CHARLES UNIVERSITY IN PRAGUE

Faculty of science

Department of Biochemistry

Study program: Biochemistry

Study specialization: Biochemistry

Mgr. Mikhail Makarov

Vliv redukce aminokyselinového repertoáru na strukturu a funkci defosfokoenzym A kinasy

Effect of amino acid alphabet reduction on structure and function of dephosphocoenzyme A kinase

MASTER THESIS

Supervisor: Mgr. Klára Hlouchová, Ph.D.

Prague 2020

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne

Abstract

It is well-known that the large diversity of protein functions and structures derives from the broad spectrum of physicochemical properties of the 20 canonical amino acids that constitute modern proteins. According to the generally accepted coevolution theory of the genetic code, evolution of protein structures and functions was continuously associated with enrichment of the genetic code, with aromatic amino acids being considered the latest addition to the genetic code to increase structural stability of proteins and diversification of their catalytic functions.

The main objective of this master thesis was to test whether enzymatic catalysis could precede the appearance of aromatic amino acids in the standard genetic code. For that purpose, the effect of amino acid alphabet reduction on structure and function of dephosphocoenzyme A kinase (DPCK) was studied. Dephosphocoenzyme A kinase catalyses the final step in the biosynthesis of coenzyme A, a very conserved cofactor.

Two aromatic amino acid-lacking mutants of DPCK from a thermophilic bacterium, *Aquifex aeolicus*, were designed by substituting aromatic amino acid residues by (i) leucines and (ii) various non-aromatic amino acids to best preserve the structural stability of the protein. Wild type protein and the two mutants were cloned and successfully expressed in *Escherichia coli*, and the recombinantly produced proteins were characterized regarding the preservation of the secondary/tertiary structure and enzymatic activity. Structural characterization suggests that substitution of aromatic amino acids by non-aromatic residues can support rich secondary structure but leads to drastic loss of a firm globular arrangement, that resulted in a significant decrease or loss of enzymatic activities. Enzyme assays demonstrated that one of the mutants did not exhibit any catalytic activity. The other (where all aromatics were substituted by leucine residues) is capable of efficient ATP hydrolysis unlike the wild type and also less efficient phosphotransferase activity when compared with the wild type protein.

Using this exemplary study, the results presented in this thesis suggest that formation of protein structure and some catalysis could at some cost precede incorporation of aromatics into the genetic code.

Key words: protein reverse evolution, coenzyme A, dephosphocoenzyme A kinase, CD spectroscopy, limited proteolysis.

Abstrakt

Je dobře známo, že velká rozmanitost proteinových funkcí a struktur je umožněna širokým spektrem fyzikálně-chemických vlastností 20 standardních aminokyselin, které tvoří moderní proteiny. Podle obecně přijímané koevoluční teorie genetického kódu, evoluce proteinových struktur a funkcí byla nepřetržitě spojena s rozšiřováním genetického kódu, přičemž aromatické aminokyseliny jsou považovány za nejvíce pozdní členy genetického kódu. Jejich inkorporace zřejmě přispěla ke zvýšení strukturní stability proteinů a diverzifikaci jejich katalických funkcí.

Hlavním cílem této diplomové práce bylo otestovat, zda enzymová katalýza mohla předcházet inkorporaci aromatických aminokyselin do standardního genetického kódu. Pro tento účel byl prozkoumán vliv redukce aminokyselinového repertoáru na strukturu a funkci defosfokoenzym A kinasy (DPCK). Defosfokoenzym A kinasa katalyzuje poslední krok v biosyntéze koenzymu A, což je evolučně vysoce konzervovaný kofaktor.

V této práci byly navrženy dva mutantní proteiny DPCK z termofilní bakterie, *Aquifex aeolicus*, tak, aby aromatické aminokyseliny byly nahrazeny (i) leuciny a (ii) nearomatickými aminokyselinami pro co největší zachování strukturní stability proteinu. Přirozená varianta (z angl. wild type) proteinu a dva mutantní proteiny postrádající aromatické aminokyseliny byly zaklonovány a úspěšně exprimovány v *Escherichia coli*. Purifikované proteiny pak byly charakterizovány z hlediska zachování sekundární a terciární struktury a enzymové aktivity v mutantech. Strukturní charakterizace ukazuje, že substituce aromatických aminokyselin nearomatickými aminokyselinami může zachovat bohatou sekundární strukturu, ale vede k výrazné ztrátě pevného globulárního uspořádání, což má za následek významné snížení nebo ztrátu enzymové aktivity. Zatímco jeden z mutantů nevykazoval žádnou námi měřitelnou katalytickou aktivitu, leucinový mutant byl na rozdíl od přirozené formy proteinu schopen ATP hydrolýzy (i v nepřítomnosti druhého substrátu) a měl výrazně sníženou fosfotransferasovou aktivitu.

Tato práce tedy na zvoleném příkladu proteinu DPCK ukazuje, že i při absenci aromatických aminokyselin může být protein schopen tvořit strukturní uspořádání a vykazovat enzymatickou aktivitu, i když se znatelně sníženou účinností.

Klíčová slova: proteinová reverzní evoluce, koenzym A, defosfokoenzym A kinasa, CD spektroskopie, omezená proteolýza.

Acknowledgements

I would like to express my gratitude to Mgr. Klára Hlouchová, Ph.D. for the opportunity given to join the laboratory, and for the patient guidance, encouragement and support throughout this project. I would also like to thank all members of the laboratory, Mgr. Filip Buchel, Valerio Guido Giacobelli, M.Sc., Ph.D. and Mgr. Vjačeslav Treťjačenko, for creating a friendly work environment, sharing valuable knowledge and training. I would additionally like to thank RNDr. Anna Březinová for performing HPLC/MS analysis and Radko Souček, Ph.D. for amino acid analysis measurements.

Table of contents

List of abbreviations

ADP	adenosine-5'-diphosphate
AMP	adenosine-5'-monophosphate
APS	ammonium persulphate
ATP	adenosine-5'-triphosphate
BLAST	basic local alignment search tool
bp	base pair
°C	Celsius degree
CD	circular dichroism
CDNN	circular dichroism analysis using neural networks
cm	centimetre
CTP	cytidine-5'-triphosphate
dCoA	3'-dephosphocoenzyme A
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DPCK	dephosphocoenzyme A kinase
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
ESI	electrospray ionization
g	gram
GTP	guanosine-5'-triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOMO	highest occupied molecular orbital
HPLC	high-performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase
k_{cat}	turnover number
kDa	kilodalton
kHz	kilohertz
K_m	Michaelis−Menten constant
L	litre
LUCA	last universal common ancestor
LUMO	lowest unoccupied molecular orbital
M	mole per litre
MADLS	multi-angle dynamic light scattering
MALDI	matrix-assisted laser desorption/ionization
mAU	milli-absorbance unit
ml	millilitre
mm	millimetre
mM	millimole per litre

M_r	relative molecular mass
MS	mass spectrometry
MWCO	molecular weight cut-off
ng	nanogram
nm	nanometre
nM	nanomole per litre
OD_{600}	optical density measured at a wavelength of 600 nanometres
PAGE	polyacrylamide gel electrophoresis
PanK	pantothenate kinase
PCR	polymerase chain reaction
PDB	protein data bank
pH	potential of hydrogen
PPAT	phosphopantetheine adenylyltransferase
PPCDC	4'-phosphopantothenoylcysteine decarboxylase
PPCS	4'-phosphopantothenoylcysteine synthetase
RNA	ribonucleic acid
RNase	ribonuclease
RPM	revolutions per minute
SDS	sodium dodecyl sulphate
SE	standard error
TCEP	tris(2-carboxyethyl)phosphine
TEMED	N,N,N',N'-tetramethylethylenediamine
TOF	time of flight
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
TrpRS	tryptophanyl-tRNA synthetase
TTP	thymidine-5'-triphosphate
TyrRS	tyrosyl-tRNA synthetase
U	enzyme unit (µmol/min)
UV	ultraviolet
V	volt
v/v	volume per volume
V_e	elution volume
V_{max}	maximum velocity
w/v	weight per volume
µg	microgram
µl	microlitre
µM	micromole per litre

1 Introduction

The theoretical part of this thesis consists of two major sections. The first part concerns evolution of the amino acid alphabet of extant proteins. Its effect on protein structure and function will be studied as part of this thesis on an exemplary protein target and its selection will be reasoned in the second part.

1.1 Amino acid alphabet evolution

Extant life on Earth is based on a series of biochemical reactions catalysed by enzymes encoded in DNA by the genetic code. It is well-agreed that the standard genetic code of 20 amino acids universally employed for protein synthesis in ribosomes consisted of a fewer number of amino acids in the very early history of life. Evolution of life at its very early stages was continuously associated with the enrichment of the genetic code that regulated the amino acid composition of proteins. The mechanism and history of the genetic code evolution are one of the most important questions surrounding the origin of life on Earth [1].

1.1.1 Possible scenarios of origin of life

It is generally accepted that life arose on the Earth early in its history. The order of events that led to emergence of life on Earth started with simple organic compounds, which reacted to form polymers, which in turn reacted to produce structures of greater and greater complexity until the first living organisms were formed [2].

The origin of the first living organisms and the initial conditions on primitive Earth is still hotly debated, which led to two general theories of origin of life [3,4]. Chemoautotrophic theory claims that life began with the first chemoautotrophic organisms capable of producing organic compounds by primitive carbon dioxide fixation pathways [5,6]. However, the more prevailing hypothesis suggests that first living organisms were heterotrophic. According to the heterotrophic theory the first organisms used prebiotically synthesized organic compounds available in the prebiotic environment [7]. They had become dependent on them and developed metabolic pathways for their biosynthesis after their depletion. These first heterotrophic organisms still had to build proteins, nucleic acids and other complex compounds, but they did not have to synthesize *de novo* amino acids, nitrogen bases and saccharides that were available in the prebiotic environment [2].

