Chapter 1 - Introduction

1.1 Protein structure and function

Proteins are probably the most important group of macromolecules when it comes to biological function: Amongst other things these macromolecules function as biocatalysts in cellular metabolism, mediate DNA amplification and gene transcription, participate in protein biosynthesis, carry molecules across membranes, function as molecular motors inside cells, act as signal transducers and selectivity filters, and function as structural components in all biological systems. All proteins are linear chains assembled from a set of 20 different building blocks, the amino acids, which differ in size, structure, hydrophobicity, and polarity. There is good evidence that most proteins adopt a single main conformation (the native state), which can occupy only a limited number of conformational substates. It is the three dimensional structure of this native state, which allows specific interactions with other molecules and enables protein function.

The driving force for adopting the native-state conformation (the folding of a protein chain) is the difference in Gibbs free energy between the natively folded and the unfolded state [21]. The Gibbs free energy of a system depends on its molecular interactions and on the degrees of freedom (the entropy) of all molecules in the system: The change in Gibbs free energy associated with the folding of a protein therefore depends on its intramolecular interactions (interactions within the protein molecule), as well as its intermolecular interactions (interactions between the protein molecule and water molecules, ligands, and other macromolecules), in the unfolded and folded state. As the chemical composition of the protein backbone is conserved, protein structure is mainly defined by chemical properties of the amino acid sidechains and their resulting molecular interactions. In a suitable chemical environment the folding of a protein into its native state is a spontaneous process, which like the native state is an intrinsic property of the amino-acid sequence [22]. A random or systematic survey of the complete conformational space of even a short protein chain would require times beyond the limit of the universe [23]. Instead, protein folding is governed by a progressive decrease in Gibbs free energy. The typical progression of folding does not follow a single folding pathway through conformational space, but rather progresses in a cumulative way, which depends on the establishment of cooperative interactions at various structural organization centers. Their formation and destruction is driven by random fluctuations of the folding chain, which are governed by

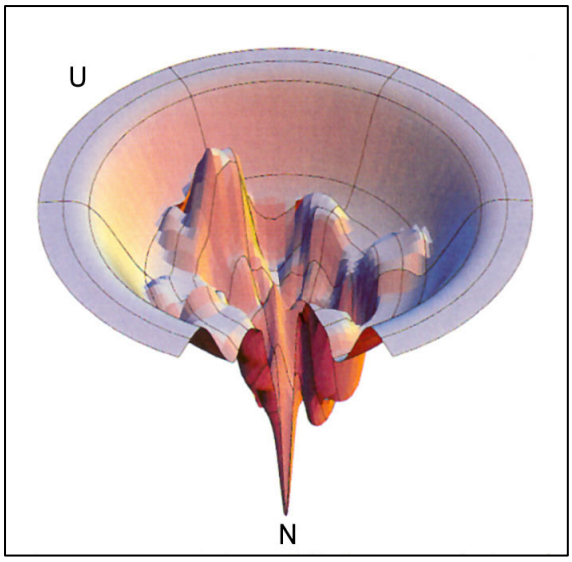


Figure 1.1: Folding can be multi-state. Schematic energy landscape of the folding space from a protein, with kinetic traps, energy barriers and some narrow throughway paths to the native state. U - unfolded state surrounding the funnel's rim, N - single native state. Figure by Dill & Chan [4].

the thermal energy. The conformational / energy landscapes that folding protein chains traverse resemble the shape of funnels. One example is given in Figure 1.1. Several possible folding paths originate at unfolded high energy states (on the funnel's rim), progress through several states of folded intermediates (within the funnel's cone), and end in the lowest-energy native state (at the funnel's bottom) [4]. The competition between loss of conformational entropy and energetic stabilization of inter-residue contacts results in an effective energy barrier separating the unfolded and folded states. The ruggedness of the landscape is responsible for the transient trapping of structures that are either partially folded or misfolded. If the landscape is sufficiently smooth the traps are shallow and there is no significant accumulation of intermediate structures. In this case only two states are observed in time-resolved protein folding experiments: The folded and the unfolded state (Figure 1.2). Observable folding pathways through the conformational landscape are not completely arbitrary, however: Important steps involve the clustering of hydrophobic groups and the formation of backbone secondary structure elements (α-helices, β-sheets), which precede the formation of tertiary structure elements [24]. The succession of folding events can differ for different proteins. But not all folding proteins reach their native fold on their own: Some proteins require additional factors, the chaperones and chaperonins, which stabilize folding intermediates by providing cavities and interaction surfaces for the stabilization of energy-rich transition states [25, 26], and hence promote folding. Other factors required to complete the folding of certain proteins in biologically significant time spans are peptidyl-prolyl *cis/trans*

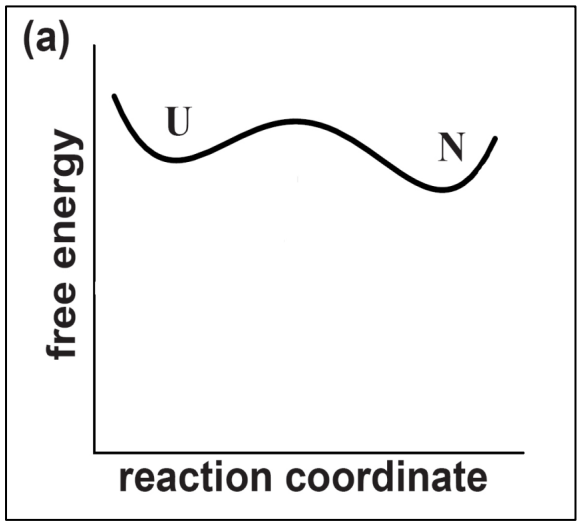


Figure 1.2: Energy diagram of a protein which folds according to a two-state mechanism. U - unfolded state, N - folded state. A transition state is located between the two states (local energy maximum). Figure after Pande *et al.* [5].

isomerases and disulfide isomerases, which catalyze chemical isomerizations of certain protein backbone and sidechain groups.

In some Proteins the entropic penalty of the folded state, which arises from the loss of conformational freedom in the unfolded state, is not completely compensated by favorable interactions provided by the native fold. Their Gibbs free energy of their unfolded state is lower than that of the folded state and hence they do not fold even under physiological conditions. These proteins are referred to as "natively unfolded" [5, 27] and require the addition of further stabilizing factors (often nucleic acids) or changes in the physicochemical environment, like downshifts in temperature, to adopt their native conformation [28].

