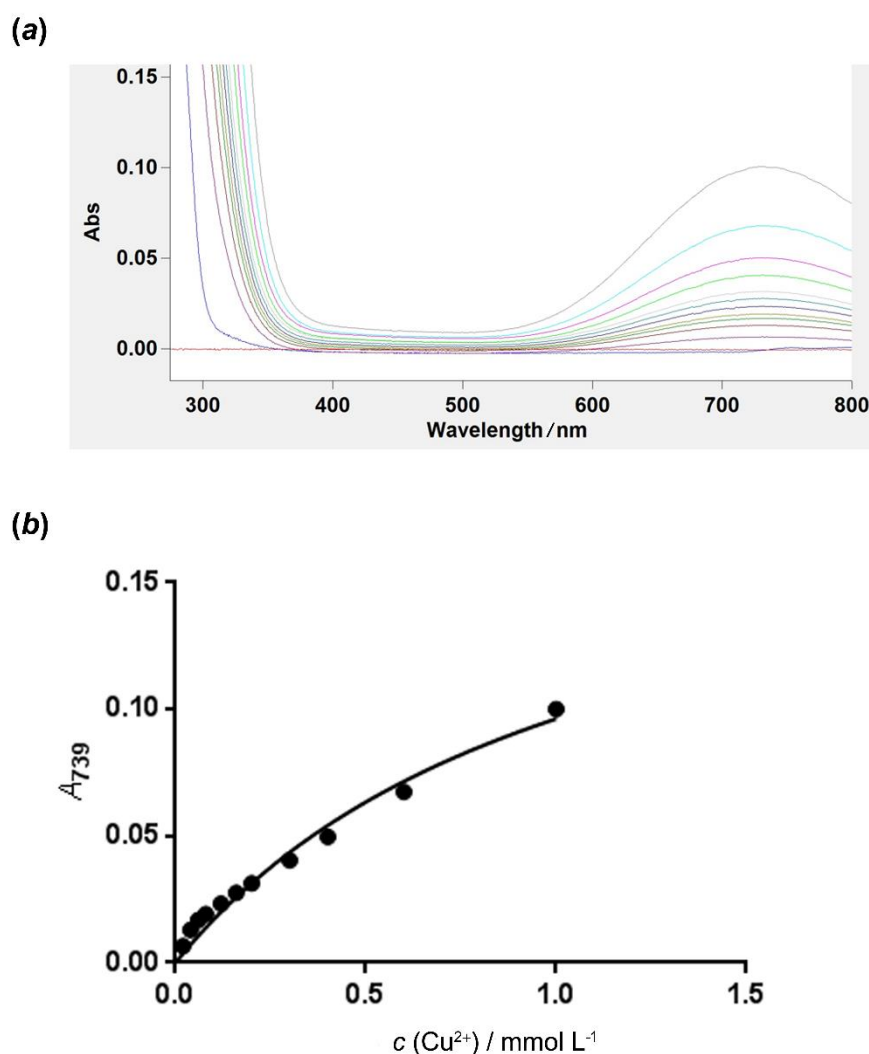
= 0.4 mmol L<sup>-1</sup>) and DDM ( $c = 0.034$  mmol L<sup>-1</sup>) on the dispersity of CrdA was also tested (results not shown), however they only slightly affected the dispersity profile of CrdA.



**Figure 40.** The volume distributions as a function of particle diameter: (a) CrdA and (b) CrdA treated with nOG detergent,  $c = 3.8$  mmol L<sup>-1</sup>.

#### 4.2.3. Affinity of CrdA towards copper(II) ions

The copper binding properties of the CrdA protein were investigated by UV-VIS spectroscopy in the wavelength range 200–800 nm. Fig. 41a shows presence of a broad absorption band at 739 nm but unfortunately in these experiments saturation was not reached, Fig. 41b. The dissociation constant ( $K_d$ ) for metal binding to the protein was determined from the titration to be  $1.076 \times 10^{-3}$  mol L<sup>-1</sup>. Such a high value of the dissociation constant is characteristic of copper trafficking proteins rather than metalloproteins with an enzymatic function.<sup>146</sup> Low affinity binding sites allow metal transfer to the recipient protein or enzyme while the high affinity sites prevent loss of the metal during redox cycling. Another possible explanation of such a high  $K_d$  is that the observed binding of copper(II) ions is unspecific, and maybe the preferable ligand for the CrdA protein is the Cu(I) ion. Based on the known crystal structures and the predictions for the Cu(II) and Cu(I) sites (Fig. 35) of similar copper binding proteins, CrdA could more likely accommodate Cu<sup>+</sup> than Cu<sup>2+</sup> since its sequence is rich in methionines and lacks histidines and cysteines.<sup>78,79</sup>

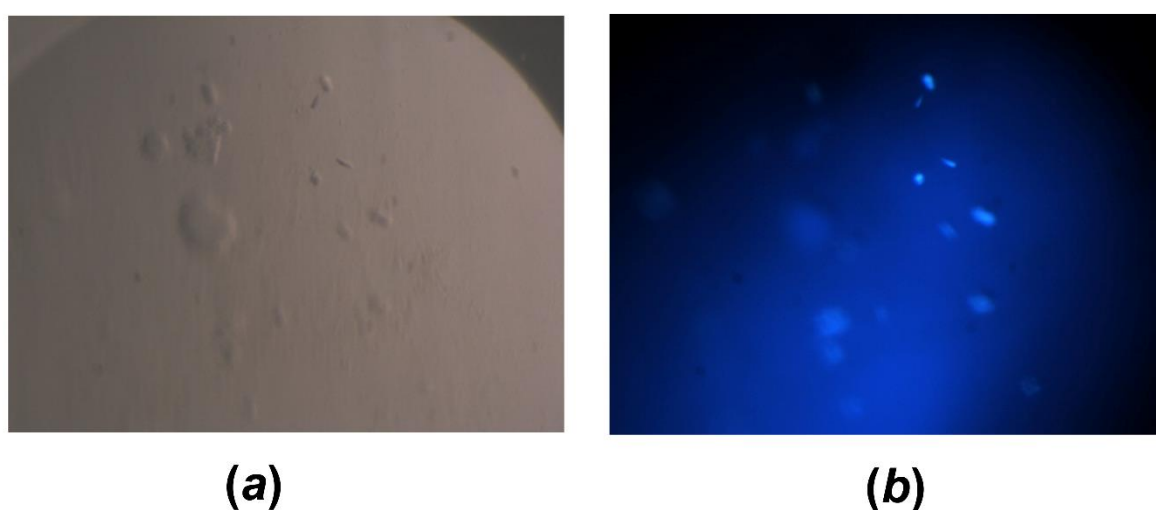


**Figure 41.** (a) Change in the UV-VIS spectrum upon titration of CrdA ( $c = 152 \text{ mmol L}^{-1}$ ) in the phosphate buffer ( $30 \text{ mmol L}^{-1} \text{ Na}_2\text{HPO}_4$  pH 7,  $50 \text{ mmol L}^{-1} \text{ NaCl}$ ) with an increasing concentration of  $\text{CuSO}_4$ . (b) Plot of  $A_{739}$  versus  $c(\text{Cu}^{2+})$ .

#### 4.2.4. Crystallization of CrdA

Purified samples of CrdA concentrated to  $9 \text{ mg mL}^{-1}$  in complex with  $\text{Cu(II)}$  ions and supplemented with nOG were submitted to sparse matrix crystallization trials, applying the sitting drop vapor diffusion method, using an automated crystallization platform (Oryx8 robot). Each of the 686 independent crystallization conditions (Structure screen I & II, PEG'S II SUITE, PACT SUITE, JCSG CORE I-IV) were screened using the sitting drop setup on the MRC 2-well plates. Prior to crystallization, the freshly purified protein was incubated with a 1

mmol L<sup>-1</sup> CuSO<sub>4</sub> solution and 3.8 mmol L<sup>-1</sup> nOG solution. In case of CrdAHis-tag microcrystals were observed (Fig. 42*a, b*) in the screening solution 32 of the PEG'S II SUITE (Qiagen) (0.1 mol L<sup>-1</sup> TRIS pH 8.5, 16 % PEG 4000) at 293 K. Diffraction data were measured at beamline ID23-1 of the ESRF synchrotron facility (Grenoble, France) but the small size of crystals limited the diffraction experiment (diffraction until the maximum resolution of 20 Å was observed). Other crystallization trials yielded drops containing aggregated forms or the phase-separation could be observed, so further optimization should be carried out.



**Figure 42.** (*a, b*) Microcrystals of CrdA incubated with Cu(II) ions. Picture (*b*) was captured under the microscope using a fluorescence excitation filter (CWL / BW = 450 / 50 nm).