1.1.2 Prebiotic sources of amino acids

Modern proteins are composed of N-unsubstituted α-monoalkylamino acids. The number of possible structural isomers of such amino acids increases exponentially with increasing number of carbon atoms. While there is only one option for two carbon atoms – glycine, there are 654 possible structural isomers of amino acids made of six carbon atoms, however, only two of them, leucine and isoleucine, constitute modern proteins [7]. This observation, among others, leads to the suggestion that there were certain selection factors to form a set of 20 canonical amino acids encoded by the genetic code.

The prebiotic availability appears to be the main factor for amino acids selection [7]. The first strong biochemical evidence of heterotrophic hypothesis was provided by electric discharge experiments [8,9]. Passing electric discharges through a mixture of methane, ammonia, water vapor and molecular hydrogen that simulated the primitive reducing Earth's atmosphere, Miller and Urey (1959) obtained a complex mixture of substances, from which they identified 10 amino acids (among other non-canonical amino acids) that constitute half of the amino acid alphabet of modern proteins. This half was represented by non-polar aliphatic amino acids (glycine, alanine, valine, isoleucine, leucine, proline), polar amino acids with hydroxyl group (serine, threonine) and negatively charged amino acids (glutamate, aspartate). Several possible mechanisms of their prebiotic synthesis have been proposed and are illustrated in Fig. 1.1.

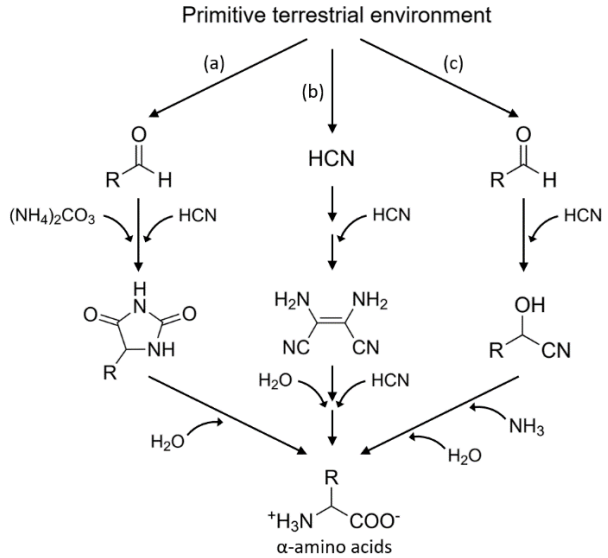


Figure 1.1. Possible mechanisms of α-amino acids synthesis in primitive terrestrial environment. (a) The Strecker synthesis of amino acids by the addition of ammonia and hydrogen cyanide to aldehydes and ketones [10]; (b) Generation of amino acids by oligomerization of hydrogen cyanide [11,12,13]; (c) The Bucherer-Bergs synthesis of amino acids from hydantoins which could have been formed from hydrogen cyanide, ketones (or aldehydes) and ammonium carbonate [14].

The yield of amino acids in prebiotic synthesis experiments decreases significantly with increasing number of carbon atoms, suggesting that the smaller amino acids were more abundant than the more complex amino acids in the prebiotic environment [7]. The other canonical amino acids including sulphur-containing amino acids (cysteine, methionine), positively charged amino acids (arginine, lysine, histidine) and aromatic amino acids (phenylalanine, tyrosine, tryptophan) were not identified in the reaction mixture. Clearly, prebiotic synthesis seems to favour the smaller and less complex amino acids, however, somewhat more convoluted syntheses have been introduced also for the bulkier and complex canonical amino acids. These include the aromatic amino acids, sulphur-containing amino acids and histidine [15,16,17,18].

Prebiotic synthesis experiments assume the absence of free molecular oxygen and strong reducing conditions of primordial Earth atmosphere which could have directed prebiotic processes towards the synthesis of organic compounds rather than their oxidative degradation. Even if the conditions were different than assumed, which compromises endogenous synthesis of organic compounds, various exogenous sources of organic material such as dust particles, meteorites and comets were reported [19,20]. A rich variety of organic compounds including amino acids have been found in carbonaceous chondrites, primitive and undefined meteorites, which are believed to form the Earth [21]. They were formed in the interstellar or circumstellar medium and were later incorporated into planetesimals during the formation of the solar system [22,23,24,25]. All amino acids, which were produced in Miller-Urey experiment, except serine and threonine, have been found in the Murchison meteorite that fell to Earth in 1969 [26,27,28]. The amino acids detected in meteorites are 4,5 billion years old, and certain amino acids might have been exposed to radiation and heat and undergone alteration processes [23].

1.1.3 Factors that led to selection of amino acids into the genetic code

Theoretical and statistical investigations clearly indicate that the choice of amino acids to be incorporated in the standard genetic code was nonrandom and adaptive [29,30,31]. These observations combined with analyses of prebiotic amino acid sources lead to a suggestion that the more complex amino acids were chosen based on other criteria than prebiotic availability, and were included in the genetic code and recruited for protein synthesis after the establishment of functioning replicating biochemistry [32,33]. Those

amino acids, whose synthesis was not enabled in the prebiotic environment, might have properties which provided evolutionary advantages for development of more complex functions and organisms. Some of the factors that could have played a role in their selection are summarized below.

Extreme pH and high temperatures may have exerted a significant selection pressure since amino acids are generally more stable near the neutral pH and at lower temperatures [34,35]. Depending on the environment where life evolved, these factors could have contributed to the synthesis and recruitment of different amino acids [7].

It is generally accepted that the Earth at the time of emergence of life was exposed to much more ultraviolet (UV) light from the Sun because of low content of free molecular oxygen and thus ozone in the primordial Earth atmosphere [36,37]. It has been shown that cysteine, methionine, phenylalanine, tyrosine, tryptophan and histidine are rapidly degraded by ultraviolet irradiation [38]. Thus, UV irradiation may have been one of the earliest selection pressures, which could have caused the late addition of the above-mentioned amino acids, particularly aromatic amino acids because of their strong absorbance of UV light [39]. Further recruitment of aromatic amino acids for protein synthesis may play a key role in initiating hydrophobic collapse in protein folding and would facilitate folding of increasingly larger proteins over the course of development of life [7,40].

Functional rather than structural amino acid properties could have contributed to the expansion of the early version of the genetic code. The selective incorporation of the last amino acids in the genetic code could be initiated by the appearance of free molecular oxygen in the Earth atmosphere [41]. The energetic distance between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), HOMO–LUMO gap, constitutes the basis of chemical reactivity and particularly reflects the kinetic stability of a substance toward electron transfer reactions [42,43]. Comparative analysis of HOMO–LUMO gaps indicated that amino acids, which were produced in prebiotic synthesis experiments and detected in the Murchison meteorite, have higher values of HOMO–LUMO gaps and lower redox reactivity than amino acids that are considered to be incorporated into the genetic code lately [41]. That fact demonstrates that the appearance of free molecular oxygen in the habitat of the early cells led to incorporation of novel amino acids with increased redox reactivity for protection from destruction by oxygen free radicals, namely tyrosine and tryptophane with the lowest values of HOMO–LUMO gaps.

1.1.4 Coevolution theory of the genetic code

Clearly, prebiotic synthesis could not produce all 20 canonical amino acids. This observation leads to a conclusion that the very first proteins could be constituted from a limited set of amino acids which were included in the original version of the genetic code [32]. The original genetic code was then gradually expanded with time by addition of more complex amino acids after developing the pathways for their biosynthesis [29]. This idea forms a basis of coevolution theory, according to which the evolution of the standard genetic code was directed primarily by coevolving amino acid biosynthetic pathways [44]. The other main theories of the genetic code origin and evolution include error minimization theory, according to which the genetic code evolved by selecting and placing biochemically similar amino acids at neighbouring positions in the genetic code to minimize the harmful effect of point mutations, and stereochemical theory, under which codons were assigned to amino acids based on physicochemical affinity between an encoded amino acid and its anticodons [45]. These three theories are not mutually exclusive and complement each other since all of them imply purposive and adaptive development of the genetic code.

According to the coevolution theory, the 20 canonical amino acids can be classified into two different groups: early (phase 1) amino acids which were prebiotically available and included first into the genetic code – glycine, alanine, serine, aspartic acid, glutamic acid, valine, leucine, isoleucine, proline and threonine; and late (phase 2) amino acids which could not be obtained from prebiotic sources and were incorporated into the genetic code only after establishment of their biosynthetic pathways – arginine, lysine, glutamine, asparagine, cysteine, methionine, phenylalanine, tyrosine, tryptophan and histidine [44]. The coevolution theory is strongly supported by studies that demonstrated that a set of 7–13 amino acids is sufficient to produce foldable and catalytically active polypeptides [46,47,48]. However, it cannot be excluded that the original set of the genetic code may have been larger and included diverse non-canonical prebiotically available amino acids due to low selectivity of early biocatalysis [49]. Nevertheless, using large meta-analyses of different computational and experimental studies, at least the order of canonical amino acid incorporation into genetic code could be approximated, as described in the following chapter.

1.1.5 Consensus chronology of amino acids appearance in the genetic code

The chronology in which amino acids were incorporated in the genetic code at the very early stages of life evolution has been derived from massive comprehensive meta-analyses of all the factors that could be involved in their recruitment. One of the first comprehensive meta-analyses ranked amino acids in order of their decreasing occurrence in the prebiotic context. Based on different versions of prebiotic formation of amino acids, the consensus chronological order of amino acid appearance in the genetic code was derived by comparison of relative concentrations of amino acids measured in meteorites, produced in prebiotic synthesis experiments with different compositions of primitive Earth atmosphere, formed in hydrothermal synthesis simulations and various chemical synthesis experiments [50]. Those amino acids, which had not been formed in any of the prebiotic synthesis experiments were ranked according to the free energies of formation from carbon dioxide, hydrogen and ammonium ions in surface seawater conditions (18 °C, 1 atmosphere) and under hydrothermal conditions (100 °C, 250 atmospheres). The resulting chronological order of amino acids based on their prebiotic availability is illustrated in Fig. 1.2.