Each protein has a thermal optimum at which it is maximally stable (Figure 1.3). Beyond (above and below) this temperature its stability decreases. Above and below certain threshold temperatures the Gibbs free energy of the unfolded state exceeds that of the folded state and the protein unfolds (heat & cold denaturation) [6, 29]. Imperfect thermal conditions are often associated with crowding of unfolded proteins and folding intermediates, and can result in the formation of protein aggregates which consist of misfolded protein chains. Misfolded conformations are derived from local energetic minima in the conformational space, surrounded by high energetic barriers which prevent unfolding [30]. Aggregated proteins are usually recognized and degraded by the cellular machinery. In some cases, however, these aggregates associate forming insoluble inclusions, colloidal particles, or filamentous fibers, which deposit in the cytosol and in other cellular compartments. Although under physiological conditions protein aggregation occurs sporadically involving wildtype forms of proteins, it is very often associated with specific mutations that destabilize the native fold. Such destabilizing mutations are causally linked to various diseases [30], such as Creutzfeld-Jakob disease, Kuru, Alzheimer's diseases, Parkinson's disease, and Fatal Familial insomnia. X-ray diffraction studies [31, 32] and cryo-electron microscopy studies [33] have provided structural insight into amyloid fibers, and have revealed their regular cross-β spine architectures. This type of architecture is formed by swiveled linear protofilament fibers, composed of stacked extended β-sheets, oriented perpendicular to the fiber axes, which as-

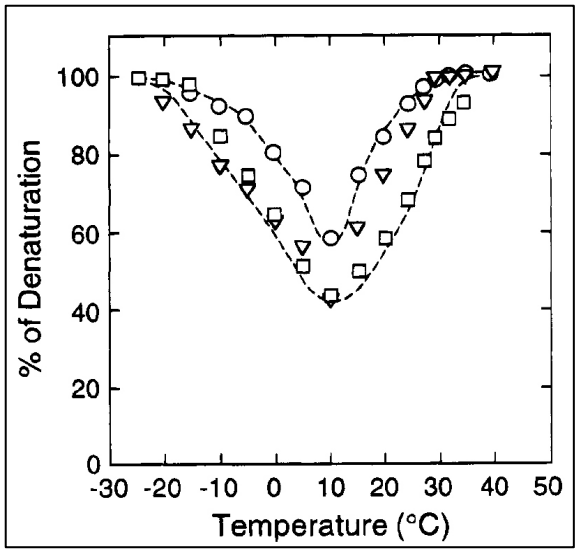


Figure 1.3: Temperature denaturation curves for histidines of Ribonuclease A from NMR spectra at 4 kbar of pressure, pH 2.0. Experimental points for individual histidine residues: ○ His 12, □ His48, ▽ His119+105. Figure by Zhang *et al.* [6].

semble forming hollow fibers. The structure of amyloids according to this model is solely defined by the peptide backbone; all tertiary structure of the native state is lost. This observation does not appear to be generally true, however: Recent structural and functional studies on ribonuclease T1 amyloid fibers suggest that the process of domain swapping, an exchange of subdomains from two or more different protein chains which promotes protein oligomerization but apart from a local reorientation does not influence the overall native fold, can result in the formation of cross-β-spine amyloids with retained biological activity [34].

1.2 Nucleic-acid structure and function

Nucleic acids are composed of four different nucleosides, which assemble forming long unbranched chains [35, 36]. Phosphorylated Nucleosides are referred to as Nucleotides. Nucleosides are composed of a sugar molecule, β-D-2'-deoxyribose in DNA (deoxyribonucleic acids) and β-D-ribose in RNA (ribonucleic acids), and one of the four nucleobases adenine, thymine, guanine, and cytosine (Figure 1.4). In RNA uracil replaces thymine. The information content of nucleic acids is encoded in their nucleotide sequences. Two nucleosides are linked by phosphodiester bonds involving the 3' and 5' hydroxyl groups of their sugar moiety (Figure 1.4).

The DNA are stabile long-lasting molecules, which function as carriers of the genetic information. DNA is typically organized as a helical double-stranded molecule, whose single strands are attached to each other by noncovalent forces [37]. The interior of the double strand is comprised of the nucleobases from two antiparallel strands, which assemble through intermolecular hydrogen bonding. This assembly follows a strict principle: The complementary chemical nature of the nucleobases favors the interaction of adenine with thymine (uracil in RNA) by forming two hydrogen bonds, and the interaction of guanine with cytosine by forming three hydrogen bonds. This kind of association is referred to as Watson-Crick basepairing. Hence, for forming a double-stranded helical molecule both single strands must have reverse-complementary sequences. The basepairs are oriented roughly perpendicular to the helix axis. The bases from adjacent basepairs are stacked on top of each other through hydrophobic interactions. Consecutive basepairs are rotated by about 32 – 36° with respect to each other along the helical axis. The outer part of DNA double strands is composed of its sugarphosphate backbone. In nucleic acids two predominant double-stranded conformations are observed, which both have right-handed screw senses [38] (Figure 1.5). B-DNA is the most abundant form of DNA. It has an elongated shape with a wide and deep major groove. This allows ac-