### 4.3. HP1026 protein from *H. pylori*

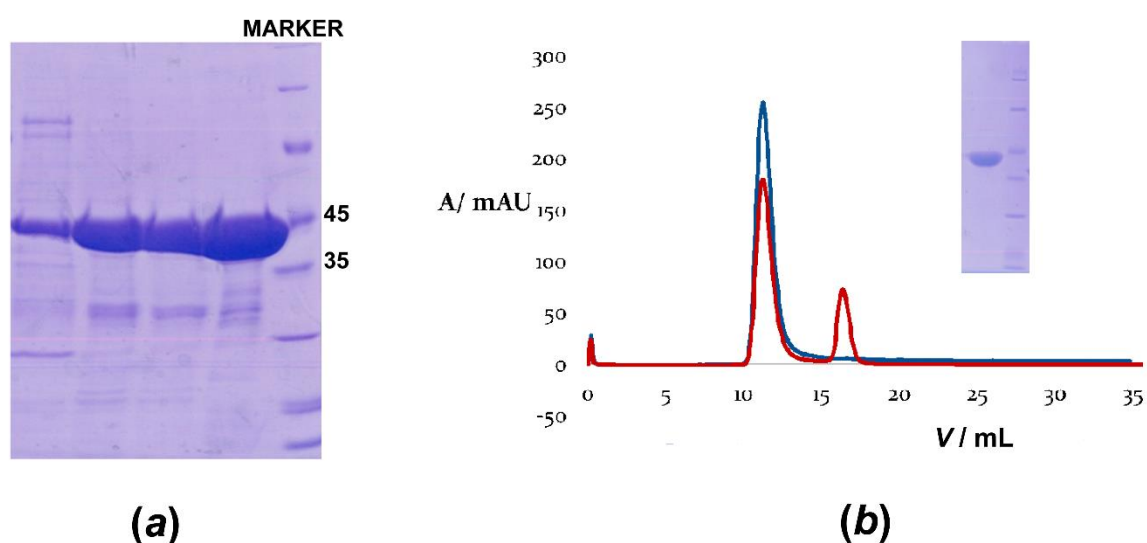
#### 4.3.1. Expression, purification and size of HP1026 in solution

His\_HP1026 protein was overproduced in *E. coli* cells by growing the cell culture at 28 °C and successfully purified by the Ni-NTA affinity chromatography followed by SEC (*Superdex 200 10/300*), Fig. 43*a, b*. The resulting His\_HP1026 from the SEC was >95% pure as clarified by SDS-PAGE gel (Fig. 43*b*) and was concentrated to more than 40 mg mL<sup>-1</sup>.

Fig. 43*b* shows that recombinant HP1026 was eluted in a single peak at the volume of 11.2 mL which corresponded to the size of 96.9 kDa. Since the size of the HP1026 monomer is 43.8



kDa the calculated molecular mass from the analytical size-exclusion chromatography analysis correlates with a dimeric species in the solution. Even though most of proteins that belong to the AAA<sup>+</sup> superfamily are present as hexamers in solution, some examples of lower oligomerization arrangement like tetrameric (*EcMgsA*),<sup>89</sup> and dimeric (*PfRFC*)<sup>147</sup> are known from the literature. Moreover, it is known for the *EcMgsA* protein and other similar clamp loader proteins that the C-terminal domain is responsible for the protein oligomerization.

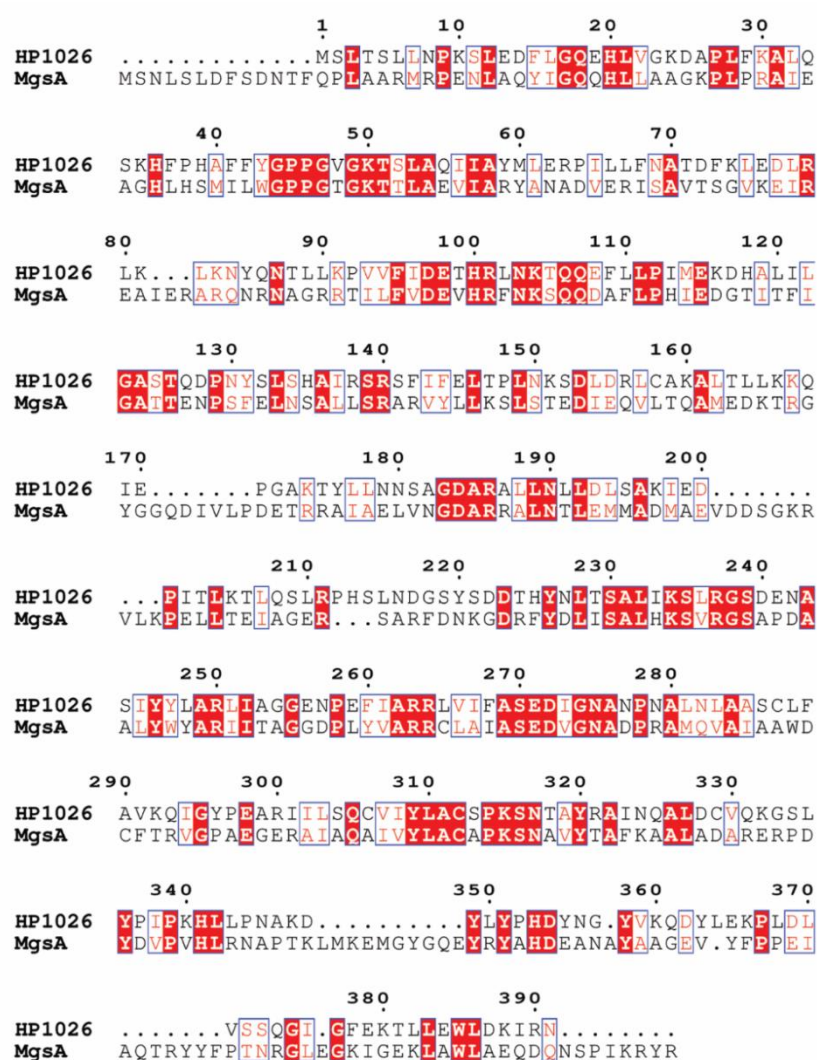


**Figure 43.** (a) SDS-PAGE of the HP1026 fractions eluted after the affinity chromatography; (b) Overlaid analytical gel filtration chromatograms of Apo\_HP1026 (blue curve; molecular mass of the monomer ~43.8 kDa) which elutes as a single peak at the volume of 11.2 mL corresponding to the molecular mass of 96.9 kDa and of HP1026 incubated with ATP-γ-S (red curve) with a major peak eluted at 11.2 mL. SDS-PAGE of a pure and concentrated fraction of HP1026 after the SEC.

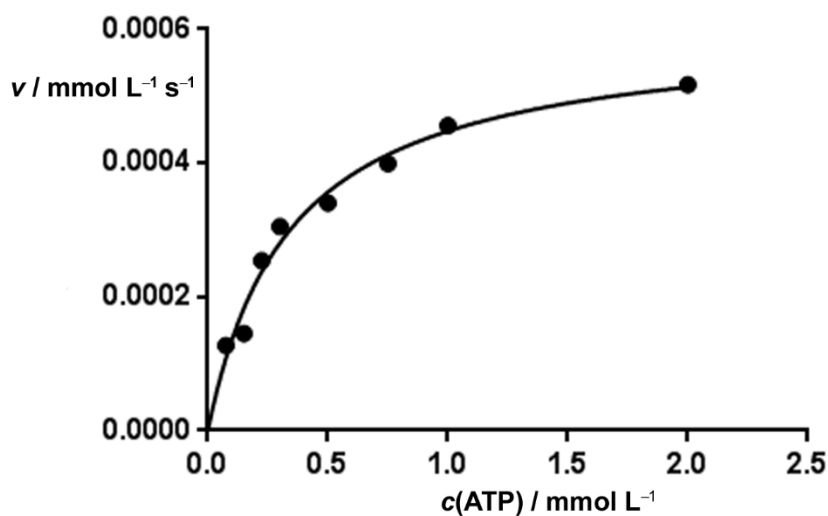
#### 4.3.2. Activity of HP1026

The HP1026 sequence contains a conserved AAA<sup>+</sup> motif (aa 36–295, Fig. 44) which shows identity  $\geq 94.2\%$  with 250 *H. pylori* proteins classified as belonging to the AAA superfamily. The highest sequence similarity is found with the *EcMgsA* protein that contains a nucleotide binding site located at the N-terminus which is strongly conserved also in the sequence of

HP1026 (aa 44–51, Fig. 44). To test whether HP1026 possesses an ATPase activity, ATP hydrolysis assay was carried. Performed experiments showed that HP1026 contains an ATPase activity (Fig. 45) within kinetic parameters as follows:  $K_m = 344 \mu\text{mol L}^{-1}$ ,  $V_{\text{max}} = 0.6019 \mu\text{mol L}^{-1} \text{ s}^{-1}$ ,  $k_{\text{cat}} = 0.02 \text{ s}^{-1}$ . HP1026 hydrolyzed ATP at a lower level ( $0.02 \text{ s}^{-1}$ ) then it is measured for the *EcMgsA* protein in the absence of DNA ( $0.056 \text{ s}^{-1}$ ).<sup>89</sup> However, in the case of *EcMgsA*, activity tests performed with addition of DNA significantly stimulated the ATPase activity,<sup>89</sup> so in our case future studies will be required to determine whether HP1026 activity is also DNA dependent.