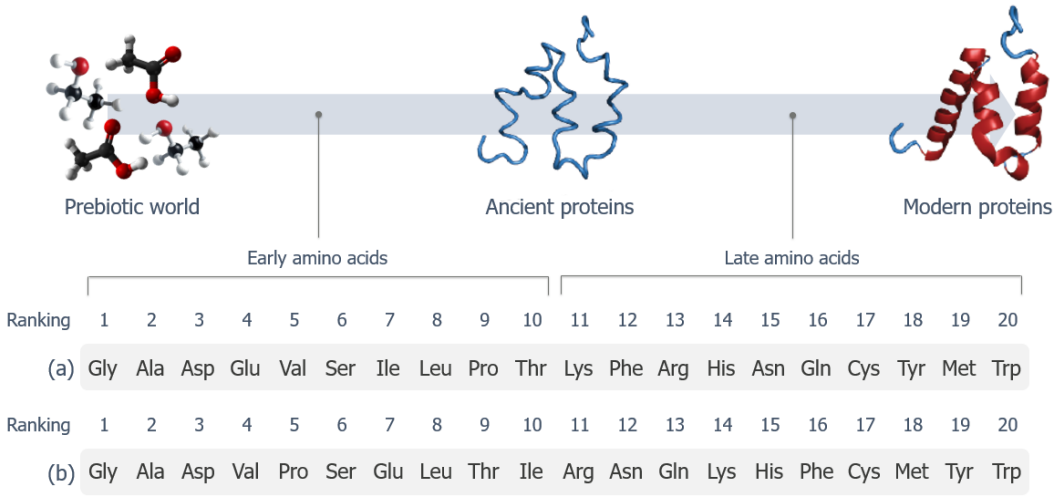


Figure 1.2. The correlation of 20 canonical amino acids in different chronological orders. Chronological order of amino acids (a) based on the prebiotic availability [50]; (b) derived by building a consensus among the various theories of the evolution of the genetic code [51].

Alternative consensus chronological order was derived from 101 different criteria based on published hypotheses about the chronology of amino acid emergence in the genetic code [51]. Expressing 101 published hypotheses as 20D ranking vectors with the

highest rank corresponding to earliest amino acids and their averaging led to an alternative yet similar consensus order (Fig. 1.2, page 16).

The first half of the order derived by Trifonov (2008) is represented by amino acids with the highest relative abundances in prebiotic environment according to Higgs and Pudritz (2009). That indicates that all the amino acids included in the early version of the genetic code were just those ones that were available in the primordial chemical environment [33]. The relative abundances of these early amino acids most likely determined the composition of the first proteins at the time of the genetic code emergence [50].

The second half of the order consists of amino acids encoded by codons which are thought to be captured from the repertoires established earlier. According to the coevolution theory, all 64 triplets were available for the assignment for the earliest amino acids, and after exhaustion of the whole triplet repertoire by early amino acids, the late amino acids could have been included in the genetic code only by codon reassignment [51]. For example, cysteine, tryptophan and tyrosine are thought to be encoded by codons that previously belonged to termination repertoires, lysine is proposed to acquire its codons from asparagine repertoire, and methionine – from isoleucine repertoire [52].

For the subject of this thesis, it is worth stressing that aromatic and sulphur-containing amino acids have been denoted as the latest additions to the genetic code by the meta-analyses mentioned above. Some recent evidence suggests that at least some of the aromatic amino acids were included even after the last universal common ancestor (LUCA), as described in the following section.

1.1.6 Amino acid composition of the primordial translational apparatus

The time between the origin of life and the LUCA is one of dramatic evolutionary change that led to the invention of the core cellular machinery, including the translational apparatus [1]. The translational machinery is one of the most conserved components of cellular metabolism [53]. Its key components, ribosomes and aminoacyl-tRNA synthetases, were the earliest systems to evolve into the extant form before the formation of three domains of life from the LUCA [53,54]. They belong to the oldest fold families that were traced back to the LUCA by extensive reconstructive bioinformatic studies.

Ancestral sequence reconstruction is an *in silico* approach for inferring the amino acid sequences of primordial proteins from the sequences of homologous extant proteins

[55]. Ancestral sequence reconstructions of universally conserved ribosomal proteins revealed a compositional bias that likely provides evidence of earlier stages in the evolution of the genetic code [56,57,58]. Compositional analysis indicates that glycine, alanine, threonine, aspartic acid and asparagine are among the most ancient amino acids in the genetic code, while glutamic acid, glutamine, cysteine, serine phenylalanine, tyrosine and tryptophan, are later additions. These results satisfactorily agree with the results of prebiotic synthesis experiments and consensus meta-studies.

Except ribosomes, three major groups of proteins including aminoacyl-tRNA synthetases proteins, amino acids biosynthesis enzymes and tRNA modification proteins could potentially have played a role in the development of the genetic code via their early evolution and diversification [1]. Evolution of novel aminoacyl-tRNA synthetases and amino acid biosynthesis enzymes could enable the incorporation of new amino acids in the genetic code and their use in polypeptide synthesis [59,60].

Ancestral sequence reconstructions of TyrRS and TrpRS protein families demonstrated that pre-LUCA ancestor of these aminoacyl-tRNA synthetases has several sites containing tyrosine, but completely lacks sites containing tryptophan. It indicates that the pre-LUCA ancestor was specific for the utilization of tyrosine, and tryptophan was added to the genetic code afterwards by divergence and neofunctionalization of ancestral aminoacyl-tRNA synthetases [1 - Fournier G. P.; Alm E. J. 2015]. Structural analyses of TrpRS and TyrRS suggest that these aminoacyl-tRNA synthetases may not have been present in LUCA, and therefore incorporation of tyrosine and tryptophan in the genetic code could have occurred only at the time of speciation of the first bacterial lineages [61].

1.2 Target selection

1.2.1 The thioester world hypothesis

Evolution of life on Earth has gone through certain stages from the prebiotic synthesis of the first organic molecules to the appearance of the modern “protein world” (Fig. 1.3, page 19). Discovery of naturally occurring ribozyme catalysts in 1980s led to the emergence of the most accepted and widespread hypothesis of the evolution of life on Earth according to which RNA molecules preceded ribosomal protein synthesis and mediated all the processes necessary for the functioning of prebiotic living systems in the early history of life [62,63]. Despite the high chemical complexity of the building blocks

of RNA, ribonucleotides, pyrimidine and 8-oxo-purine nucleotides could have been synthesised prebiotically from cyanoacetylene and cyanide, and further polymerized on the surface of clay minerals to produce RNA [64,65]. However, the RNA world stage is not considered to directly follow the stage of prebiotic synthesis mostly because of highly probable instability of long RNA molecules in the Earth's primordial environment and enormous complexity of RNA required to perform catalytic reactions and store and transfer genetic information [66].

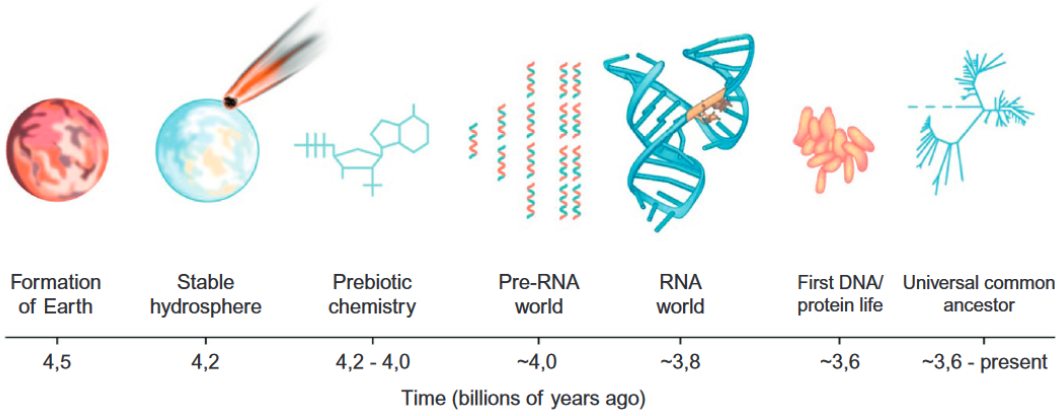


Figure 1.3. The main stages in the origin and evolution of life on Earth according to the RNA world hypothesis. Adopted and modified from [67].

According to one of the most recent hypotheses of the evolution of life, the appearance of RNA was preceded by non-coded synthesis of short and catalytically active peptides with beta hairpin structure from prebiotically available amino acids. These peptides formed ancestral aminoacyl-tRNA synthetases which were coupled with lately emerged proto-tRNA to give rise to the primitive version of the genetic code [68]. Since the formation of a peptide bond does not occur spontaneously, the synthesis of these small peptides at the earliest stage of the evolution of life could have been mediated by simple energy carrier molecules such as thioesters (thioester world) or inorganic pyrophosphate (pyrophosphate world) [69].

Thioesters rather than pyrophosphate could serve as a driving force for peptide bond formation, since phosphate minerals which might be the main source of prebiotic phosphorus, are either insoluble in water or have low reactivity [70,71]. The use of thioesters, most likely, derived from naturally abundant pantetheine and coenzyme A in the initial stage of life is strongly supported by the ease of prebiotic synthesis of organosulphur compounds in the electric discharge experiments, in which hydrogen sulphide was included in the simulated primitive Earth atmosphere [72]. The presence of hydrogen sulphide besides methane, ammonia and carbon dioxide leads to formation of

methionine, ethionine, homocysteic acid, S-methyl-cysteine, methionine sulfoxide, methionine sulfone and cysteamine under prebiotic conditions [73]. Mild heating of the mixture of cysteamine, β-alanine (synthetized in most prebiotic experiments) and pantoyl lactone (which could have been formed prebiotically from α-ketoisovaleric acid, formaldehyde and cyanide) was reported to yield the pantetheine moiety of coenzyme A [74,75,76]. Prebiotic availability of its precursor, pantetheine, and the central role in cellular metabolism suggest that coenzyme A emerged very early in the evolutionary history of life, dating back to the LUCA [77,78]. It was estimated that about 4 % of all known enzymes utilize coenzyme A or its thioesters as substrates [79]. Many coenzyme A dependent enzymes, namely citrate synthase, succinyl coenzyme A synthetase and acetyl coenzyme A synthetase have universal phylogenetic distribution and are thought to have been present in the LUCA, which indicates that coenzyme A must have been available earlier on, possibly already in the RNA world [78].