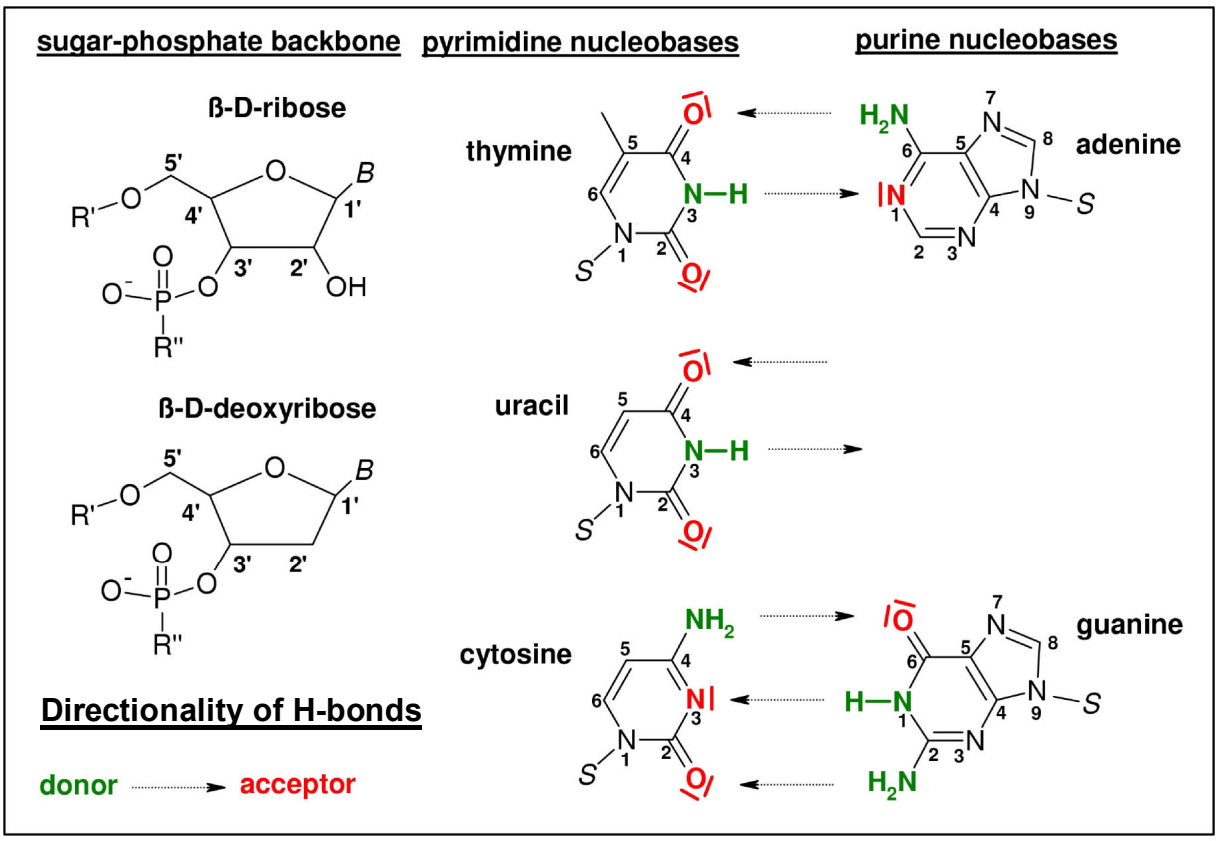


Figure 1.4: Nucleic-acid building blocks and Watson-Crick base pairing. Nucleosides are composed of a sugar moiety (β-D-ribose in RNA, β-D-2-deoxyribose in DNA) and a nucleobase, connected via an N-glycosidic link. The sites of connection are indicated by *italic* characters; *S* and *B* indicate connections with sugar and nucleobase groups, respectively. Nucleosides are linked via 5'-3' phosphodiester bonds. adjacent nucleosides are indicated by R' and R''. Watson-Crick base pairing between thymine and adenine / cytosine and guanine is indicated by arrows. Hydrogen-bond donor and acceptor groups are colored green and red, respectively. In RNA uracil replaces thymine. All numbers refer to the positional schemes of purines, pyrimidines, and sugars groups.

cess to groups which discriminate the nucleobases (positions 2, 3, and 4 in pyrimidines, positions 1, 2, and 6 in purines, Figure 1.4) and will be referred to as headgroups in the following text.

Due to steric hindrance between the 2'-hydroxyl group and oxygens from the adjacent phosphate group at its O3' position, as well as unfavorable interactions between the 2'-hydroxyl group and the C8 / C6 group of the nucleobase from the nucleoside on its 3' side, the ribose of RNA cannot pucker in the C2'-endo state, which is essential for the B-conformation [38]. In the A-conformation the sugar moiety is in the C3'-endo state. This conformation of nucleic acids is more compact and allows only limited access to the characteristic nucleobase headgroups. It is observed in DNA-RNA hybrid molecules and in RNA double strands. In DNA the A-conformation is not observed under physiological conditions; it can be induced, however, under high-salt conditions [39] or low humidity [40].

With the exception of certain viruses which use double-stranded RNA as the carrier of genetic information [41], RNA is not organized as a long double-stranded molecule in most living

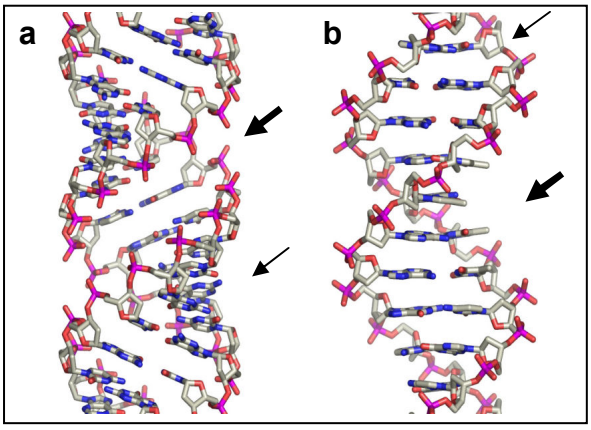


Figure 1.5: Two conformations of the DNA double helix. a) A-DNA, b) B-DNA. The minor and major grooves are indicated by bold and fine arrows.

systems and can be assigned into three different functional groups [36]: The rRNA, which is part of the ribosome (the organelle of protein biosynthesis), makes up about 80% of the total cellular RNA. This is followed by the tRNA (transfer RNA), which comprise about 16% of the total cellular RNA and serve as carriers of activated amino acids in protein biosynthesis. rRNA and tRNA molecules adopt discrete three-dimensional folds through intramolecular interactions, which involve double-stranded regions based on Watson-Crick as well as other types of basepairs and single-stranded regions. Like protein folding, the folding of these RNA molecules is governed by a progressive decrease in Gibbs free energy. Despite the facts that in contrast to proteins the chemical repertoire of nucleic acids is rather limited and the individual building blocks are rather large, certain rRNA and other folded RNA molecules can catalyze chemical reactions like enzymes and are therefore referred to as ribozymes [42].

The remainder of the cellular RNA is largely comprised of mRNA (messenger RNA), a group of elongated molecules which are predominantly single-stranded. mRNA contains transcribed genetic information, which is translated into protein sequences by the ribosome in protein biosynthesis. In contrast to rRNA and tRNA, the turnover of mRNA is quite fast.

1.3 Cellular adaptation to thermal environments and changing temperatures

Temperature is an important environmental parameter, which can greatly vary for different biological niches. Within one niche, temperatures can change considerably over time. The cellular temperature of most organisms is close to the temperature of the ecological niche they are living in. Only the classes *Mammalia* and *Aves*, which have a metabolism that allows the extensive production of heat and efficient cooling, are capable of keeping the body temperature constant over a limited range of temperature. Amazingly, in the struggle for biological resources life has developed strategies to colonize ecological niches covering a temperature range from -20 °C [43] to more than 120 °C [44] using basically the same cellular machinery. Adaptation to different thermal environments involves various adjustments of the cellular instrumentation, of which some examples shall be given here.