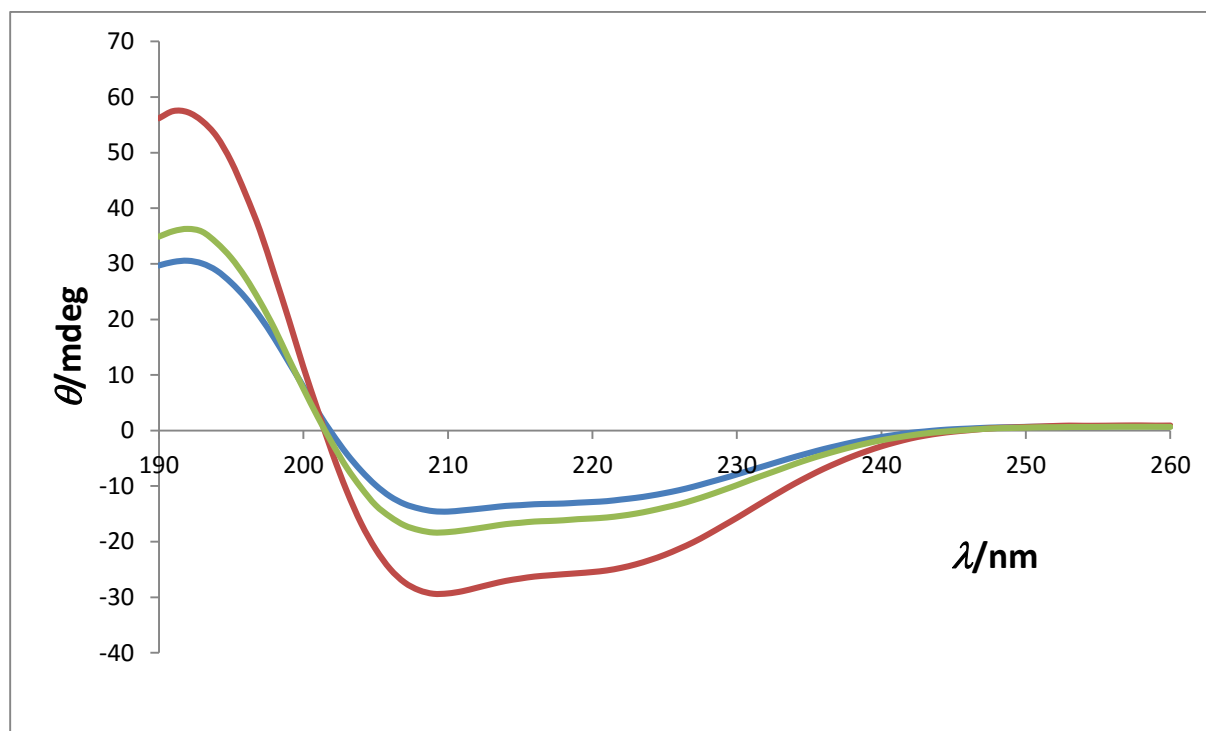
**Figure 44.** Sequence alignment of HP1026 (HP, O25670) and MgsA (EC, P0A731) using CLUSTAL OMEGA.<sup>130</sup> The alignment figure was generated using ESPrnt 3.0.<sup>136</sup> The numbering system is based on the HP1026 sequence. Residues conserved in both sequences are labelled in white and highlighted in red, while similar residues are labelled in red.



**Figure 45.** ATPase activity of HP1026. Calculated parameters by applying the Michaelis-Menten model were  $K_m = 344 \mu\text{mol L}^{-1}$ ,  $V_{\text{max}} = 0.6019 \mu\text{mol L}^{-1} \text{s}^{-1}$  and  $k_{\text{cat}} = 0.02 \text{s}^{-1}$ .

#### 4.3.3. Stability of HP1026

CD measurements showed that both Apo\_HP1026 and HP1026 protein incubated with ATP- $\gamma$ -S demonstrate the typical profile of a protein rich in  $\alpha$ -helices having negative bands at 222 nm and 208 nm and a positive band around 190 nm (Fig. 46, Table 37).<sup>108</sup> This result together with SEC profiles (Fig. 43b) indicates that addition of ATP- $\gamma$ -S did not alter the secondary structure profile nor the oligomerization profile of the HP1026 protein.



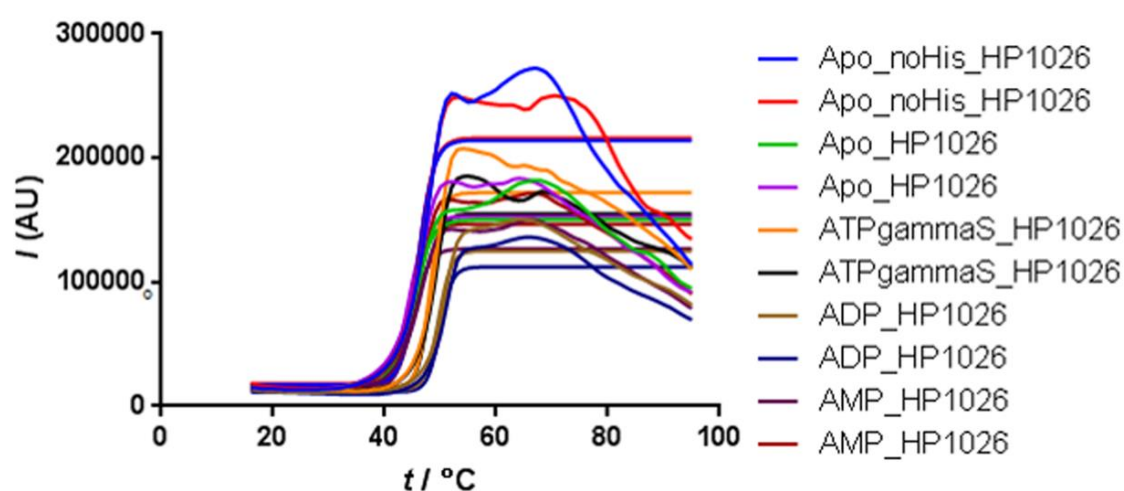
**Figure 46.** CD spectrum of the HP1026 in the far UV region (190–260 nm) presented as ellipticity in milidegrees. Apo\_HP1026, HP1026 incubated with ATP- $\gamma$ -S in a molar ratio 1:1 and HP1026 incubated with ATP- $\gamma$ -S in a molar ratio 1:2 are presented as blue, red and green curves, respectively.

**Table 37.** CD data of the HP1026 analyzed by the secondary structure analysis software, CDNN. Deconvoluted results are shown as contributions of the various components to the protein secondary structure.

SECONDARY STRUCTURE ELEMENT	%
Helix	44.5
Antiparallel $\beta$ sheet	1.1
Parallel $\beta$ sheet	8.8
$\beta$ turn	14.9
Random coil	30.7

Thermal stability of HP1026 after addition of different additives, as well as without a His-tag, was also investigated while the buffer condition was kept constant.  $T_m$  value for the HP1026:ADP, HP1026:ATP- $\gamma$ -S and HP1026:AMP complexes was at 49.75 ( $\pm$ 0.29) °C, 48.09

( $\pm 0.03$ ) °C and 44.85 ( $\pm 0.09$ ) °C, respectively (Table 38). According to these values addition of AMP didn't change the  $T_m$  value of the HP1026 protein while the removal of His-tag slightly increased  $T_m$  by +1.38 °C. From Fig. 47 and Table 38 it is evident that ATP- $\gamma$ -S and ADP strongly affected the protein stability by shifting the protein-complex denaturation to higher  $T_m$  values by +3.52 °C and +5.18 °C, respectively. The data after the ATP hydrolysis assay were another confirmation that HP1026 belongs to the class of ATPase proteins. The results of this thermofluor assay were also applied in the crystallization experiments.



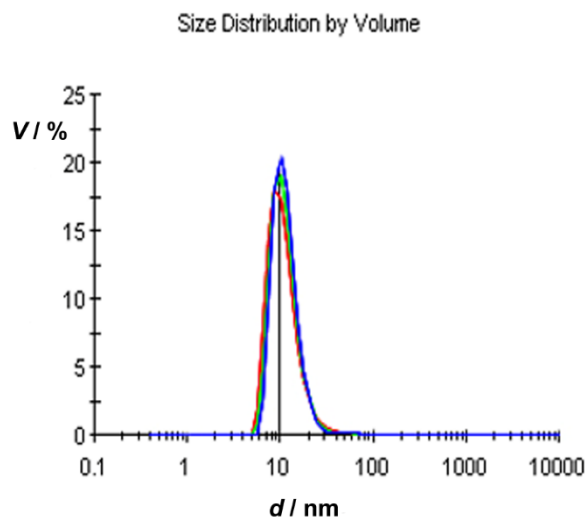
**Figure 47.** Effect of different additives and His-tag on the HP1026 thermal stability.

**Table 38.** Melting temperatures of ApoHP1026 without His-tag, ApoHP1026 and HP1026 after addition of different additives.

PROTEIN	HP1026noHis	HP1026	HP1026	HP1026	HP1026
ADDITIVE	–	–	ATP- $\gamma$ -S	ADP	AMP
$T_m$ / °C	45.95 ( $\pm 0.03$ )	44.57 ( $\pm 0.57$ )	48.09 ( $\pm 0.03$ )	49.75 ( $\pm 0.29$ )	44.85 ( $\pm 0.09$ )

#### 4.3.4. Crystallization of HP1026

Good quality of the sample for the crystallization studies was confirmed by the DLS experiment ( $PDI < 0.12$ , Figure 50) for the HP1026 protein treated with 10 % (v/v) glycerol.



**Figure 48.** The volume distributions as a function of particle diameter of HP1026 after addition of 10 % (v/v) glycerol.