Several aminoacyl-tRNA synthetases are able to catalyse aminoacylation of cysteine and other thiols (pantetheine and coenzyme A) to yield aminoacyl-S-thioesters and cysteine-containing dipeptides [80]. Aminoacylation reactions of pantetheine and coenzyme A might have represented a primitive amino acid activating system before the appearance of proto-tRNA since both coenzyme A and tRNA can carry amino acids for peptide bond synthesis [80,81]. Thus, ancestral aminoacyl-tRNA synthetases might have acquired amino acid : pantetheine or amino acid : coenzyme A ligase activity before gaining ability to aminoacylate tRNA. Phylogenetic analyses of structural domains propose that catalytic domains of aminoacyl-tRNA synthetases could have emerged even before ribosomes about 3,6–3,7 billion years ago [82]. These observations constitute basis for the possible genetic code origin model, in which ancestral aminoacyl-tRNA synthetases might have mediated thioester-based non-coded peptide synthesis before the emergence of ribosomal translational machinery (Fig. 1.4, page 21) [80].

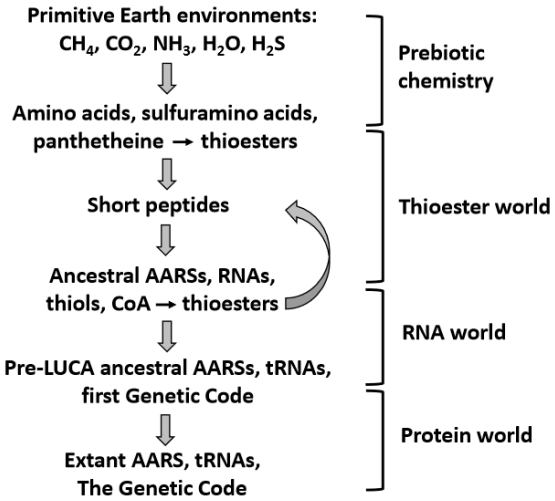


Figure 1.4. The possible scenario of the genetic code origin. Adopted from [80].

1.2.2 Coenzyme A biosynthesis

Coenzyme A is an essential cofactor that participates in numerous metabolic pathways and regulates key metabolic enzymes. It is crucial for reactions of central metabolism, where it functions as the major acyl group carrier (mainly for acetyl and succinyl groups) and participates in the biosynthesis and degradation of fatty acids, tricarboxylic acid cycle, synthesis of phospholipids, polyketides and nonribosomal peptides [83,84].

Coenzyme A is synthesized from pantothenate (vitamin B5), cysteine and ATP by a pathway of five enzymatic steps that is highly conserved across all domains of life except for certain variations in Archaea [84]. First, pantothenate kinase (PanK) catalyses the ATP-dependent phosphorylation of pantothenate to 4'-phopshopantothenate. 4'-phopshopantothenate is subsequently condensed with cysteine, which is in turn decarboxylated to generate 4'-phopshopantetheine by 4'-phosphopantothenoylcysteine synthetase (PPCS) and 4'-phosphopantothenoylcysteine decarboxylase (PPCDC), respectively. Phosphopantetheine adenylyltransferase (PPAT) transfers the AMP moiety of ATP to 4'-phopshopantetheine to yield 3'-dephosphocoenzyme A (dCoA). Finally, coenzyme A is formed by transfer of ATP-derived phosphate group to 3'-hydroxyl group of dCoA catalysed by dephosphocoenzyme A kinase (DPCK) (Fig. 1.5, page 22). It was shown that the first and last steps of coenzyme A biosynthesis, catalysed by PanK and DPCK, are rate limiting and allosterically regulated [83,85]. Both kinases utilize ATP as a phosphate donor and are crucial for many organisms [85].

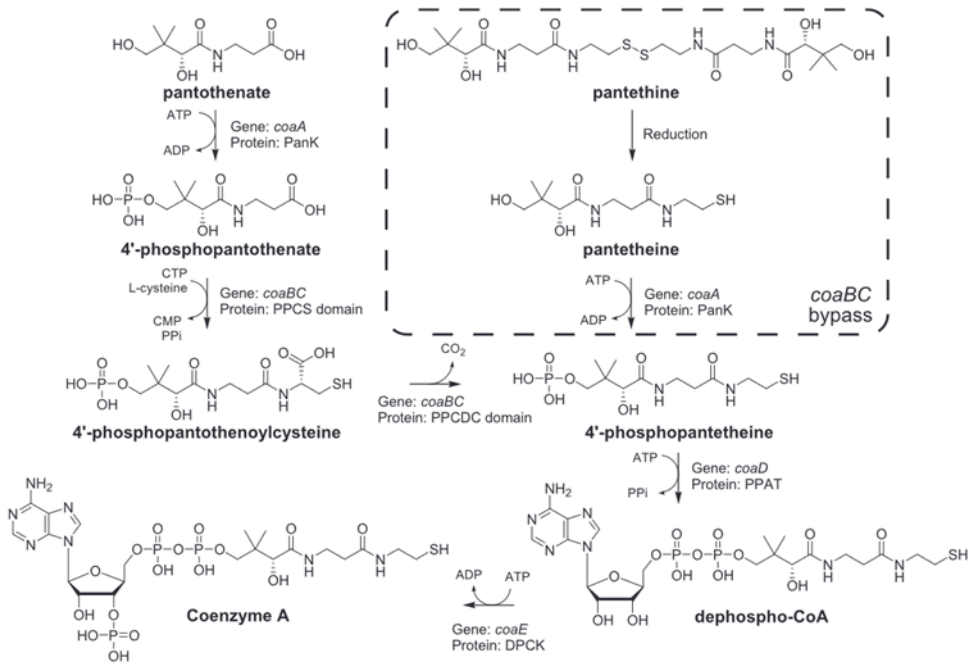


Figure 1.5. Biosynthetic pathway of coenzyme A in *Escherichia Coli*. Adopted from [86].

While many mammalians depend on uptake of exogenous pantothenate to fulfil their coenzyme A requirement, most bacteria, plants and fungi are capable of *de novo* biosynthesis of pantothenate by condensation of β-alanine with pantoate, which in turn is synthetized in two steps from α-ketoisovalerate [87,88]. Instead of pantothenate, coenzyme A can be biosynthesized directly from pantetheine, degradation product of coenzyme A, via the coenzyme A salvage pathway [86]. This pathway is parallel to coenzyme A biosynthesis pathway and bypasses the second and third steps of coenzyme A biosynthesis catalysed by PPCS and PPCDC.

In bacteria, coenzyme A plays a key role in biogenesis of cell envelope. Bacterial bifunctional enzyme GlmU, that catalyses two final steps of UDP-*N*-acetylglucosamine biosynthesis from fructose-6-phosphate, utilizes acetyl coenzyme A for the acetylation of glucosamine-1-phosphate to produce *N*-acetylglucosamine-1-phosphate, which is subsequently converted into UDP-*N*-acetylglucosamine by the transfer of uridine-5'-monophosphate [89] (Fig. 1.6, page 23). UDP-*N*-acetylglucosamine is a precursor to the main components of bacterial cell walls, namely peptidoglycan, essential bacterial structure that maintains the mechanical integrity of the cell, lipid A in Gram-negative bacteria and teichoic acid in Gram-positive bacteria, both of which are surface-exposed molecules that effect the permeability of the bacterial cell wall [90,91,92].

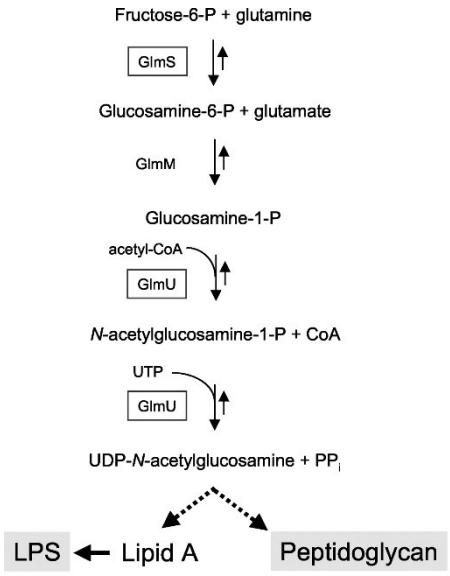


Figure 1.6. Biosynthetic pathway of UDP-*N*-acetylglucosamine in *Escherichia Coli*.

Adopted from [93].

The pathway of coenzyme A biosynthesis is considered an attractive target for developing novel antimicrobial agents due to the distinctive differences between mammalian and bacterial enzymes involved in this pathway [86]. There are three different types of PanK, bacterial types I and III and mainly eukaryotic type II, that differ in structural fold and catalytic and inhibition properties [94]. Eukaryotic PPCS and PPCDC are individual enzymes encoded by separate *coaB* and *coaC* genes, meanwhile, in prokaryotes PPCS and PPCDC activities are carried by single bifunctional PPCS/PPCDC protein encoded by *coaBC* gene [95]. The opposite situation is observed for the two final steps of coenzyme A biosynthesis, that are catalysed by two separate enzymes, PPAT (*coaD* gene) and DPCK (*coaE* gene), in bacteria, and single bifunctional coenzyme A synthase (*COASY* gene) in most eukaryotes.