All biological cells are enclosed in a plasma membrane, which shields and protects their interior from the environment. In addition, cells contain specialized membrane-surrounded compartments in their cytosol. Most membranes are impermeable for polar substances and hence allow the formation of local concentration gradients, which fuel endergonic processes. Membranes function as barriers against mechanical stress and participate in many biological functions such as cellular motility, signal transduction, uptake and disposal of matter. Biological membranes are composed of amphiphilic lipid molecules with elongated bilayer-forming hydrophobic tails, and polar head groups which interact with the surrounding solvent molecules. Important classes of lipid molecules in membranes are the diacylphosphatidylglycerides (diglycerides) and the sphingosines. Membranes are biologically functional only in a fluid state. The temperature range where a membrane is in this physical condition strongly depends on molecular interactions, which can be influenced their by its molecular composition. While biological membranes of thermophilic bacteria contain long regular hydrocarbon chains which have large interaction surfaces, the chain length is reduced and the level of mono- and polyunsaturated fatty acid chains increases in mesophilic bacteria with decreasing temperatures [45, 46]. Shorter fatty-acid chains form smaller interaction surfaces with adjacent lipids, while double bonds present in unsaturated lipid chains give rise to kinked or curved lipid chains. These disturb the regular arrangement of surrounding hydrophobic chains, contribute to a higher membrane flexibility, and hence lower membrane stability. Other adaptations include alterations of the lipid head groups and in the protein content of membranes [47]. Ani-

mals and plants have developed additional strategies: They synthesize sterols, further lipid compounds [48], which increase membrane flexibility without affecting membrane stability a lot.

Proteins are a different kind of molecules which need to be specifically adapted to environmental temperatures to sustain cellular function. Most proteins adopt single native (biologically active) conformations, which are stabilized by molecular interactions (salt bridges, hydrogen bonds, hydrophobic interactions). The stabilizing interactions are opposed by entropic penalties, arising from the limited motilities of flexible groups in the native state, which greatly depend on temperature:

$$\Delta G = \Delta H - T \cdot \Delta S$$

Equation 1.1

ΔG - Gibbs free energy (driving force of a reaction), ΔH - enthalpie (thermodynamic energy), $T \cdot \Delta S$ – (energy from the change of disorder).

Various strategies have been identified, which help proteins to adapt to different thermal environments: The stability of proteins from thermophilic organisms is greatly extended by removing destabilizing charge interactions [49] and / or by providing additional attractive intermolecular interactions, as well as by improving their conformational structure (higher rigidity, tighter packing, reduced entropy of unfolding, release of conformational strain), as compared to their orthologs from mesophilic and cryophilic organisms [50, 51]. Additional stability, however, cannot be considered to be beneficial *per se*: it limits structural flexibility and hence decreases biological activity of proteins. This is especially important at cold temperatures, where kinetic rate constants of all chemical reactions are strongly decreased:

$$k = A \cdot e^{-\frac{E_a}{RT}}$$

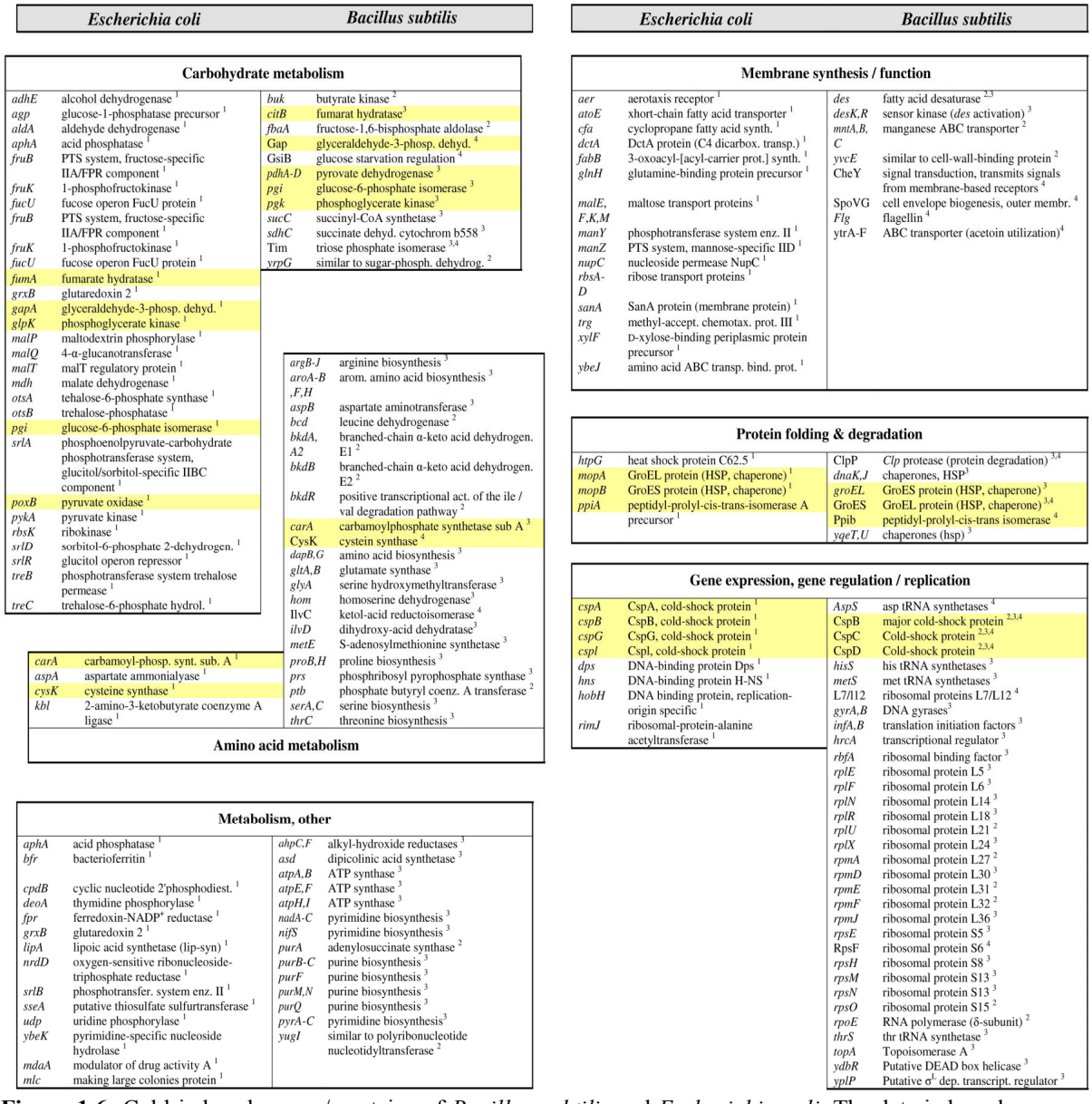
Equation 1.2

k is the reaction rate constant, A is a pre-exponential factor that depends on reaction parameters, E_a is the activation energy, R is the gas constant (8.314 J mol^{-1} K^{-1}) and T is the temperature.