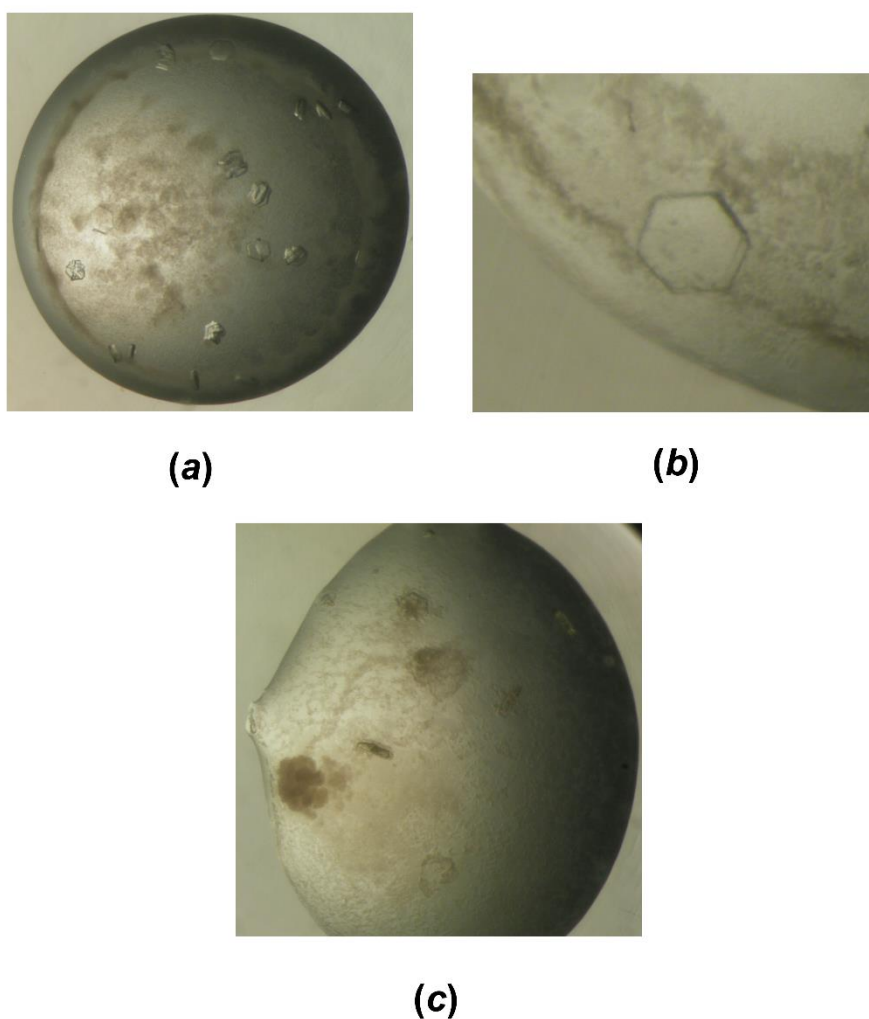
Since the HP1026 protein was very soluble ( $> 40 \text{ mg mL}^{-1}$ ) it was hard to estimate which concentration was appropriate for the crystallization trials so the solubility profile of HP1026 was explored. Using the solubility tool kit an optimal concentration of  $15 \text{ mg mL}^{-1}$  was estimated to be suitable for the crystallization experiments. Screening of different crystallization conditions was performed by using commercial screening solutions, Structure screen I & II, PEG'S II SUITE, PACT SUITE and PGA Screen, by the sitting drop vapour diffusion method using an automated crystallization platform (Oryx8 robot). Crystallization drops were prepared by mixing the protein and precipitant solutions in a ratio 1:1, and equilibrated against  $75 \mu\text{L}$  of the mother liquor in the reservoir. The initial hexagonal shaped crystals were grown in a week at  $20 \text{ }^\circ\text{C}$  from different screening solutions of the PACT SUITE (Table 39). Crystals were very sensitive and had tendency to dissolve in time (Fig. 49c). Crystals were tested at the PXIII beamline of the Swiss Synchrotron Light Source (SLS, Villigen, Switzerland) with and without cryoprotectant solutions. Cryoprotectant solutions were prepared by mixing the precipitant solution and ethylene glycol / glycerol / oil in different percentages (v/v, 20 %, 25 %, 30 %). Crystals without cryoprotectant contained ice rings and didn't diffract because they were destroyed during the freezing step in liquid nitrogen. Crystals that were frozen in the cryoprotectant solution were without ice rings but, unfortunately, diffracted only to  $15 \text{ \AA}$ . During the diffraction experiment annealing was also tried but it didn't help in ordering the molecules inside the crystal. A number of different ways were explored to

improve the diffraction quality. Microseeding resulted in too many small crystals while dehydration of the crystal, carried out by transferring the crystal from the crystallization drop into a drop of the dehydrating solution (containing 30 % of PEG 3350) which was then equilibrated against a 75  $\mu\text{L}$  volume of dehydrating solution, didn't improve crystal quality. Crystallization at lower temperature was also tested but it resulted in drops containing only precipitation.

**Table 39.** The PACT SUITE composition table of the conditions that gave crystals of HP1026.

Condition number	Composition
43	0.2 mol L <sup>-1</sup> NaCl, 0.1 mol L <sup>-1</sup> TRIS pH 8.0, 20 % PEG 6000
78	0.2 mol L <sup>-1</sup> NaOOCOH, 0.1 mol L <sup>-1</sup> BIS TRIS PROPANE pH 7.5, 20 % PEG 3350
86	0.2 mol L <sup>-1</sup> NaBr, 0.1 mol L <sup>-1</sup> BIS TRIS PROPANE pH 8.5, 20 % PEG 3350
88	0.2 mol L <sup>-1</sup> KSCN, 0.1 mol L <sup>-1</sup> BIS TRIS PROPANE pH 8.5, 20 % PEG 3350

Taking into account all of the previous results together with the results of the protein stability investigation, the crystallization trials were carried out for HP1026 without His-tag and for HP1026 in complex with ATP- $\gamma$ -S or ADP. ATP- $\gamma$ -S was used instead of ATP since it is a non-hydrolyzable ATP analogue. The HP1026 protein solution, at the final concentration of 15 mg mL<sup>-1</sup>, was incubated with ADP or ATP- $\gamma$ -S and MgCl<sub>2</sub> at the final concentrations of 2 mmol L<sup>-1</sup> and 5 mmol L<sup>-1</sup>, respectively, for 2 h prior to the crystallization experiments. HP1026 lacking the His-tag failed to give crystals while HP1026:ATP- $\gamma$ -S and HP1026:ADP gave crystals. Optimization of HP1026:ATP- $\gamma$ -S and HP1026:ADP crystals was performed by varying the protein and precipitant ratio and by screening the effect of different additives. The best quality crystals were obtained for HP1026:ADP complex in a drop containing 2.5  $\mu\text{L}$  protein and 2 or 1.5  $\mu\text{L}$  precipitant solution (0.2 mol L<sup>-1</sup> NaBr, 0.1 mol L<sup>-1</sup> Bis-tris propane pH 8.5, 20 % PEG 3350, 0.015 mmol L<sup>-1</sup> CYMAL-7) at 293(2) K (Fig. 49a,b).



**Figure 49.** Hexagonal crystals of HP1026-ADP (*a*, *b*) after 5 days, (*c*) after 3 weeks.

Since the crystals were multiple crystals they were separated before freezing in liquid nitrogen. Crystals were also tested at the PXIII beamline (SLS, Villigen, Switzerland) and diffracted until  $7 \text{ \AA}$  which is still not enough for solving the phase problem, so further improvement should be done.





## § 5. CONCLUSIONS

Crystal structures of truncated forms of *Helicobacter pylori* FlgD from two different strains in two space groups, *I422* and *P2*, have been solved, at 2.2 Å and 2.8 Å resolution, respectively. Analogously to *Pseudomonas aeruginosa* and *Xanthomonas campestris* FlgD proteins, crystallization experiments set up for the full-length protein resulted in crystals of a truncated form, lacking both N- and C-terminus ends. The crystal structures of the truncated FlgD show that the monomer is composed of a tudor and a fibronectin type III domain. The spatial orientation of the two domains in *HpFlgD* differs from that of the homologous FlgD family members, *P. aeruginosa* and *X. campestris*. Even though the full-length *HpFlgD* contains a long N-terminal region which does not show the typical features of a canonical export signal, results within this thesis, together with the bioinformatic studies, and the previously reported studies on the essential role of the N-terminal region for secretion in both *S. enterica* and *E. coli*, support the idea that the N-terminal part of the protein acts like an exporting signal. Another interesting aspect is that the C-terminal part of *HpFlgD* could represent a recognition motif for its partner. All these findings, together with the crystal structures of *HpFlgD* and the crystal structures of FlgD from other bacterial homologues, imply that functioning of FlgD requires proteolysis of the N-terminal pro-peptide. Finally, the structural differences of *HpFlgD* with respect to the homologous FlgD family members present in other bacteria suggest that significant differences exist in the molecular organization of the flagella in different bacterial species.

Recombinant CrdA protein (His/Strep/GST tagged) was successfully overexpressed. In case of CrdA\_His and Strep\_CrdA proteins high expression levels were obtained but the protein had poor solubility, so denaturation and refolding methods were implemented with success. Circular dichroism and 1D NMR confirmed the presence of secondary structure elements with a major contribution of  $\beta$  sheets (60.3 %). Dynamic light scattering showed that nOG detergent prevents protein aggregation and reduces sample polydispersity. SEC of Cu:CrdA suggests that addition of copper(II) ions stabilize formation of uniform species shifting the equilibrium toward monomeric species. UV-VIS titration with copper(II) sulphate resulted with low Cu(II) affinity ( $K_d = 1.076 \times 10^{-3} \text{ mol L}^{-1}$ ), a property of copper trafficking proteins, but it also implies that

since the CrdA protein is rich in methionines the affinity of Cu(I) should be tested. Crystallization trials gave only microcrystals that were not suitable for diffraction experiment so further optimization is needed.