1.2.3 Dephosphocoenzyme A kinase

Dephosphocoenzyme A kinase, the final enzyme in the biosynthesis of coenzyme A, catalyses MgATP-dependant phosphorylation of the 3'-hydroxyl group of the ribose moiety of dCoA to produce the final product, coenzyme A [96]. Bacterial DPCKs are highly conserved among different bacterial species and demonstrate good structural and sequence homology with DPCK domains of eukaryotic bifunctional coenzyme A synthases [97,98]. In contrast, the recently identified archaeal DPCK from *Thermococcus kodakarensis* does not show significant homology to previously identified bacterial DPCKs and DPCK domains of eukaryotic coenzyme A synthetases [99].

1.2.3.1 Catalytic properties

The first DPCK was identified and isolated from *Corynebacterium ammoniagenes,* and the N-terminal sequence of this enzyme was subsequently used to identify the gene encoding DPCK from *Escherichia coli* by a basic local alignment search tool (BLAST) search [100]. Dephosphocoenzyme A kinase from *Escherichia coli* of 206 amino acid residues has a molecular weight of 22,6 kDa and demonstrates a high level of activity over a broad pH range with the maximum activity at pH 8,5. Michaelis–Menten constants estimated for dCoA and ATP are 0,74 and 0,14 mM, respectively. The reported kinetic parameters of the other bacterial DPCKs and eukaryotic coenzyme A synthases are presented in Tab. 1.1.

Table 1.1. Kinetic parameters of DPCKs and coenzyme A synthases from different organisms

Organism	Substrate	V_{max} ($\mu mol \cdot min^{-1} \cdot mg^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1} \cdot M^{-1}$)	Reference
DPCKs						
Escherichia coli	dCoA	–	740	–	–	[100]
	ATP	–	140	–	–	
Mycobacterium tuberculosis	dCoA	–	34,9	0,029	830	[101]
	ATP	–	56,8	0,048	840	
Entamoeba histolytica (isotype 1)	dCoA	3,71	114	1,48	13500	[102]
	ATP	3,54	19,6	1,41	70500	
Entamoeba histolytica (isotype 2)	dCoA	2,48	57,9	0,96	16500	
	ATP	2,71	15	1,05	70000	
Coenzyme A synthases						
Pig	dCoA	4,37	4,1	–	–	[103]
	ATP	–	330	–	–	
Human	dCoA	5,11	5,2	–	–	
	ATP	5,18	192	–	–	

Dephosphocoenzyme A kinases demonstrate relatively low phosphate acceptor specifity. For example, DPCK from *Escherichia coli* is able to transfer phosphate group from ATP to adenosine, AMP and adenosin phosphosulphate with 4, 5 and 8 % activity compared with natural substrate, dCoA, respectively [100]. It was reported that DPCKs from *Escherichia coli* and *Aquifex aeolicus* are active towards short chain acyl thioesters of dCoA, namely dephosphoacetyl coenzyme A, dephosphosuccinyl coenzyme A and dephosphomalonyl coenzyme A [104]. In the case of DPCK from *Aquifex aeolicus* partial

activity was detected even with a long chain acyl thioester of dCoA, dephospho-*n*-heptadecanoyl coenzyme A.

Dephosphocoenzyme A kinases are considered to be generally MgATP-dependant, however, the archaeal DPCK from *Thermococcus kodakarensis* was reported to predominantly utilize GTP as the phosphate donor [99]. Enzyme assay with DPCK isolated from *Entamoeba histolytica* and various nucleoside triphosphates showed that ATP can be replaced by TTP or GTP with retaining 4,9–24,1 and 33 % activity compared with ATP, respectively [102]. Phosphotransferase activity of DPCK from *Entamoeba histolytica* is Mg^{2+}-dependent, however, Mg^{2+} can be replaced by various divalent metal cations such as Zn^{2+}, Cu^{2+} and Mn^{2+} with relative activity higher than 50 %, but no activity was observed after replacing Mg^{2+} by monovalent metal cations.

In the case of DPCK from *Mycobacterium tuberculosis* it was shown that CTP strongly binds to the enzyme at a site that overlaps with that of dCoA and, therefore, prevents binding of dCoA and limits catalysis [101]. CTP binding is considered to be one of the two regulatory mechanisms of DPCK along with oligomerization of the enzyme that will be discussed in section 1.2.3.3. More detailed kinetic studies proposed a sequential ordered mechanism of substrate binding for DPCK from *Mycobacterium tuberculosis*, upon which dCoA binds first to the enzyme followed by ATP binding.

1.2.3.2 Structural properties

To date, 17 different solved structures of DPCKs from 9 prokaryotic organisms are available in the Protein Data Bank (PDB), but only 4 of them are described in literature: DPCKs from *Haemophilus influenzae* (PDB code: 1JJV), *Escherichia coli* (PDB code: 1N3B), *Thermus thermophilus* (PDB code: 1UF9), and *Legionella pneumophila* (PDB codes: 4TTP, 4TTQ, 4TTR) [98,105,106,107].

The first three-dimensional structure was obtained for DPCK from *Haemophilus influenzae* in complex with ATP and provided a detailed insight into the catalytic mechanism of DPCKs (Fig. 1.7, page 26). Dephosphocoenzyme A kinases are members of the P-loop nucleotide triphosphate hydrolase superfamily, that comprises the various nucleoside and nucleotide kinases [108]. Similarly to the other members of this family, DPCKs comprise three domains: the nucleotide (ATP) binding domain, the substrate binding (CoA) domain and the lid domain [105].

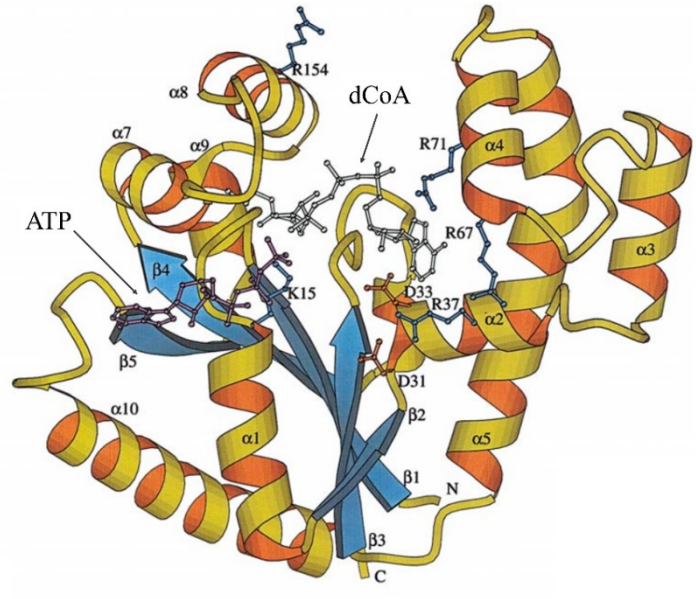


Figure 1.7. Three-dimensional X-ray structure of DPCK from *Haemophilus influenzae*. Ribbon diagram of polypeptide fold with both substrates, ATP and dCoA (indicated by arrows). Adopted from [105].

The ATP-binding domain is the core domain of DPCK and formed by five parallel β-strands (β5–β4–β1–β3–β2), flanked by α-helices on both sides [105]. This domain includes the P-loop, which is also called Walker A motif, that is responsible for the binding of β-phosphate of ATP. The P-loop has the sequence of GXXXXGKS/T (where X stands for any amino acid) and is found in all previously identified and isolated DPCKs. Gly14 and Lys15 residues of the P-loop are strictly conserved in DPCKs sequences. The P-loop creates a large anion hole with a positive electrostatic potential provided by the amino groups of the loop and the dipole of the helix α1. The oxygen atoms of the phosphate groups form numerous hydrogen bonds to the backbone amino groups of the P-loop. The ribose moiety of ATP does not interact with DPCK, meanwhile, the adenine base is accommodated between the side chains of Arg140 and Leu179 and forms hydrogen bonds with Asn175, Ala177 and Leu179. The P-loop usually does not undergo a conformational change during ATP binding in P-loop kinases, however, unusual drastic conformational change of the P-loop from 3_{10}-helix to extended loop conformation upon ATP binding was observed for DPCK from *Legionella pneumophila* [107].

Dephosphocoenzyme A kinases, like the other phosphotransferases, require Mg^{2+} for stabilizing the negative charges on the phosphates to facilitate the phosphoryl transfer [105]. In P-loop kinases coordination of Mg^{2+} is usually supported by the hydroxyl group of serine in the P-loop and the carboxylate of aspartate or glutamate in the strand β3 [109].

However, in DPCK from *Haemophilus influenzae* Mg^{2+} seems to be coordinated by the carboxylate of Asp31 residue in the end of the strand β2 since no equivalent amino acid in the strand β3 was identified [105].

The CoA domain consists of four α-helices (α2–5), inserted between the strands β2 and β3 [105]. The ATP-binding and the CoA domains create the deep cleft that appears to be the putative site for dCoA binding and catalysis. The adenine base of dCoA is positioned in the hydrophobic environment created by the side chains of Ile92, Met96, Val112, Leu114 and Leu120, and forms hydrogen bonds with His89 and Asn119 at the bottom of the binding site. The pantothenate and cysteamine moieties of dCoA are accommodated in the narrow pocket between the ATP-binding domain and the lid domain. The resulting orientation of the adenine base ensures the spatial proximity of 3'-hydroxyl group of the ribose moiety to Asp33 in the catalytic centre. The highly conserved Asp33 is a key catalytic amino acid residue, that is suggested to activate dCoA for nucleophilic attack on the γ-phosphate of ATP. In most DPCKs, the catalytic aspartate residue is placed in the sequence in the tandem with another aspartate forming a DXD motif, where X stands for any amino acid [101]. The neighbouring aspartate (Asp31) participates in coordinating Mg^{2+} in DPCK from *Haemophilus influenzae* [105]. Apart from Asp33, several arginine residues, namely Arg37, Arg67, Arg71 from the CoA domain and Arg144 and Arg154 from the lid domain, form the catalytic site of the enzyme and are involved in substrate binding and the reaction intermediate stabilization.