Experimentally determined Q_{10} values of most reactions are around 2 - 3 (the rate constants of a reaction decreases 2 - 3 fold with a decrease in temperature by 10 °C). To compensate this loss in reaction velocities, psychrophilic enzymes have been selected for highest specific activities (larger k_{cat} values and higher overall turnover rates), and for enhanced catalytic efficiency (larger k_{cat} / K_m values - higher turnover rates at lower half saturating substrate concentrations), and have been shown to be considerably destabilized as compared to their mesophilic orthologs [52]. Higher catalytic activity is achieved through a reduction in activation energy. In the case of psychrophilic proteins this decrease is expected to originate from an

increased flexibility of either selected parts (subdomains) or the overall enzyme structure [53, 54].

Rapid changes of cellular temperature can be perilous for organisms, as consecutive changes influence further physicochemical parameters such as solubility, solvent polarity, pH, vapor pressure, solvent viscosity, surface tension, rate of diffusion, and thus compromise cellular homeostasis. Key problems occurring after a rapid decrease in temperature (≥ 10°C), which will be referred to as a cold shock in the following text, are: the loss of membrane fluidity, a lowering in the rates of enzymatic and transport processes, the stabilization of nucleic-acid secondary structures, their subsequent inhibitory effects on DNA replication, transcription and



Escherichia coli		*Bacillus subtilis*	
Carbohydrate metabolism			
adhE	alcohol dehydrogenase [1]	*buk*	butyrate kinase [2]
agp	glucose-1-phosphatase precursor [1]	*citB*	fumarat hydratase[3]
aldA	aldehyde dehydrogenase [1]	*fbaA*	fructose-1,6-bisphosphate aldolase [2]
aphA	acid phosphatase [1]	Gap	glyceraldehyde-3-phosp. dehyd. [4]
fruB	PTS system, fructose-specific IIA/FPR component [1]	GsiB	glucose starvation regulation [4]
fruK	1-phosphofructokinase [1]	*pdhA-D*	pyrovate dehydrogenase [3]
fucU	fucose operon FucU protein [1]	*pgi*	glucose-6-phosphate isomerase [3]
fruB	PTS system, fructose-specific IIA/FPR component [1]	*pgk*	phosphoglycerate kinase[3]
fruK	1-phosphofructokinase [1]	*sucC*	succinyl-CoA synthetase [3]
fucU	fucose operon FucU protein [1]	*sdhC*	succinate dehyd. cytochrom b558 [3]
fumA	fumarate hydratase [1]	Tim	triose phosphate isomerase [3,4]
grxB	glutaredoxin 2 [1]	*yrpG*	similar to sugar-phosph. dehydrog. [2]
gapA	glyceraldehyde-3-phosp. dehyd. [1]		
glpK	phosphoglycerate kinase [1]		
malP	maltodextrin phosphorylase [1]		
malQ	4-α-glucanotransferase [1]		
malT	malT regulatory protein [1]		
mdh	malate dehydrogenase [1]		
otsA	tehalose-6-phosphate synthase [1]		
otsB	trehalose-phosphatase [1]		
pgi	glucose-6-phosphate isomerase [1]		
srlA	phosphoenolpyruvate-carbohydrate phosphotransferase system, glucitol/sorbitol-specific IIBC component [1]		
poxB	pyruvate oxidase [1]		
pykA	pyruvate kinase [1]		
rbsK	ribokinase [1]		
srlD	sorbitol-6-phosphate 2-dehydrogen. [1]		
srlR	glucitol operon repressor [1]		
treB	phosphotransferase system trehalose permease [1]		
treC	trehalose-6-phosphate hydrol. [1]		
argB-J	arginine biosynthesis [3]		
aroA-B ,F,H	arom. amino acid biosynthesis [3]		
aspB	aspartate aminotransferase [3]		
bcd	leucine dehydrogenase [2]		
bkdA, A2	branched-chain α-keto acid dehydrogen. E1 [2]		
bkdB	branched-chain α-keto acid dehydrogen. E2 [2]		
bkdR	positive transcriptional act. of the ile / val degradation pathway [2]		
carA	carbamoylphosphate synthetase sub A [3]		
CysK	cystein synthase [4]		
dapB,G	amino acid biosynthesis [3]		
gltA,B	glutamate synthase [3]		
glyA	serine hydroxymethyltransferase [3]		
hom	homoserine dehydrogenase[3]		
IlvC	ketol-acid reductoisomerase [4]		
ilvD	dihydroxy-acid dehydratase[3]		
metE	S-adenosylmethionine synthetase [3]		
proB,H	proline biosynthesis [3]		
prs	phosphoribosyl pyrophosphate synthase [3]		
ptb	phosphate butyryl coenz. A transferase [2]		
serA,C	serine biosynthesis [3]		
thrC	threonine biosynthesis [3]		
carA	carbamoyl-phosp. synt. sub. A [1]		
aspA	aspartate ammonialyase [1]		
cysK	cysteine synthase [1]		
kbl	2-amino-3-ketobutyrate coenzyme A ligase [1]		
Amino acid metabolism			
Metabolism, other			
aphA	acid phosphatase [1]	*ahpC,F*	alkyl-hydroxide reductases [3]
bfr	bacterioferritin [1]	*asd*	dipicolinic acid synthetase [3]
		atpA,B	ATP synthase [3]
cpdB	cyclic nucleotide 2'phosphodiest. [1]	*atpE,F*	ATP synthase [3]
deoA	thymidine phosphorylase [1]	*atpH,I*	ATP synthase [3]
fpr	ferredoxin-NADP⁺ reductase [1]	*nadA-C*	pyrimidine biosynthesis [3]
grxB	glutaredoxin 2 [1]	*nifS*	pyrimidine biosynthesis [3]
lipA	lipoic acid synthetase (lip-syn) [1]	*purA*	adenylosuccinate synthase [2]
nrdD	oxygen-sensitive ribonucleoside-triphosphate reductase [1]	*purB-C*	purine biosynthesis [3]
		purF	purine biosynthesis [3]
srlB	phosphotransfer. system enz. II [1]	*purM,N*	purine biosynthesis [3]
sseA	putative thiosulfate sulfurtransferase [1]	*purQ*	purine biosynthesis [3]
udp	uridine phosphorylase [1]	*pyrA-C*	pyrimidine biosynthesis[3]
ybeK	pyrimidine-specific nucleoside hydrolase [1]	*yugI*	similar to polyribonucleotide nucleotidyltransferase [2]
mdaA	modulator of drug activity A [1]		
mlc	making large colonies protein [1]		