HP1026 was overexpressed, purified to homogeneity and biophysically characterized. Gel filtration profile indicated the presence of dimeric species in the solution for both Apo- and ATP- $\gamma$ -S treated HP1026, so no oligomerization step was observed after the addition of ATP- $\gamma$ -S. CD analysis showed that  $\alpha$ -helices are present as major secondary structure elements (44.5 %). ATP hydrolysis assay confirmed the ATPase activity of a putative helicase like HP1026 protein within kinetics parameters as follows:  $K_m = 344 \mu\text{mol L}^{-1}$ ,  $V_{\text{max}} = 0.6019 \mu\text{mol L}^{-1} \text{s}^{-1}$ ,  $k_{\text{cat}} = 0.02 \text{s}^{-1}$ . These data together with the thermofluor assay for the first time demonstrate the role of the HP1026 protein as an ATPase. Good quality of the sample for the crystallization studies was confirmed by the DLS experiment ( $PDI < 0.12$ ). Crystallization trials were carried out for Apo\_HP1026, ATP- $\gamma$ -S:HP1026 and ADP:HP1026. The hexagonal shaped crystals were grown in a week at 20 °C. Though not of good quality for diffraction analysis, these results outline a starting point for further crystallization trials.

## § 6. LIST OF ABBREVIATIONS

ADP – Adenosine 5'-diphosphate

AMP - Adenosine 5'-monophosphate

ATP- $\gamma$ -S - Adenosine 5'-[ $\gamma$ -thio]triphosphate

APS - Ammonium persulfate

CV- Column volume

DDM – n-Dodecyl- $\beta$ -D-maltopyranoside

dNTP- Deoxynucleotides

DTT - 1,4-Dithio-D-threitol

EDTA - Ethylenediaminetetraacetic acid

HABA - 2-[4'-Hydroxy-benzeneazo] benzoic acid

HEPES - 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid

IPTG - Isopropyl- $\beta$ -D-1-thiogalactopyranoside

$K_m$  - Michaelis-Menten constant

LDAO – *N, N*-Dimethyldodecylamine *N*-oxide

MES - 2-(*N*-Morpholino)ethanesulfonic acid

MPD - 2-Methyl-2,4-pentandiol

nOG – n-Octyl- $\beta$ -D-glucopyranoside

PEG - Poly(ethylene glycol)

PES – Polyethersulfone

PMSF - Phenylmethylsulfonyl fluoride

PTFE - Poly(tetrafluoroethylene)

SDS – Sodium dodecyl sulfate

SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis

TBS – Tris-buffered saline TEMED – Tetramethylethylenediamine

TRIS-base - Tris(hydroxymethyl)aminomethane

TRIS-HCl - Tris(hydroxymethyl)aminomethane hydrochloride

$V_{max}$  - maximum enzyme velocity



## § 7. REFERENCES

1. D. Rothenbacher and H. Brenner, *Microbes. Infect.* **5** (8) (2003) 693–703.
2. J. F. Tomb, O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W.S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter, *Nature* **388** (6642) (1997) 539–547.
3. R. A. Alm, L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust, *Nature* **397** (1999) 176–180.
4. J. D. Oh, H. Kling-Backhed, M. Giannakis, J. Xu, R. S. Fulton, L. A. Fulton, H. S. Cordum, C. Wang, G. Elliott, J. Edwards, E. R. Mardis, L. G. Engstrand, and J.I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **103** (2006) 9999–10004.
5. D. A. Baltrus, M. R. Amieva, A. Covacci, T. M. Lowe, D. S. Merrell, K. M. Ottemann, M. Stein, N. R. Salama, and K. Guillemin, *J. Bacteriol.* **191** (1) (2009) 447–448.
6. S. Suerbaum and C. Josenhans, *Nat. Rev. Microbiol.* **5** (2007) 441–452.
7. M. J. Blaser, *Br. Med. J.* **316** (1998) 1507–1510.
8. R. M. Jr Peek and M. J. Blaser, *Nature Rev. Cancer* **2** (1) (2002) 28–37.
9. S. Suerbaum and P.N. Michetti, *Engl. J. Med.* **347** (2002) 1175–1186.
10. J. C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne, and P. Legrain, *Nature* **409** (2001) 211–215.
11. L. Cendron and G. Zanotti, *Febs J.* **278** (8) (2011) 1223–1231.
12. H. Zhou, M. Luo, X. Cai, J. Tang, S. Niu, W. Zhang, Y. Hu, Y. Yin, A. Huang, and D. Wang, *Proteins* **79** (7) (2011) 2346–2351.

13. O. A. Soutourina and P. N. Bertin, *FEMS Microbiol. Rev.* **27** (2003) 505–523.
14. W. T. Kuo, K. H. Chin, W. T. Lo, A. H. Wang, and S.H. Chou, *J. Mol. Biol.* **381** (1) (2008) 189–199.
15. B. Waidner, K. Melchers, F. N. Stähler, M. Kist and S. Bereswill, *J. Bacteriol.* **187** (2005) 4683–4688.
16. B. Waidner, K. Melchers, I. Ivanov, H. Loferer, K. W. Bensch, M. Kist, and S. Bereswill, *J. Bacteriol.* **184** (2002) 6700–6708.
17. D. Roncarati, A. Danielli, and V. Scarlato, *J. Bacteriol.* **193** (20) (2011) 5629–5636.
18. C. S. Goodwin, M. M. Mendall, and T. C. Northfield, *Lancet.* **349** (1997) 265–269.
19. J. G. Kusters, A. H. van Vliet, and E.J. Kuipers, *Clin. Microbiol. Rev.* **19** (2006) 449–490.
20. B. J. Marshall and J.R. Warren, *Lancet.* **1** (1984) 1311–1315.
21. A. Covacci, J. L. Telford, G. Del Giudice, J. Parsonnet, and R. Rappuoli. *Science* **284** (1999) 1328–1333.
22. D. Kersulyte, A. K. Mukhopadhyay , B. Velapatiño, W. Su, Z. Pan, C. Garcia, V. Hernandez, Y. Valdez, R. S. Mistry, R. H. Gilman, Y. Yuan, H. Gao, T. Alarcón, M. López-Brea, G. Balakrish Nair, A. Chowdhury, S. Datta, M. Shirai, T. Nakazawa, R. Ally, I. Segal, B. C. Wong, S. K. Lam, F. O. Olfat, T. Borén, L. Engstrand, O. Torres, R. Schneider, J. E. Thomas, S. Czinn, and D. E. Berg, *J. Bacteriol.* **182** (11) (2000) 3210–3218.
23. H. Yoshiyama and T. Nakazawa, *Microbes Infect.* **2** (2000) 55–60.
24. F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao, *Science* **277** (1997) 1453–1462.
25. C. T. Parker, S. Huynh, L. Gorski, K. K. Cooper, and W. G. Miller, *Genome Announc.* **3** (6) (2015) e01365–15.
26. J. G. Kusters, A. H. M. van Vliet, and Ernst J. Kuipers, *Clin. Microbiol. Rev.* **19** (2006) 449–490.
27. R. D. Leunk, P. T. Johnson, B. C. David, W. G. Kraft, and D. R. Morgan, *J. Med. Microbiol.* **26** (1988) 93–99.

28. N. S. Akopyants, S. W. Clifton, D. Kersulyte, J. E. Crabtree, B. E. Youree, C. A. Reece, N. O. Bukanov, E. S. Drazek, B. A. Roe, and D. E. Berg, *Mol. Microbiol.* **28** (1998) 37–53.
29. S. C. Lange Censini, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci, *Proc. Natl. Acad. Sci. USA* **93** (1996) 14648–14653.
30. M. Asahi, T. Azuma, S. Ito, Y. Ito, H. Suto, Y. Nagai, M. Tsubokawa, Y. Tohyama, S. Maeda, M. Omata, T. Suzuki, and C. Sasakawa, *J. Exp. Med.* **191** (2000) 593–602.
31. P. J. Christie and J. P. Vogel, *Trends Microbiol.* **8** (2000) 354–360.
32. L. Terradot and G. Waksman, *FEBS J.* **278** (2011) 1213–1222.
33. S. Backert, N. Tegtmeyer, and M. Selbach, *Helicobacter* **15** (2010) 163–176.
34. M. Hatakeyama, *Curr. Opin. Microbiol.* **11** (2008) 30–37.
35. J. Viala, C. Chaput, I. G. Boneca, A. Cardona, S. E. Girardin, A. P. Moran, R. Athman, S. Memet, M. R. Huerre, A. J. Coyle, P. S. DiStefano, P. J. Sansonetti, A. Labigne, J. Bertin, D. J. Philpott, and R. L. Ferrero, *Nat. Immunol.* **5** (2004) 1166–1174.
36. T. L. Cover and M. J. Blaser, *J. Biol. Chem.* **267** (1992) 10570–10575.
37. J. C. Atherton, P. Cao, R. M. Peek, Jr., M. K. Tummuru, M. J. Blaser, and T. L. Cover, *J. Biol. Chem.* **270** (1995) 17771–17777.
38. K. Ogura, S. Maeda, M. Nakao, T. Watanabe, M. Tada, T. Kyutoku, H. Yoshida, Y. Shiratori, and M. Omata, *J. Exp. Med.* **192** (2000) 1601–1610.
39. D. Falush, C. Kraft, N. S. Taylor, P. Correa, J. G. Fox, M. Achtman, and S. Suerbaum, *Proc. Natl. Acad. Sci. USA* **98** (2001) 15056–15061.
40. E. J. Kuipers, D. A. Israel, J. G. Kusters, M. M. Gerrits, J. Weel, A. Van Der Ende, R. W. van Der Hulst, H. P. Wirth, J. Hook-Nikanne, S. A. Thompson, and M. J. Blaser, *J. Infect. Dis.* **181** (2000) 273–282.
41. R. A. Burne and Y. M. Chen, *Microbes Infect.* **2** (2000) 533–542.
42. H. L. Mobley, M. D. Island and R. P. Hausinger, *Microbiol. Rev.* **59** (1995) 451–480.
43. S. H. Phadnis, M. H. Parlow, M. Levy, D. Ilver, C. M. Caulkins, J. B. Connors, and B. E. Dunn, *Infect. Immun.* **64** (1996) 905–912.
44. D. R. Scott, D. Weeks, C. Hong, S. Postius, K. Melchers, and G. Sachs, *Gastroenterology* **114** (1998) 58–70.
45. D. L. Weeks, S. Eskandari, D. R. Scott, and G. Sachs, *Science* **287** (2000) 482–485.