The lid domain is formed by two α-helices (α7–8), deposited between the strands β4 and β5, and shields the active site of the protein to prevent transfer of the phosphate group to water [105].

The calculated distance between 3'-hydroxyl group of dCoA and the γ-phosphate of ATP is about 7 Å, and, therefore, transfer of the phosphate group requires significant conformational change of the protein [105]. Both the CoA and the lid domains are mobile and close over the catalytic site during catalysis to direct phosphate group transfer after binding of dCoA. In the absence of dCoA, DPCK exists in an open conformation. Two proline residues, Pro90 and Pro 134, may play an important role in the domains closure. Pro90 causes the kink in the middle of the helix α5 and might serve as a hinge for the movement of the CoA domain. Transition to the closed conformation after ATP binding might be driven by electrostatic interactions between positively charged arginine residues in the CoA and the lid domains and negatively charged ATP phosphate groups [106].

In conclusion, there are only five residues strictly conserved among DPCKs from different sources [105]. These include Gly14 and Lys15 of the P-loop; Arg140, which interacts with the adenine base of ATP; Thr8, which is involved in binding of dCoA; and Asp33, which is thought to be a key catalytic amino acid.

1.2.3.3 Oligomerization

Unlike the monomeric DPCK from *Haemophilus influenzae*, DPCK from *Escherichia coli* was crystallized in trimeric form stabilized by the presence of sulphate ions [98]. The sulphate ion in the centre of the trimer interface is surrounded by Ser119, Tyr121, and Lys122 from each monomer, but only Tyr121 is involved in hydrogen-bonding interactions with the sulphate. These residues are positioned in the second half of the loop between strands β3 and β4 and are not present in the sequence of DPCK from *Haemophilus influenzae*, which explains the difference in oligomeric states of the two proteins. Another group of residues that contributes to the trimerization of DPCK from *Escherichia coli*, is a five-residues extension at the C-terminus of the protein that wraps around the neighbouring monomers. The other sulphate ions mimic the phosphate groups of ATP and occupy positions approximate to those of the β-phosphate of ATP in all three chains of the trimer. The arrangement of the three subunits enables the free movement of the CoA and the lid domains to close over the catalytic site and leaves the putative catalytic site oriented to the exterior of the trimeric domain.

The central sulphate ion has been shown to stabilize the loop between strands β3 and β4 and was suggested to be replaced by a phosphate ion *in vivo*, however, the biological role of the trimeric arrangement of DPCK from *Escherichia coli* has not been determined [98]. Similar sulphate-dependent oligomerization is also observed for DPCKs from *Thermus thermophilus* and *Aquifex aeolicus*.

The presence of sulphate ions also induces the formation of trimers in solution and leads to dynamic equilibrium between monomeric and trimeric forms of the protein, which was demonstrated by size-exclusion chromatography with purified protein in buffer with and without 0,2 M ammonium sulphate. In the absence of sulphate, DPCK from *Escherichia coli* is eluted in a single peak with a volume equivalent to a monomeric molecular mass of 22 kDa, whereas the presence of ammonium sulphate leads to emergence of an additional peak corresponding to a molecular mass of 66 kDa which indicates oligomerization of the protein.

It was shown that emergence of a dynamic equilibrium between trimeric and monomeric forms of DPCK from *Mycobacterium tuberculosis* plays an important role in regulation of enzyme activity [110]. The native mycobacterial DPCK exists in solution as a mixture of more active monomeric and less active trimeric forms of equal abundance, but incubation with dCoA increases the proportion of the monomeric protein as confirmed by size-exclusion chromatography. These observations allow to conclude that the presence of dCoA in cells induces the conversion of the enzyme from trimeric to monomeric form and thus regulates biosynthesis of coenzyme A. No effect on the oligomeric status of DPCK was observed in the case of the second substrate, ATP.

Enzyme assays established that the monomeric form is the more active form of the enzyme [110]. It was assumed that in the cell, DPCK is possibly sequestered in an inactive trimeric form which is subsequently converted to the active, monomeric form in the presence of dCoA. This conversion then facilitates the binding of the second substrate, ATP. This could be one of possible regulatory mechanisms employed by the cell to regulate the biosynthesis of coenzyme A.

2 Aims of the thesis

- Recombinant expression of wild type protein and two aromatic amino acid-lacking mutants of DPCK from *Aquifex aeolicus*.
- Purification and biophysical characterization of recombinantly expressed proteins.
- Characterization of enzyme activity of recombinantly expressed proteins.

3 Materials and methods

3.1 Instruments and equipment

Analytical balances	
M254Ai	BEL Engineering, Italy
Apparatus for agarose gel electrophoresis	
Mini-Sub Cell GT Cell	Bio-Rad, USA
Apparatus for SDS-PAGE	
Mini-PROTEAN Tetra Cell	Bio-Rad, USA
Centrifuges	
Centrifuge 5424R equipped with FA-45-24-11 fixed angle rotor	Eppendorf, Germany
Minicentrifuge C1301-B	Labnet International, USA
Ultracentrifuge Avanti JXN-30 equipped with JA-30.50 Ti fixed angle rotor and JA-10 fixed angle rotor	Beckman Coulter, USA
Chromatography systems	
Chromatography system NGC Quest 10 Plus	Bio-Rad, USA
HPLC system Dionex Ultimate 3000 RS equipped with TSQ Quantiva MS detector	Thermo Fisher Scientific, USA
Columns	
Column HiLoad 16/600 Superdex 200 pg	GE Healthcare, USA
Column HisTrap HP, 5 ml	GE Healthcare, USA
Column HiTrap Capto Q, 5 ml	GE Healthcare, USA
Column Superdex 75 10/300 GL	GE Healthcare, USA
ZIC-HILIC HPLC column	Millipore, USA
Electrophoresis power supply	
PowerPac Basic	Bio-Rad, USA
Incubators	
Incubator MyTemp Mini	Benchmark scientific, China
Incubator shaker Innova 44	New Brunswick, Canada

Thermocell mixing block MB-102	Hangzhou Bioer Technology, China
Magnetic stirrer	
HSC	VELP Scientifica, Italy
pH meter	
pH/ORP benchtop meter HI3220	Hanna instruments, USA
Shaker	
ProBlot Rocker 35 EZ	Labnet International, USA
Scales	
Adventurer AX2202	OHAUS, USA
Sonicator	
S-3000 equipped with 1/16 microtip probe tapered 418	Misonix, USA
Spectrophotometer	
Microplate reader CLARIOstar Plus	BMG LABTECH, Germany
Nanodrop 2000	Thermo Fisher Scientific, USA
Other equipment	
Amino Acid Analyser Biochrom 30+ Series	Biochrom, UK
Centrifugal filters Ultrafree-MC, 0,22 μm	Millipore, USA
Centrifugal unit Amicon Ultra, MWCO 10000, 4 ml	Millipore, USA
Circular dichroism spectrophotometer Jasco J-1500	JASCO, Japan
Gel documentation system InGenius 3	Syngene, India
Thermal cycler GeneQ TC-24/H(b)	Hangzhou Bioer Technology, China
UltrafleXtreme MALDI-TOF/TOF mass spectrometer	Bruker, Germany
Zetasizer Ultra	Malvern Panalytical, UK

3.2 Chemicals

3.2.1 Chemicals, enzymes and commercial kits

Bio-rad, USA: Precision Plus Protein Dual Xtra Prestained Protein Standard, Bio-Rad Protein Assay Kit II.

Biotium, USA: Gelred nucleic acid gel stain (10000x).

Eurofin DiscoverX Corporation, USA: ADP Quest Assay kit.

GE Healthcare, USA: Gel filtration low molecular weight calibration kit.

Jena Bioscience, Germany: Pfu-X Polymerase from *Pyrococcus furiosus* (2,5 U/µl), Pfu-X buffer (10x).

Linde, Czech Republic: Liquid nitrogen.

New England BioLabs, USA: Antarctic phosphatase (5 U/µl), restriction endonuclease *XhoI* (20 U/µl), restriction endonuclease *NdeI* (20 U/µl), CutSmart buffer (10x), antarctic phosphatase (5 U/µl), antarctic phosphatase reaction buffer (10x), instant sticky-end master mix based on T4 DNA ligase (2x).

PENTA, Czech republic: Isopropanol (≥ 99,8 %), acetic acid (≥ 99,8 %).

Qiagen, Germany: GelPilot DNA Loading dye (5x), GelPilot 100 bp Plus Ladder, GelPilot 1 kb Ladder.

SERVA Electrophoresis GmbH, Germany: Acrylamide/bis solution (40% w/v, 29:1).

Sigma-Aldrich, USA: Kanamycin sulphate from *Streptomyces kanamyceticus* (molecular biology grade), IPTG (≥ 99 %), agarose (molecular biology grade), Tris (≥ 99,8 %), Tris-HCl (≥ 99 %), SDS (≥ 99,5 %), APS (≥ 98 %), TEMED (≥ 99 %), imidazole (≥ 99,5 %), zinc sulphate heptahydrate (≥ 99 %), EASYpack protease inhibitor cocktail, dCoA (≥ 90 %), ATP disodium salt hydrate (≥ 99 %), acetonitrile (HPLC grade, ≥ 99,9 %), ammonium carbonate (≥ 99,9 %), ammonium hydrocarbonate (≥ 99,5 %), lysozyme from hen egg white (≥ 98 %), RNase A from bovine pancreas (molecular biology grade), endoproteinase Lys-C from *Lysobacter enzymogenes* (sequencing grade), EDTA (≥ 99 %), glycerol (≥ 99,5 %), bromophenol blue (molecular biology grade), β-mercaptoethanol (≥ 99 %), glycine (≥ 99 %), sodium chloride (≥ 99 %), Tween-20 (molecular biology grade), TCEP (≥ 98 %), potassium chloride (≥ 99 %), magnesium chloride hexahydrate (≥ 99 %), DTT (molecular biology grade), hydrochloric acid (molecular biology grade), trifluoroacetic acid (HPLC grade, ≥ 99,9 %), sinapinic acid (≥ 99,5 %).