Escherichia coli		*Bacillus subtilis*	
Membrane synthesis / function			
aer	aerotaxis receptor [1]	*des*	fatty acid desaturase [2,3]
atoE	xhort-chain fatty acid transporter [1]	*desK,R*	sensor kinase (*des* activation) [3]
cfa	cyclopropane fatty acid synth. [1]	*mntA,B, C*	manganese ABC transporter [2]
dctA	DctA protein (C4 dicarbox. transp.) [1]		
fabB	3-oxoacyl-[acyl-carrier prot.] synth. [1]	*yvcE*	similar to cell-wall-binding protein [2]
glnH	glutamine-binding protein precursor [1]	CheY	signal transduction, transmits signals from membrane-based receptors [4]
malE, F,K,M	maltose transport proteins [1]	SpoVG	cell envelope biogenesis, outer membr. [4]
manY	phosphotransferase system enz. II [1]	*Flg*	flagellin [4]
manZ	PTS system, mannose-specific IID [1]	ytrA-F	ABC transporter (acetoin utilization)[4]
nupC	nucleoside permease NupC [1]		
rbsA-D	ribose transport proteins [1]		
sanA	SanA protein (membrane protein) [1]		
trg	methyl-accept. chemotax. prot. III [1]		
xylF	D-xylose-binding periplasmic protein precursor [1]		
ybeJ	amino acid ABC transp. bind. prot. [1]		
Protein folding & degradation			
htpG	heat shock protein C62.5 [1]	ClpP	*Clp* protease (protein degradation) [3,4]
mopA	GroEL protein (HSP, chaperone) [1]	*dnaK,J*	chaperones, HSP[3]
mopB	GroES protein (HSP, chaperone) [1]	*groEL*	GroES protein (HSP, chaperone) [3]
ppiA	peptidyl-prolyl-cis-trans-isomerase A precursor [1]	GroES	GroEL protein (HSP, chaperone) [3,4]
		Ppib	peptidyl-prolyl-cis-trans isomerase [4]
		yqeT,U	chaperones (hsp) [3]
Gene expression, gene regulation / replication			
cspA	CspA, cold-shock protein [1]	*AspS*	asp tRNA synthetases [4]
cspB	CspB, cold-shock protein [1]	CspB	major cold-shock protein [2,3,4]
cspG	CspG, cold-shock protein [1]	CspC	Cold-shock protein [2,3,4]
cspI	CspI, cold-shock protein [1]	CspD	Cold-shock protein [2,3,4]
dps	DNA-binding protein Dps [1]	*hisS*	his tRNA synthetases [3]
hns	DNA-binding protein H-NS [1]	*metS*	met tRNA synthetases [3]
hobH	DNA binding protein, replication-origin specific [1]	L7/l12	ribosomal proteins L7/L12 [4]
		gyrA,B	DNA gyrases[3]
rimJ	ribosomal-protein-alanine acetyltransferase [1]	*infA,B*	translation initiation factors [3]
		hrcA	transcriptional regulator [3]
		rbfA	ribosomal binding factor [3]
		rplE	ribosomal protein L5 [3]
		rplF	ribosomal protein L6 [3]
		rplN	ribosomal protein L14 [3]
		rplR	ribosomal protein L18 [3]
		rplU	ribosomal protein L21 [2]
		rplX	ribosomal protein L24 [3]
		rpmA	ribosomal protein L27 [2]
		rpmD	ribosomal protein L30 [3]
		rpmE	ribosomal protein L31 [2]
		rpmF	ribosomal protein L32 [2]
		rpmJ	ribosomal protein L36 [3]
		rpsE	ribosomal protein S5 [3]
		RpsF	ribosomal protein S6 [4]
		rpsH	ribosomal protein S8 [3]
		rpsM	ribosomal protein S13 [3]
		rpsN	ribosomal protein S13 [3]
		rpsO	ribosomal protein S15 [2]
		rpoE	RNA polymerase (δ-subunit) [2]
		thrS	thr tRNA synthetase [3]
		topA	Topoisomerase A [3]
		ydbR	Putative DEAD box helicase [3]
		yplP	Putative σ^{54} dep. transcript. regulator [3]

Figure 1.6: Cold-induced genes / proteins of *Bacillus subtilis* and *Escherichia coli*. The data is based on gene-expression analyses ([1] [8], [2] [18], [3][19]) and changes in protein concentrations analyzed by 2D-protein gel electrophores ([4] [20]) of cold-incubated cells (1h after cold shock) and controls. Upregulated genes / increases in gene products which are common in the two bacterial species are highlighted.

translation of mRNA, and the formation of crystalline ice and its associated damage to cellular structures if the drop goes below the freezing temperature of the cytoplasm. Various adaptive mechanisms which help organisms to cope with these changes have been identified and designated as cold shock response.

Whereas the bacterial response to increasing heat is understood in great detail [55], large parts of the initiation and regulation of their cold shock response still need to be elucidated. The consequence of a cold shock on culture growth and protein expression are very similar for many species: First, cell growth is halted and expression of most cellular proteins is down-regulated for an adaptive period, which can last a few hours. During this time, the expression of cold-shock response genes, whose gene products mediate cellular adaptation to cold stress [56], is significantly upregulated [57]. These genes vary between bacterial species [8, 18-20, 58]. Only one family, that of the major cold shock proteins, are highly conserved in *Eubacteria* [7, 59].

Although individual mechanisms contributing to thermal adaptation can differ considerably between organisms and apparently even for the same organism in different experimental set-ups (Figure 1.6), three strategies appear to be of general significance [60]: 1) expression of protein isoforms with modified kinetic properties and stabilities, 2) shifts in individual protein concentrations through changes in gene expression and protein turnover, and 3) changes in the milieu of the cell which compensate the effects of a temperature change on macromolecules. When temperatures are lowered, a prerequisite for organisms is the adjustment of vital metabolic processes to the decreased reaction rates. In addition, cellular cold adaptation requires energy and building materials. It is therefore not surprising that fundamental catabolic processes (the glycolysis, the tri-carboxylic acid cycle, the degradation of storage-lipids, and the urea pathway to name but the most important) and certain anabolic processes (mainly sugar-, amino acid-, lipid-, and nucleoside metabolisms) are commonly upregulated (see Figure 1.6 for *E. coli* and *B. subtilis*). As previously discussed for life in different thermal environments, a critical step in cellular adaptation to a temperature change are adjustments for the maintenance of membrane-fluidity, in order to maintain the biological functions of the lipid bilayer. In bacteria a temperature downshift is often followed by an upregulation of lipid-desaturases which introduce chemical double-bonds into membrane lipids. In addition, the composition of membrane proteins is adjusted to the new physiological situation by favoring the uptake of simple, energy-rich compounds, such as oligosaccharides, which can easily be digested. Apart from their roles as energy donors or cellular building blocks certain compounds, such as sucrose, trehalose, some amino acids, and other osmolytes [61], stabilize cellular proteins. Although their stabilizing effects are not understood in full detail, the sugar trehalose was shown

to preferentially interact with proteins in the folded state and to increase the surface tension of the solvent. Both mechanism significantly stabilize the native fold by increasing the energetic penalty of the unfolded state [62]. In addition, certain osmolytes lower the freezing-point and prevent the formation of ice crystals.