46. N. C. Ha, S. T. Oh, J. Y. Sung, K. A. Cha, Hyung Lee, M., and B. H. , *Nat. Struct. Biol.* **8** (2001) 505–509.
47. S. McNicholas, E. Potterton, K. S. Wilson, and M. E. M. Noble, *Acta Crystallogr. D Biol. Crystallogr.* **67** (2011) 386–394.
48. T. Boren, P. Falk, K. A. Roth, G. Larson, and S. Normark, *Science* **262** (1993) 1892–1895.
49. M. Gerhard, N. Lehn, N. Neumayer, T. Boren, R. Rad, W. Schepp, S. Miehlke, M. Classen, and C. Prinz, *Proc. Natl. Acad. Sci. USA* **96** (1999) 2778–12783.
50. H. Ota, J. Nakayama, M. Momose, M. Hayama, T. Akamatsu, T. Katsuyama, D. Y. Graham, and R. M. Genta, *Virchows Arch.* **433** (1998) 419–426.
51. R. de Jonge, R. G. Pot, R. J. Loffeld, A. H. van Vliet, E. J. Kuipers, and J. G. Kusters, *Helicobacter* **9** (2004) 158–164.
52. P. W. O’Toole, M. C. Lane, and S. Porwollik, *Microbes Infect.* **2** (2000) 1207–1214.
53. M. R. Amieva and E. M. El-Omar, *Gastroenterology* **134** (2008) 306–323.
54. C. Josenhans and S. Suerbaum, *Int. J. Med. Microbiol.* **291** (2002) 605–614.
55. P. Lertsethtakarn, K.M. Ottemann, and D. R. Hendrixson, *Annu. Rev. Microbiol.* **65** (2011) 389–410.
56. K. M. Ottemann and A. C. Lowenthal, *Infect. Immun.* **70** (2002) 1984–1990.
57. D. R. Thomas, N. R. Francis, C. Xu, and D. J. DeRosier, *J. Bacteriol.* **188** (2006) 7039–7048.
58. F. F. V. Chevance and K. T. Hughes, *Nat. Rev. Microbiol.* **6** (2008) 455–465.
59. K. Paul, G. Gonzalez-Bonet, A. M. Bilwes, B. R. Crane, and D. Blair, *EMBO J.* **30** (2011) 2962–2971.
60. I. Pulić, V. Loconte, G. Zanotti, *Am. J. Biochem. Biotechnol.* **10** (3) (2014) 143–161.
61. S. Roure, M. Bonis, C. Chaput, C. Ecobichon, A. Mattox, C. Barrière, N. Geldmacher, S. Guadagnini, C. Schmitt, M. C. Prévost, A. Labigne, S. Backert, R. L. Ferrero, and I. G. Boneca, *Mol. Microbiol.* **86** (2012) 845–856.
62. F. A. Samatey, H. Matsunami, K. Imada, S. Nagashima, T. R. Shaikh, D. R. Thomas, J. Z. Chen, D. J. DeRosier, A. Kitao, and K. Namba, *Nature* **431** (2004) 1062–1068.
63. R. M. Macnab, *Annu. Rev. Microbiol.* **57** (2003) 77–100.
64. N. Kamal, N. Dorrell, A. Jagannathan, S. M. Turner, C. Constantinidou, D. J.

- Studholme, G. Marsden, J. Hinds, K. G. Laing, B. W. Wren, and C. W. Penn, *Microbiology* **153** (2007) 3099–3111.
65. K. A. Ryan, N. Karim, M. Worku, C. W. Penn, and P.W. O'Toole, *J. Bacteriol.* **187** (2005) 5742–5750.
66. M. Erhardt, T. Hirano, Y. Su, K. Paul, D. H. Wee, S. Mizuno, S. I. Aizawa, and K. T. Hughes, *Mol. Microbiol.* **75** (5) (2010) 1272–1284.
67. T. Kubori, N. Shimamoto, S. Yamaguchi, K. Namba, and S. Aizawa, *J. Mol. Biol.* **226** (1992) 433–446.
68. K. Ohnishi, Y. Ohto, S. Aizawa, R. M. Macnab, and T. Iino, *J. Bacteriol.* **176** (1994) 2272–2281.
69. N. Moriya, T. Minamino, K. Imada, and K. Namba, *Microbiology* **157** (2011) 1354–1362.
70. C. Weber-Sparenberg, P. Pöplau, H. Brookman, M. Rochón, C. Möckel, M. Nietschke, and H. Jung, *Arch. Microbiol.* **186** (2006) 307–316.
71. S. L. Benoit, A. L. Zbell, and R. J. Maier, *Microbiology* **153** (2007) 3748–3756.
72. N. Mehta, J. W. Olson, and R. J. Maier, *J. Bacteriol.* **185** (2003) 726–734.
73. J. Velayudhan, N. J. Hughes, A. A. McColm, J. Bagshaw, C. L. Clayton, S. C. Andrews, and D. J. Kelly, *Mol. Microbiol.* **37** (2000) 274–286.
74. B. Waidner, S. Greiner, S. Odenbreit, H. Kavermann, J. Velayudhan, F. Stahler, J. Guhl, E. Bisse, A. H. M. van Vliet, S. C. Andrews, J. G. Kusters, D. J. Kelly, R. Haas, M. Kist, and S. Bereswill, *Infect. Immun.* **70** (2002) 3923–3929.
75. G. L. Mendz, E. M. Holmes, and R. L. Ferrero, *Biochim. Biophys. Acta* **1388** (1998) 465–477.
76. D. Bayle, S. Wangler, T. Weitzenegger, W. Steinhilber, J. Volz, M. Przybylski, K. P. Schafer, G. Sachs, and K. Melchers, *J. Bacteriol.* **180** (1998) 317–329.
77. S. Bereswill, S. Greiner, A. H. M. van Vliet, B. Waidner, F. Fassbinder, E. Schiltz, J. G. Kusters, and M. Kist, *J. Bacteriol.* **182** (2000) 5948–5953.
78. L. Zhang, M. Koay, M. J. Maher, Z. Xiao, and A.G. Wedd, *J. Am. Chem. Soc.* **128** (2006) 5834–5850.
79. T. J. Lawton, G. E. Kenney, J. D. Hurley, and A. C. Rosenzweig, *Biochemistry* (2016), doi: 10.1021/acs.biochem.6b00175
80. D. Ronacarti, A. Danielli, G. Spohn, I. Delany, and V. Scarlato, *J. Bacteriol.* **189** (2007)

- 7234–7243.
81. D. J. Evans, Jr. D. G. Evans, L. Engstrand, and D. Y. Graham, *Infect. Immun.* **60** (1992) 2125–2127.
82. G. Spohn, A. Danielli, D. Roncarati, I. Delany, R. Rappuoli, and V. Scarlato, *J. Bacteriol.* **186** (2004) 2956–2965.
83. A. Chastanet, J. Fert, and T. Msadek, *Mol. Microbiol.* **47** (2003) 1061–1073.
84. C. Ueguchi, M. Kakeda, H. Yamada, and T. Mizuno, *Proc. Natl. Acad. Sci. USA* **91** (1994) 1054–1058.
85. G. Spohn and V. Scarlato, *Mol. Microbiol.* **34** (1999) 663–674.
86. T. Ogura and A. J. Wilkinson, *Genes Cells* **6** (2001) 575–597.
87. J. Snider, G. Thibault, and W. A. Houry, *Genome Biol.* **9** (2008) 216.1–216.8.
88. J. P. Erzberger, M. L. Mott, and J. M. Berger, *Nat. Struct. Mol. Biol.* **13** (2006) 676–683.
89. A. N. Page, N. P. George, A. H. Marceau, M. M. Cox, and J. L. Keck, *J. Biol. Chem.* **286** (2011) 12075–12085.
90. A. Seybert, M. R. Singleton, N. Cook, D. R. Hall, and D. B. Wigley, *EMBO J.* **25** (2006) 2209–2218.
91. J. P. Erzberger, M. L. Mott, and J. M. Berger, *Nat. Struct. Mol. Biol.* **13** (2006) 676–683.
92. A. Bazin, M. el V. Cherrier, I. Gutsche, J. Timmins, and L. Terradot, *Nucleic Acids Res.* **43** (2015) 8564–8576.
93. pET System Manual, 11. edition, Novagen (Merck), 2007.
94. <https://www.gelifesciences.com> (downloaded on March 4<sup>th</sup>, 2016.)
95. <https://www.lucigen.com> (downloaded on March 4<sup>th</sup>, 2016.)
96. <https://www.thermofisher.com> (downloaded on March 4<sup>th</sup>, 2016.)
97. One shot BL21(DE3), BL21(DE3)pLysS, BL21(DE3)pLysE competent cells Manual, part 28-0812, Invitrogen (Life technologies), 2010.
98. J. Sambrook and D. Russell: *Molecular Cloning: A Laboratory Manual*, 3<sup>th</sup> edition, Cold Spring Harbor Laboratory Press (2001).
99. <https://www.thermofisher.com/hr/en/home/life-science/dna-rna-purification-analysis/nucleic-acid-gel-electrophoresis/dna-stains/etbr.html> (downloaded on March 5<sup>th</sup>, 2016.)
100. Expresso T7 Cloning and Expression System Manual, Lucigen, 2014.

101. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, and A. Bairoch; *Protein Identification and Analysis Tools on the ExPASy Server*; (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press (2005). pp. 571–607.
102. <http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Bulletin/protsil2bul.pdf> (downloaded on March 5<sup>th</sup>, 2016.)
103. Z. Li, M. Nimtz, and U. Rinas, *Appl Microbiol Biotechnol* **92** (2011) 823–833.
104. Tools for protein expression and purification, IBA.
105. [http://scd-theses.u-strasbg.fr/776/01/html/these\\_body.html](http://scd-theses.u-strasbg.fr/776/01/html/these_body.html) (downloaded on March 9<sup>th</sup>, 2016.)
106. <https://www.bioke.com> (downloaded on March 9<sup>th</sup>, 2016.)
107. L. F. Vallejo and U. Rinas, *Microb Cell Fact* **3:11** (2004) 1–12.
108. S. M. Kelly and N. C. Price, *Biochim. Biophys. Acta* **1338** (1997) 161–185.
109. <http://web.nmsu.edu/~kburke/Instrumentation/CD1.html> (downloaded on March 10<sup>th</sup>, 2016.)
110. G. Bohm, R. Muhr, and R. Jaenicke, *Protein Eng.* **5** (1992) 191–195.
111. <http://www.lsinstruments.ch> (downloaded on March 11<sup>th</sup>, 2016.)
112. B. Lorber, F. Fischer, M. Bailly, H. Roy, and D. Kern, *Biochem. Mol. Biol. Educ.* **40** (6) (2012) 372–382.
113. S. Boivin, S. Kozak and R. Meijers, *Protein Expres. Purif.* **91** (2013) 192–206.
114. K. Samizo, R. Ishikawa, A. Nakamura, and K. Kohama, *Anal. Biochem.* **293** (2001) 212–215.
115. GraphPad Prism version 6.00 for Windows, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com)
116. B. Rupp, *Protein crystallization, Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, Taylor and Francis Group, New York, USA, 2010, pp. 77–137.
117. T. G. G. Battye, L. K. Ontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie, *Acta Crystallogr D Biol Crystallogr* **67** (2011) 271–281.
118. W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **66** (2010) 133–144.
119. P. Evans, *Acta Crystallogr D Biol Crystallogr* **62** (2006) 72–82.
120. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M.

- Keegan, E. B. Rissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson, *Acta Crystallogr D Biol Crystallogr* **67** (2011) 235–242.
121. G. M. Sheldrick, *Acta Crystallogr A Found Crystallogr* **64** (2008) 112–122.
122. K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **62** (2006) 1002–1011.
123. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. Mccoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P.H. Zwart, *Acta Crystallogr D Biol Crystallogr* **66** (2010) 213–221.
124. P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **66** (2010) 486–501.
125. A. A. Lebedev, A. A. Vagin, and G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr* **64** (2007) 733–39.
126. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, and A.A. Vagin, *Acta Crystallogr D Biol Crystallogr* **67** (2011) 355–367.
127. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, *Nucleic Acids Res.* **25** (1997) 3389–3402.
128. The UniProt Consortium, The Universal Protein Resource (UniProt): a hub for protein information. *Nucleic Acids Res.* **43** (2015) D204–D212.
129. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins, *Bioinformatics* **23** (2007) 2947–2948.
130. H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, and R. Lopez. *Nucleic acids research* **41** (2013) (Web Server issue) W597–600.
131. D.T. Jones, *J. Mol. Biol.* **292** (1999) 195–202.
132. A. S. Juncker, H. Willenbrock, G. von Heijne, H. Nielsen, S. Brunak, and A. Krogh, *Protein Sci.* **12** (2003) 1652–1662.
133. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P.E. Bourne, *Nucleic Acids Research*, **28** (2000) 235–242.
134. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC
135. E. Krissinel and K. Henrick, *J. Mol. Biol.* **372** (2007) 774–797.

136. X. Robert, P. Gouet, *Nucleic Acids Res.* **42** (2014) W320–324.
137. C. D. Dickinson, B. Veerapandian, X. P. Dai, R. C. Hamlin, N. H. Xuong, E. Ruoslahti, and K. R. Ely, *J. Mol. Biol.* **236** (1994) 1079–1092.
138. C. K. Brown, Z. Y. Gu, Y. V. Matsuka, S. S. Purushothaman, L. A. Winter, P. P. Cleary, S. B. Olmsted, D. H. Ohlendorf, and C. A. Earhart, *Proc. Natl. Acad. Sci. USA* **102** (2005) 18391–18396.
139. N. Tegtmeyer, R. Hartig, R. M. Delahay, M. Rohde, S. Brandt, J. Conradi, S. Takahashi, A. J. Smolka, N. Sewald, and S. Backert, *J. Biol. Chem.* **285** (30) (2010) 23515–23526.
140. J. D. Dubreuil, G. Del Giudice, and R. Rappuoli, *Microbiol. Mol. Biol. R.* **66** (4) (2002) 617–629.
141. C. R. Courtney, L. M. Cozy, and D. B. Kearns, *J. Bacteriol.* **194** (2012) 4619–4629.
142. D. Bumann, S. Aksu, M. Wendland, K. Janek, U. Zimny-Arndt, N. Sabarth, T. F. Meyer, and P. R. Jungblut, *Infect. Immun.* **70** (2002) 3396–3403.
143. G. Zanotti, *World J. Gastroenterol.* **20** (2014) 1402–1423.
144. H. M. Singer, C. Kühne, J. A. Deditius, K. T. Hughes, and M. Erhardt, *J. Bacteriol.* **196** (2014) 1448–1457.
145. A. G. Evdokimov, J. Phan, J. E. Tropea, K. M. Routzahn, H. K. Peters, M. Pokross, and D. S. Waugh, *Nat. Struct. Biol.* **10** (2003) 789–793.
146. J. T. Rubino and K. J. Franz. *J. Inorg. Biochem.* **107** (2012) 129–143.
147. T. Oyama, Y. Ishino, I. K. O. Cann, S. Ishino, and K. Morikawa, *Mol. Cell.* **8** (2001) 455–463.



## § 8. CURRICULUM VITAE

*Name:* Ivana Kekez (Pulić)

*Date of birth:* May 14, 1986

*Place of birth:* Zagreb, Croatia

### **Academic Degree:**

2009 Diploma Degree in Chemistry and Biology, University of Zagreb, Faculty of Science, Croatia

Diploma Thesis: „Studies on the reaction between hesperetin and titanium(IV) with spectrophotometric method“ (original in Croatian; *supervisor:* Dr. Vlasta Vojković, Associate Professor)

### **Education:**

2011 – now Dr. Sc. student in Chemistry (field: Inorganic and Structural Chemistry), University of Zagreb, Faculty of Science, Croatia (*supervisors:* Dr. D. Matković-Čalogović, Professor and Dr. G. Zanotti, Professor)

2004 – 2009 Undergraduate student in Chemistry and Biology, University of Zagreb, Faculty of Science, Croatia

2000 – 2004 High school (*VII. gimnazija*, Zagreb, Croatia)

1992 – 2000 Elementary school (Zagreb, Croatia)

### **Working experience:**

2011 – now Research / Teaching Assistant, University of Zagreb, Faculty of Science, Division of General and Inorganic Chemistry, Croatia

2009 – 2011 Teacher of Chemistry and Biology in high schools: “Lucijan Vranjanin” and “Gornjogradska gimnazija” in Zagreb, Croatia

### **Teaching:**

2011 – now Teaching Assistant for undergraduate courses (University of Zagreb, Faculty of Science, Division of General and Inorganic Chemistry)

---



- Laboratory Exercises in General and Inorganic Chemistry (*for students in Biology*),
- Laboratory Exercises in Chemistry 1 (*for students in Geology*),
- Laboratory Exercises in General Chemistry 1 (*for students in Chemistry, Biology and Chemistry*),
- Laboratory Exercises in General Chemistry 2 (*for students in Chemistry, Biology and Chemistry*),
- Introduction to Inorganic Chemistry (*laboratory exercises for students in Environmental Science*),
- General and Inorganic Chemistry (*seminar for students in Physics, Geophysics, Physics and Polytechnics, Physics and Informatics*),
- Chemistry 1 (*seminar for students in Geology*),
- General and Inorganic Chemistry (*seminar for students in Biology*)
- Methods in Chemistry Teaching (*seminar for students in Biology and Chemistry, Chemistry, Physics and Chemistry*)

Trained and directed 1 undergraduate student during his research for diploma thesis.