Thermo Fisher Scientific, USA: dNTP mix (10 mM each), B-PER bacterial protein extraction reagent, DNase I (RNase-free, 1 U/µl).

ZymoResearch, USA: Zymoclean DNA gel recovery kit, DNA Clean & Concentrator kit, Zyppy plasmid miniprep kit.

3.2.2 Cell strains

- BL21 (DE3) *Escherichia Coli* chemically competent cells (Thermo Fisher Scientific, USA).
- One Shot TOP10 *Escherichia Coli* chemically competent cells (Thermo Fisher Scientific, USA).

3.2.3 Cultivation media

- LB Broth (Sigma Aldrich, USA).
- LB Agar (Sigma Aldrich, USA).

3.2.4 Plasmid DNA

- pET-24a (+) expression vector (Novagen, Germany).

3.2.5 Primers

3.2.5.1 PCR primers

DPCK forward primer	5′-AAAAACATATGAAACGTATCGGTCTGACC-3′
DPCK reverse primer	5′-AAAAACTCGAGTTCCAGCGGGTCACGG-3′

3.2.5.2 Sequencing primers

T7 promoter primer	5′-TAATACGACTCACTATAGGG-3′
T7 terminator primer	5′-GCTAGTTATTGCTCAGCGG-3′

3.2.6 Buffers

- TAE buffer: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA.
- SDS-PAGE sample buffer (5x): 250 mM Tris, 10 % (w/v) SDS, 0,5 % (w/v) bromophenol blue, 25 % (v/v) β-mercaptoethanol, 50 % (v/v) glycerol, pH 6,8.
- SDS-PAGE running buffer: 25 mM Tris, 192 mM glycine, 0,1 % (w/v) SDS, pH 8,3.

- PBS buffer: 10 mM Na_2HPO_4, 1,8 mM KH_2PO_4, 137 mM NaCl, 2,7 mM KCl, pH 7,4.
- LYS buffer: 20 mM Tris, 20 mM NaCl, 1 mM β-mercaptoethanol, pH 8,0.
- IonA buffer: 20 mM Tris, 20 mM NaCl, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 8,0.
- IonB buffer: 20 mM Tris, 1 M NaCl, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 8,0.
- NiA buffer: 20 mM Tris, 500 mM NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 7,6.
- NiB buffer: 20 mM Tris, 500 mM NaCl, 500 mM imidazole, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 7,6.
- GF buffer: 50 mM Tris, 500 mM NaCl, 20 mM KCl, 10 mM $MgCl_2$, 0,5 mM DTT, pH 7,6.
- DIAL buffer: 50 mM Tris, 500 mM NaCl, 20 mM KCl, 10 mM $MgCl_2$, 0,5 mM TCEP, pH 7,6.
- Lys-C buffer: 25 mM Tris, 300 mM NaCl, 1 mM EDTA, 0,5 mM TCEP, pH 8,0.
- HPLC buffer: 25 mM NH_4HCO_3, 300 mM NaCl, 20 mM KCl, 10 mM $MgCl_2$, pH 7,4.

3.3 Methods

3.3.1 DNA manipulation

3.3.1.1 Polymerase chain reaction

The DPCK synthetic genes were dissolved in distilled water to the final concentration of 50 ng/µl and amplified using the standard protocol for DNA amplification by Pfu-X polymerase. Briefly, the reaction mixture was prepared by mixing on ice 50 ng of gene, 1 µl of 10 mM dNTP mix, 2 µl of 10 mM DPCK forward primer, 2 µl of 10 mM DPCK reverse primer, 5 µl of 10x Pfu-X buffer and water to adjust the volume to 50 µl. After 0,5 µl of 2,5 U/µl Pfu-X polymerase was added, the reaction mixture was placed into thermocycler and PCR was run according to the following program (Tab. 3.1, page 36).

Table 3.1. PCR amplification program

Step	Temperature (°C)	Time (s)	Number of cycles
Initial denaturation	95	120	1
Denaturation	95	20	32
Annealing	56	30	32
Elongation	68	30	32
Final elongation	68	180	1

3.3.1.2 Horizontal agarose gel electrophoresis

First, 1% agarose gel was prepared by resuspending 0,3 g agarose in 30 ml of TAE buffer followed by heating in microwave (600 V, 2–3 minutes) to dissolve agarose completely and cooling down to 45–50 °C. Then, 3 µl of 10000x Gelred nucleic acid gel stain was added, and the agarose solution was poured into a gel tray with the well comb placed. After agarose gel was completely solidified, it was transferred to the electrophoresis unit, and the electrophoresis unit was filled with TAE buffer to cover the surface of agarose gel completely. Then, DNA samples were mixed with 2 µl of 5x GelPilot DNA Loading dye and water to the final volume of 10 µl, and prepared samples were loaded into the gel wells. The first lane was always loaded with 3 µl of GelPilot 100 bp Plus Ladder or GelPilot 1 kb Ladder. Electrophoresis was run at 100 V for approximately 50 minutes, and after the dye line was 75–80 % of the way down the gel, electrophoresis was stopped, and the DNA bands were visualized by illuminating the gel with UV light in InGenuis3 gel documentation system.

3.3.1.3 DNA extraction from agarose gel

DNA was extracted from the agarose gel using Zymoclean DNA gel recovery kit according to the manufacturer's instructions. This commercial kit includes AD buffer, DNA wash buffer and Zymo-Spin columns. The compositions of both buffers are not specified by the manufacturer.

First, excised agarose gel slice with the desired DNA band was weighed out and AD buffer was added in proportion 300 µl of AD buffer for 100 mg of agarose, followed by incubation at 37 °C for 10 minutes to completely dissolve the gel slice. Then, the melted agarose solution was placed into a Zymo-Spin column in a collection tube and centrifuged for 60 seconds at 15000x g. The flow-through was discarded and the column was washed 2 times with 200 µl of DNA wash buffer followed by centrifugation for 30 seconds at 15000x g. Finally, the column was placed into a 1,5 ml tube and 15 µl of warm water (50 °C) was applied directly to the column matrix followed by centrifugation for 60 seconds at 15000x g.

3.3.1.4 Measurement of DNA concentration and purity

The concentration and purity of DNA samples were estimated spectrophotometrically using Nanodrop 2000. The DNA concentration was evaluated by pipetting 1 µl of the sample onto the sample surface and measuring absorbance at 260 nm. The purity of DNA was assessed by measuring and calculating the ratio of absorbance at 260 nm and 280 nm.

3.3.2 Molecular cloning

3.3.2.1 Restriction digestion of insert and linearized vector

Reaction mixtures for restriction digestion of amplified *DPCK* genes were prepared by mixing 1 µg of DNA, 0,5 µl of 20 U/µl *XhoI*, 0,5 µl of 20 U/µl *NdeI*, 2,5 µl of 10x CutSmart buffer and water to adjust the volume to 25 µl. Then, the mixtures were incubated at 37 ℃ for 2 hours and restriction endonucleases were inactivated by heating the reaction mixture at 65 °C for 20 minutes.

DNA restriction fragments were recovered using DNA Clean & Concentrator kit according to the standard procedure provided by the manufacturer. This commercial kit includes DNA binding buffer, DNA wash buffer and Zymo-Spin columns. The compositions of both buffers are not specified by the manufacturer.

Briefly, the reaction mixtures were mixed with 250 µl of DNA binding buffer, transferred into Zymo-Spin column in a collection tube and centrifuged for 30 seconds at 15000x g. The flow-through was discarded and the column was washed 2 times with 200 µl of DNA wash buffer followed by centrifugation for 30 seconds at 15000x g. Finally, the column was placed into a 1,5 ml tube and 10 µl of water was applied directly to the column matrix followed by centrifugation for 60 seconds at 15000x g. The concentration of eluted DNA was estimated spectrophotometrically, as described in section 3.3.1.4.

pET-24a (+) expression vector was digested using the same protocol, but 3 µg of DNA was used for the reaction. The digested vector was analysed by agarose gel electrophoresis, and extracted from the gel as described in section 3.3.1.3. The concentration and purity of extracted digested vector was estimated spectrophotometrically as described in section 3.3.1.4.

3.3.2.2 Dephosphorylation of vector

To prevent self-ligation, plasmid DNA after restriction digestion was dephosphorylated before ligation. Dephosphorylation was performed by preparing the reaction mixture

consisting of 1 µg of pET-24a (+) expression vector cleaved by *XhoI* and *NdeI*, 1 µl of 5 U/µl Antarctic phosphatase, 2 µl of 10x Antarctic phosphatase reaction buffer and water to adjust the volume to 20 µl. The reaction mixture was incubated at 37 ℃ for 30 minutes, then antarctic phosphatase was inactivated by heating at 80 °C for 2 minutes. Dephosphorylated plasmid DNA was recovered using DNA Clean & Concentrator kit, and the concentration of eluted plasmid DNA was estimated spectrophotometrically as described in section 3.3.1.4.

3.3.2.3 Ligation reaction

Ligation was performed using Instant sticky-end master mix based on T4 DNA ligase. 20 ng of cleaved and dephosphorylated pET-24a (+) expression vector was mixed with 5,4 ng of the cleaved gene and water to adjust the volume to 10 µl, so the molar ratio of the vector and insert was 1:6 to maximize the yield of successfully ligated product. After thawing on ice, 5 µl of 2x Instant sticky-end master mix was added, and the whole reaction mixture was thoroughly pipetted up and down 8–10 times and placed on ice. The reaction mixture was used immediately for transformation of One Shot TOP10 *Escherichia Coli* chemically competent cells.