Changes in temperature generally induce the synthesis of chaperons / chaperonins and peptidyl-prolyl *cis/trans* isomerases, which aid protein folding (see Chapter 1.1). Other classes of proteins compensate effects on the stability of nucleic acids: Like in proteins, molecular interactions in nucleic acids are strongly influenced by temperature. A downshift in temperature aggravates DNA strand separation, which is required for gene transcription and DNA replication. In cellular systems the separation of the DNA double strand requires the introduction of negative supercoils. This is performed by type II isomerases / gyrases, which cleave the DNA double strand, passage a segment of DNA through this break and reseal the break. This mechanism consumes ATP. By introducing additional negative supercoils the double strand is destabilized and hence the stabilizing effect of a temperature downshift is compensated. Type I topoisomerases catalyze the reverse reaction, which is energetically downhill. In this way, the topological state of DNA double strands and hence the adhesion of the two DNA single strands can be fine-tuned to changes in temperature.

Another problem associated with temperature decreases is the increasing tendency of nucleic acid single strands to form mispaired double strands as a consequence of reduced entropic penalties (Equation 1.1). This effect is expected to greatly impair protein biosynthesis by lowering translation efficiencies. In many bacterial organisms a group of small RNA binding proteins, the major cold shock proteins, and certain RNA helicases are expressionally upregulated shortly after a temperature downshift, which may act as RNA chaperones that prevent aberrant mRNA duplex formation.

1.4 Bacterial cold shock proteins - key players in the bacterial cold shock response

Although most proteins involved in bacterial cold adaptation differ between species, a small group of proteins is conserved in more than 400 different bacteria, which include cryophilic, mesophilic, thermophilic and hyperthermophilic species [7, 56-59]. These proteins, designated as major cold shock proteins (CSP), consist of about 60 amino acids and bind to single-stranded nucleic acids with nano- to micromolar affinity. The number of CSP isoforms differs between species: *Escherichia coli* has nine different CSP, for *Bacillus subtilis* three CSP were described, the genome of *Thermotoga maritima* harbors two *csp* genes, and for *Bacillus caldolyticus* only one *csp* gene has been described so far. The expression of certain *csp* genes such as *cspA* from *Escherichia coli* and cspB from *Bacillus subtilis* increases significantly shortly after a cold shock [63, 64]. Hence, these genes are subject to cold regulation [57]. In contrast, the expression of most other genes is downregulated upon cold shock and culture growth stops for a period of adaptation, which can last several hours. An *E. coli* strain with four inactivated *csp* genes is greatly impaired in cold adaptation and is cold susceptible [8] . Similar results were obtained for *csp* deletion strains of *Bacillus subtilis* with two deletions [63]. The deletion of all three *csp* genes in this organism appears to result in a lethal phenotype. The cold sensitive phenotypes could be complemented by overexpression of individual *csp* genes. The reason for the great redundancy of CSP in certain species such as *Escherichia coli* is unclear. Some CSP have been implicated in cellular adaptation to stationary phase growth and dry stress [65], thus the CSP may be involved in different cellular adaptive mechanisms, which compensate the consequences of changes in environmental parameters including temperature.

It is not completely understood how the CSP contribute to cold adaptation, whether they generally help to maintain cellular function under cold conditions, or specifically participate in cold adaptation, or both. Their high affinity to single-stranded nucleic acids and their ability to prevent the formation of mispaired inter- and intramolecular double strands *in vitro* suggests, that they may be involved in the control and maintenance of transcription and translation [18, 19, 66-68]. It was shown that *Ec*-CspA (*Escherichia. coli* CspA) and *Bs*-CspB (*Bacillus subtilis* CspB) bind poly-thymidine oligonucleotides with K_D values in the nanomolar range. Poly-cytidine oligonucleotides are bound with affinities reduced by one or two orders of magnitude, whereas poly-purine based oligonucleotides are bound with K_D in the millimo-

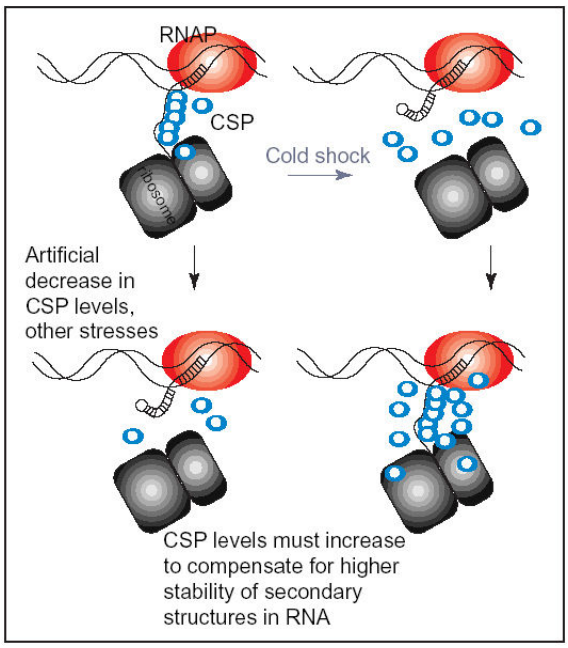


Figure 1.7: Model for the function of cold shock proteins as RNA chaperones that couple transcription to translation of mRNA. After cold shock or other stresses an increase in CSP concentration is necessary, in order to counterbalance the increased stability of RNA secondary structure. Figure by Grauman & Marahiel [7].

lar range [69-71]. The binding site of *Bs*-CspB was estimated to accommodate 6 - 7 nucleotides based on a CspB : dT_{23} stoichiometric relation of 3:1. Interestingly, dC_{23} interacts with a single *Bs*-CspB, suggesting that one end of this substrate may interact with only a part of a single CSP binding site. The sequence-independent binding to single-stranded oligonucleotides *in vitro* with K_D values in the micromolar range has lead to the speculation that the physiological role of the CSP could be an RNA chaperone function [7, 72]: By preventing the formation of mispaired mRNA double strands, an event which is expected to become more likely at cold temperatures and which may interfere with mRNA translation or even lead to mRNA degradation, the CSP may be required to sustain protein biosynthesis in the cold (Figure 1.7). Using an *in-vitro* translation system, however, the addition of a cold shock protein had the opposite effect, resulting in less efficient translation at cold temperatures in the presence of *Bs*-CspB [73].