**Projects – associate:**

- |                                |   |
|--------------------------------|---|
| June 2015 –<br>now             | Research Assistant in project financed by the Croatian Science Foundation “Essential metal ions in <i>H. pylori</i> proteins and model complexes- structure and function/property” ( <i>supervisor</i> : Dr. D. Matković-Čalogović, Professor) at the University of Zagreb, Faculty of Science, Department of Chemistry             |
| September<br>2013 – 2016       | Research Assistant in PRIN project financed by the Republic of Italy “Structural Characterization of the Pathogenic Factors from <i>H. pylori</i> ” at the University of Padua, Department of Biomedical Sciences ( <i>supervisor</i> : Dr. G. Zanotti, Professor)  |
| May 2012 –<br>December<br>2012 | Research Assistant in project financed by the Croatian Science Foundation “Structural studies on the [NiFe]hydrogenase maturation system from <i>Helicobacter pylori</i> ” at the University of Padua, Department of Biomedical Sciences ( <i>supervisors</i> : Dr. G. Zanotti, Professor and Dr. D. Matković-Čalogović, Professor) |
| 2011 – 2013                    | Research Assistant in project “Structural studies on biological macromolecules by X-ray diffraction” ( <i>supervisor</i> : Dr. D. Matković-Čalogović, Professor) at the University of Zagreb, Faculty of Science, Department of Chemistry   |
-

**Awards:**

- 2012 Grant awarded by the Croatian Science Foundation for working on the project “Structural studies on the [NiFe]hydrogenase maturation system from *Helicobacter pylori*” at the University of Padua, Department of Biomedical Sciences
- 2009 Medal awarded by the Department of Chemistry for outstanding achievements throughout the studies
- 2007 – 2008 Fellowship awarded for the gifted students by the City of Zagreb
- 2005 – 2007 Fellowship awarded for the gifted students in the A category by the Ministry of Science, Education and Sports

**Additional Training:**

- 2015 Symposium & Workshop on Microscale Thermophoresis, University of Zagreb, Faculty of Science, Zagreb, Croatia, June 30 – July 1, 2015
- 2012 – 2014 1 year and 3 months spent at the University of Padua, Department of Biomedical Sciences (Prof. Zanotti’s group)
- 2014 1<sup>st</sup> European Crystallography School, Pavia, Italy, August 28 – September 6, 2014
- International School of Crystallography, 47<sup>th</sup> Course, Structural Basis of Pharmacology: Deeper Understanding of Drug Discovery through Crystallography, Erice, Italy, May 30 – June 8, 2014
- 2012 3<sup>rd</sup> Granada International School of Crystallization, Granada, Spain, May 21 – 25, 2012
- 2011 XI School on Synchrotron Radiation: Fundamentals, Methods and Applications, Duino, Italy, September 5–16, 2011

**Membership in Professional Affiliations:**

- 2009 – now Croatian Chemical Society
- 2011 – now Croatian Crystallographic Association

**Main Scientific Interests:**

Molecular cloning, gene expression and protein purification (affinity chromatography, size-exclusion chromatography, ionic-exchange), biophysical protein characterization (UV-Vis spectroscopy and fluorimetry, CD, DLS, ITC, NMR), protein crystallization, protein X-ray crystallography, small molecule X-ray crystallography.

---

**Scientific Publications**

5. I. Pulić, L. Cendron, M. Salamina, P. Polverino de Laureto, D. Matković-Čalogović, G. Zanotti. Crystal structure of truncated FlgD from the human pathogen *Helicobacter pylori*. *J. Struct. Biol.* **194** (2) (2016) 147–155.
4. V. Vrdoljak, B. Prugovečki, I. Pulić, M. Cigler, D. Sviben, J. Parlov Vuković, P. Novak, D. Matković-Čalogović, M. Cindrić. Dioxidomolybdenum(VI) complexes with isoniazid-related hydrazones: solution-based, mechanochemical and UV-light assisted deprotonation. *New J. Chem.* **39** (2015) 7322–7332.
3. I. Pulić, V. Loconte, G. Zanotti. Structural Characterization at the Atomic Level of a Molecular Nano-Machine: The state of the art of *Helicobacter pylori* flagellum organization. *Am. J. Biochem. Biotechnol.* **10** (3) (2014) 143–161.
2. B. Foretić, I. Picek, V. Damjanović, D. Cvijanović, I. Pulić, B.-M. Kukovec, D. Matković-Čalogović. Spectroscopic and structural insights into N-substituted pyridinium-4-aldoximes and their pentacyanoferrate(II) complexes. *Polyhedron.* **52** (2013) 733–742.
1. B. Prugovečki, I. Pulić, M. Toth, D. Matković-Čalogović. High resolution structure of the manganese derivative of insulin. *Croat. Chem. Acta.* **85** (4) (2012) 435–439.

**Oral Presentation at Scientific Meetings:**

1. I. Pulić, L. Cendron, D. Matković-Čalogović, G. Zanotti, *Structural characterization of the metal homeostasis factors from Helicobacter pylori*, XXIII Croatian Meeting of Chemists and Chemical Engineers, Osijek, Croatia, April 21 – 24, 2013, Abstr. 41.

**Poster Presentations at Scientific Meetings:**

10. I. Pulić, G. Zanotti, D. Matković-Čalogović, *Cloning and expression of different constructs of HP1326 from the bacterium Helicobacter pylori*, XI Meeting of Young Chemical Engineers, Zagreb, Croatia, February 18 – 19, 2016, Abstr. 113.
  9. I. Pulić, L. Cendron, M. Salamina, D. Matković-Čalogović, G. Zanotti, *Crystal structure of truncated FlgD from the human pathogen Helicobacter pylori*, 29th European Crystallographic Meeting, Rovinj, Croatia, August 23 – 28, 2015, *Acta Crystallogr. Sect. A* **71** (2015) s253.
  8. I. Pulić, L. Cendron, M. Kekez, G. Zanotti, D. Matković-Čalogović, *Characterization of CrdA – the copper homeostasis factor from bacterium Helicobacter pylori*, Symposium & Workshop on Microscale Thermophoresis, University of Zagreb, Faculty of Science, Zagreb, Croatia, June 30 – July 1, 2015, Abstr. 27.
  7. I. Pulić, L. Cendron, M. Salamina, G. Zanotti, D. Matković-Čalogović, *Structural studies of FlgD - flagella hook capping protein from Helicobacter pylori*, 1<sup>st</sup> European Crystallography School, Pavia, Italy, August 28 – September 6, 2014.
-

6. I. Pulić, L. Cendron, G. Zanotti, D. Matković-Čalogović, *Structural and functional characterization of ATP binding HP1026 - Helicobacter pylori heat shock protein*, International School of Crystallography, 47<sup>th</sup> Course, Structural Basis of Pharmacology: Deeper Understanding of Drug Discovery through Crystallography, Erice, Italy, May 30 – June 8, 2014, Abstr. 271.
5. I. Pulić, M. Salamina, L. Cendron, G. Zanotti, M. Kekez, D. Matković- Čalogović, *Cloning, purification and crystallization of FlgD and CrdA – Helicobacter pylori survival required proteins*, Synthetic Life: Molecules, Cells and Tissues, Rehovot, Israel, October 13 – 16, 2013.
4. I. Pulić, L. Cendron, D. Matković-Čalogović, G. Zanotti, *Structural studies about HypB and CrdA, two Helicobacter pylori metal homeostasis key factors*, XI Annual Retreat - Venetian Institute of Molecular Medicine - VIMM (Advanced Biomedical Research Foundation), Marostica, Italy, October 19 – 20, 2012, Abstr. 42.
3. I. Pulić, B. Prugovečki, A. Jurković, D. Matković-Čalogović, *Crystallization of insulin derivatives*, 3<sup>rd</sup> Granada International School of Crystallization, Granada, Spain, May 21 – 25, 2012, Abstr. 123.
2. I. Pulić, B. Prugovečki, D. Matković-Čalogović, *The crystal structure of vanadium insulin derivative*, IX Meeting of Young Chemical Engineers, Zagreb, Croatia, February 16 – 17, 2012, Abstr. 163.
1. I. Pulić, B. Prugovečki, D. Matković-Čalogović, *High resolution structure of the Mn derivative of insulin*, XI School on Synchrotron Radiation: Fundamentals, Methods and Applications, Duino, Italy, September 5–16, 2011.

**Co-authored Contributions, but not the Presenting Author:**

4. E. Šimunić, S. Vuk, A. Skuhala, N. Smrečki, M. Babić, O. Jović, I. Pulić, D. Matković-Čalogović, Z. Popović, *Biologically active copper(II) complexes with N-alkyliminodiacetamide derivatives*, XI Meeting of Young Chemical Engineers, Zagreb, Croatia, February 18 – 19, 2016, Abstr. 100.
  3. I. Tkalčec, K. Bukvić, N. Smrečki, B.-M. Kukovec, I. Pulić, D. Matković-Čalogović, Z. Popović, *Reactions of nickel(II) and cobalt(II) salts with iminodiacetamide derivatives*, XI Meeting of Young Chemical Engineers, Zagreb, Croatia, February 18 – 19, 2016, Abstr. 103
  2. B. Prugovečki, V. Vrdoljak, M. Cindrić, I. Pulić, D. Matković- Čalogović, *Synthesis and structural studies of dioxomolybdenum(VI) complexes with isoniazid related hydrazones*, 29th European Crystallographic Meeting, Rovinj, Croatia, August 23 – 28, 2015, *Acta Crystallogr. Sect. A* **71** (2015) s420–421.
  1. I. Pavličić, R. Biba, N. Smrečki, I. Pulić, D. Matković-Čalogović, Z. Popović, *Co(II), Ni(II) and Cu(II) complexes with N-cyclohexyliminodiacetamide*, XXIV Croatian Meeting of Chemists and Chemical Engineers, Zagreb, Croatia, April 21 – 24, 2015, Abstr. 98.
-