In a different study, some CSP of *Escherichia coli* were shown to function as transcriptional antiterminators [74]: CspC and CspE could prevent the premature termination of transcription of the *metY-rpsO* operon both *in vitro* and *in vivo*. The expression of its *nusA*, *infB*, *rbfA*, and *pnp* genes located downstream of multiple transcription terminators is also induced upon cold shock. In addition, CspA was shown to prevent transcription antitermination of a genetic construct featuring an antibiotic-resistance located distal to a terminator site. Based on these observations it was suggested that the CSP may act as expressional regulators which fine-tune gene expression in the cold.

Structures of several CSP, *Bs*-CspB, *Bc*-Csp, *Ec*-CspA, and *Tm*-Csp (*Thermotoga maritima* Csp) were determined by X-ray crystallography [75-77] or NMR spectroscopy [78, 79]. The CSP belong to a subclass of the OB (oligonucleotide / oligosaccharide binding) fold and their chains fold into single β-barrels designated as cold shock domains (CSD). The barrel is composed of highly curved antiparallel β-strands arranged as β1-β2-β3-β5-β4-β1. β-strands 1-3 and 4-5 are arranged as two β-sheets, respectively. The β-strands are connected by three small

loops, which connect β1- β2, β2- β3 and β4- β5, and one extended loop connecting β3- β4, which is replaced in most other OB-fold family members by a long α-helix [80]. The protein surface of the CSP is highly polarized; one side carries polar and charged groups whereas on the other side several solvent-exposed aromatic sidechains, which are surrounded by polar groups, form an amphiphatic interface. Several residues from this interface are located in two conserved ribonucleoprotein motifs (RNP-1 and RNP-2), which can also be found in other RNA-binding proteins [81-83]. The replacement of aromatic sidechains from this interface by smaller sidechains significantly reduces the affinity for single-stranded nucleic acids [82, 84]. However, until now no NMR or crystal structure of cold shock proteins in complex with nucleic acids has been available, which structurally characterizes CSP·ligand interactions.

Cold shock domains can be found in most organisms with the exception of *Archaea* and *Cyanobacteria*. In *Eukaryotes* the highest sequence similarity with the bacterial CSP was found for the CSD of the Y-box proteins, which share at least 40% sequence identity with most CSP [85, 86]. The Y-box proteins have been implicated in transcriptional activation and repression, regulation of alternative splicing and mRNA stability, translational activation or repression, and RNA packaging [87]. It was also shown recently that a Y-Box protein was required for cell growth at cold temperatures [88].

1.5 Bacterial Cold shock proteins - model systems for protein folding and stability

The cold shock proteins are relatively small proteins (6 - 7.5 kDa in size) featuring single β-barrels which are composed of five antiparallel β-strands and intervening loops [75-79]. These β-barrels contain a hydrophobic core mainly composed of aromatic and large hydrophobic residues, and a polar surface, which contains several polar residues and some solvent-exposed aromatic sidechains. The CSP architecture therefore resembles that of more complex globular proteins, while their surface-exposed aromatic sidechains are a characteristic feature of nucleic-acid binding proteins. The CSP adopt their native state following a simple two-state (unfolded $\leftrightarrow$ folded) mechanism, which does not appear to involve stable folding intermediates [16, 17]. The time constant of the folding process is less than one millisecond at 25 °C [17, 89]. This fast folding is conserved in the CSP family. The transition state of folding is reached very late on a structural reaction coordinate and a Tanford β_T value of ≥ 0.9 (measure of the average degree of exposure in the transition state relative to that of the denatured state and native state) suggests that the transition state of *Bs*-CspB is almost native-like.

Further analyses of *Bs*-CspB using FRET (fluorescence resonance energy transfer) experiments suggest that the hydrophobic collapse of residues forming the core precedes the rate-limiting step in protein folding [90] .

CSP are highly conserved (typically on a level above 40% sequence identity) in the kingdom of *Eubacteria*. This includes psychrophilic, mesophilic, thermophilic, and hyperthermophilic species. The CSP of the mesophilic bacterium *Bacillus subtilis* (*Bs*-CspB) and the hyperthermophilic bacterium *Bacillus caldolyticus* (*Bc*-Csp) differ in 12 (of 67) sequence positions of which two, Arg3 and Leu66, confer increased stability to *Bc*-Csp ($\Delta\Delta G$ = 15.8 kJ·mol^{-1}) [49, 91, 92]. Structural and thermodynamic analyses revealed that the additional stabilization of *Bc*-Csp has a strong electrostatic component, which does not arise from pairwise coulombic attractions of the key residues identified by mutagenesis [49]. Instead, additional unfavorable electrostatic interactions at positions 3 and 66 appear to be predominantly responsible for the destabilization of *Bs*-CspB.

The globular architecture, small size, high level of sequence conservation, and the existence of CSP of organisms adapted to various thermal environments qualifies these proteins as ideal model systems for studying protein folding and protein stability.

1.6 Goals of this study

This dissertation aims at structural studies of cold shock proteins for a better understanding of CSP function on the molecular level as well as in biological contexts. The second chapter contains all materials and methods used in this study. It also contains a detailed description of protein structure determination by X-ray crystallography.

In the third chapter, complexes of CSP and oligonucleotides are analyzed using X-ray crystallography and further biophysical methods, in order to address the following questions: How do CSP bind to oligonucleotide ligands? How selective is nucleic-acid binding by CSP? Is the CSP oligonucleotide binding site preformed in the absence of a ligand or does ligand binding involve an induced-fit mechanism? Do different classes of cold shock proteins exist, which recognize different targets, or do all CSP feature a common binding interface? Is there evidence for a functional relationship between the bacterial CSP and the eukaryotic Y-box proteins?

In the fourth chapter the structural plasticity of the cold shock domain architecture is analyzed, by comparing the crystal structure of a CSP swapped dimer to closed monomeric structures of CSP. In addition, the structural rearrangement which enables the domain swap will be

discussed in the context of CSP folding, and the question will be addressed, whether domain swapping favors oligonucleotide-binding, or *vice versa*.

In chapter five the crystal structures of two *Bs*-CspB variants which feature five stabilizing mutations are analyzed for effects contributing to additional protein stability. These variants were generated by the *Proside in-vitro* evolution method [11], which is based on phage display techniques. Their stabilizing effect will be compared to stabilizations in a naturally occurring CSP from the thermophilic bacterium *Bacillus caldolyticus* [49, 77]. Finally, the prevalence of stabilizing residues in *Bs*-CspB and *Bc*-Csp in the sequences of CSP representatives is evaluated, in order to test whether effects stabilizing the CSP variants generated by *Proside*, may impart stability in natural CSP.