3.3.2.4 Transformation

First, 50 µl of One Shot TOP10 *Escherichia Coli* chemically competent cells were thawed on ice for 10 minutes, then 5 µl of ligation mixture was added. After incubation on ice for 25 minutes, cells were heat shocked at 42 ℃ for 60 seconds and placed immediately on ice for 2 minutes. Then, cells were diluted with 250 µl of sterile LB medium without antibiotic and grown in 37 ℃ shaking incubator at 220 rpm for 1 hour. Finally, 150 µl of cell culture was plated onto a 10 cm LB agar plate containing 50 µg/ml kanamycin sulphate and incubated overnight at 37 ℃.

3.3.2.5 Colony polymerase chain reaction

Bacterial cell colonies were screened for the presence of insert using colony PCR. Two single cell colonies were picked from the agarose plate by 10 µl pipette tips and swirled in 20 µl of sterile water. Then, PCR amplification reaction mixtures were prepared by mixing the following on ice: 1 µl of bacterial cells-water suspension, 1 µl of 10 mM dNTP mix, 2 µl of 10 mM DPCK forward primer, 2 µl of 10 mM T7 terminator primer and 5 µl of 10x Pfu-X buffer and water to adjust the volume to 50 µl. After 0,5 µl of 2,5 U/µl Pfu-X polymerase was added, the reaction mixture was placed into the thermocycler and PCR

was run according to the following program (Tab. 3.2). The initial denaturation step was prolonged until 5 minutes to lyse bacterial cells and release plasmid DNA.

Table 3.2. Colony PCR amplification program

Step	Temperature (°C)	Time (s)	Number of cycles
Initial denaturation	95	300	1
Denaturation	95	20	32
Annealing	56	30	32
Elongation	68	30	32
Final elongation	68	180	1

Products of PCR amplification were analyzed by agarose gel horizontal electrophoresis to identify the presence of the insert. The leftover bacterial cells-water suspensions of positive clones were used to start overnight liquid cultures.

3.3.2.6 Small scale plasmid DNA isolation

Plasmid DNA was isolated from the cells using Zyppy plasmid miniprep kit according to the manufacturer's protocol. This commercial kit includes 7x Lysis buffer, Neutralization buffer, Endo-Wash buffer, Zyppy Wash buffer and Zymo-Spin IIN columns. The compositions of all buffers are not specified by the manufacturer.

Leftover bacterial cells-water suspension of the positive clone was transferred into 5 ml of LB medium containing 50 μg/ml kanamycin sulphate, and bacterial culture was grown overnight at 37 °C while shaking at 220 rpm. Then, 600 μl of bacterial culture was transferred to a 1,5 ml microcentrifuge tube and gently mixed by inverting the tube 4–6 times with 100 μl of 7x Lysis buffer. After incubation for 2 minutes, when the solution became clear blue, cell lysis was stopped by adding 350 μl of cold Neutralization buffer. The solution was mixed thoroughly and centrifuged for 3 minutes at 15000x g. After centrifugation the supernatant was transferred into a Zymo-Spin IIN column in a collection tube and centrifuged for 20 seconds at 15000x g. The flow-through was discarded and the column was washed, first, with 200 μl of Endo-Wash buffer and, second, with 400 μl of Zyppy Wash buffer, each wash step was followed by centrifugation for 30 seconds at 15000x g. Finally, the column was placed into a 1,5 ml tube and 30 μl of water was applied directly to the column matrix followed by centrifugation for 30 seconds at 15000x g. The concentration and purity of eluted DNA was estimated spectrophotometrically, as described in section 3.3.1.4, and then plasmid DNA was stored at −20 °C.

3.3.2.7 Sequencing of plasmid DNA

Isolated plasmid DNA was submitted for Sanger sequencing to confirm the sequence of the insert. The reaction mixtures for Sanger sequencing were prepared by mixing 100 ng of plasmid DNA, 2 µl of 10 mM T7 promoter primer or T7 terminator primer and water to adjust the volume to 20 µl. The prepared samples were sent to DNA sequencing service laboratory (OMICS Genomics facility, BIOCEV), and the sequence of plasmid DNA was read in both directions. After receiving the resulting nucleotide sequence, the obtained data were compared to the expected sequence using the Clustal Omega Multiple Sequence Alignment online program.

3.3.3 Protein manipulation

3.3.3.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

First, 14% polyacrylamide resolving gel was prepared by mixing the following components (Tab. 3.3).

Table 3.3. Composition of 14% polyacrylamide resolving gel

Component	Volume (ml)
Tris-HCl (1,5 M, pH 8,8)	1,25
Acrylamide/bis solution (40% w/v, 29:1)	1,6
10% SDS (w/v)	0,05
10% APS (w/v)	0,05
TEMED	0,005
Water	to 5 ml

To prepare 1,0 mm thick gel two glass plates were clamped in the casting frame, and gel mixture was carefully pipetted into the gap between glass plates, leaving 2 cm of below the bottom of the comb for stacking gel. The top of the gel was layered with isopropanol, and the resolving gel was left for 30 minutes at room temperature to polymerise. After that, isopropanol was discarded, the top of the gel was washed with water and the polymerized resolving gel was overlaid by stacking gel mixture composed of the following components (Tab. 3.4).

Table 3.4. Composition of 6,6% polyacrylamide stacking gel

Component	Volume (ml)
Tris-HCl (1,0 M, pH 6,8)	0,625
Acrylamide/bis solution (40% w/v, 29:1)	0,375
10% SDS (w/v)	0,025
10% APS (w/v)	0,025
TEMED	0,005
Water	to 2,5 ml

The well comb was then inserted into the polymerizing stacking gel. After 30 minutes the well comb was removed, the glass plates were removed out of the casting frame and set into the cell buffer dam. 800 ml of SDS-PAGE running buffer was poured first into the inner chamber to completely cover the gel, and then into the outer chamber until the buffer surface reached the required level in the outer chamber.

Samples for SDS-PAGE were diluted with water to the final volume of 8 µl if necessary, mixed with 2 µl of 5x SDS-PAGE sample buffer and boiled at 95 °C for 5 minutes. After samples were cooled down to room temperature, centrifuged at 30 seconds for 15000x g and loaded into the comb wells. To estimate the molecular weight of separated proteins, 3 µl of Precision Plus Protein Dual Xtra Prestained Protein Standard was loaded into the first lane of the gel. The electrophoresis was run at a constant voltage of 150 V for approximately 50 minutes until the bromophenol blue dye front reached the bottom of the gel. After the electrophoresis was completed, the gel was removed from the glass plates, rinsed with water and visualized by imidazole-zinc reverse staining.

3.3.3.2 Imidazole–zinc reverse staining of proteins in polyacrylamide gel

Imidazole–zinc reverse staining provides rapid and sensitive (as few as 1,8 ng) detection of proteins fractionated by polyacrylamide gels or SDS denatured gels [111]. It is based on the ability of imidazole to react with Zn^{2+} to produce insoluble zinc imidazolate $ZnIm_2$ [112]. Gel after electrophoresis is first soaked into an imidazole solution, which attaches to the whole surface of the gel unoccupied by protein bands – sodium dodecyl sulphate coating on the protein prevents the stains from binding to the proteins. After removing the imidazole solution and adding zinc sulphate solution, imidazole attached on the gel surface reacts with Zn^{2+}, forming white precipitate of zinc imidazolate – that enables detection of clear unstained protein bands on a deep white stained background.

Gel was rinsed with water, placed in a container with 30 ml of 0,2 M imidazole and agitated on shaker at room temperature for 10 minutes. Then, the imidazole solution was discarded, and the gel was developed by agitating in 20 ml of 0,3 M zinc sulphate for 30 seconds. After disposing zinc sulphate solution, the gel was rinsed two times with water and photographed against black background.

3.3.3.3 Protein concentration determination

Protein concentrations during purification were determined using Bio-Rad Protein Assay Kit II according to manufacturer's instructions. Bio-Rad protein assay is based on the

Bradford method for protein quantification that relies on the shift of absorbance maximum of Coomassie Brilliant Blue G-250 dye from 465 nm to 595 nm after binding to protein under acidic conditions [113]. This kit includes bovine serum albumin standard and 5x Bio-Rad Protein Assay Dye Reagent Concentrate. The composition of the dye reagent is not specified by the manufacturer.

Protein samples were diluted 20 times with PBS buffer. Then, 10 µl of PBS buffer (blank), serially diluted bovine serum albumin standard with concentrations of 1; 2; 4; 6; 8; 10 mg/ml and diluted protein samples were pipetted into separate microtiter plate wells. After that, 190 µl of Bio-Rad Protein Assay Dye Reagent Concentrate diluted 5 times with PBS buffer was added to each well, and the samples were mixed on a shaker for 1 minute and after incubated for 10 minutes at room temperature. After incubation, the absorbance at 595 nm was measured using CLARIOstar Plus microplate reader. All measurements were performed in triplicates, and protein concentrations were calculated from the calibration curve constructed from the absorbance values of the BSA standards.

Additionally, protein concentrations after purification were determined by amino acid analysis using Biochrom 30+ Series Amino Acid Analyser. The amino acid analysis was kindly performed by Radko Souček, Ph.D. (IOCB Prague).

3.3.4 Expression and isolation of recombinant proteins

3.3.4.1 Protein expression and solubility analysis

After the presence of insert was confirmed, BL21 (DE3) *Escherichia Coli* chemically competent cells were transformed with the plasmid containing the gene of interest following the heat shock protocol for bacterial transformation. The transformed cells were plated into a LB agar plate containing 50 µg/ml kanamycine sulphate and incubated overnight at 37 °C.

Before large-scale production, recombinant proteins were first produced in a small volume of LB medium and tested for their solubility and expression level for further optimization of expression conditions if necessary. A starter culture was established overnight from a single colony selected from the agar plate and grown in 5 ml of LB medium containing 50 µg/ml kanamycin sulphate in 37 °C shaking incubator at 220 rpm. The next day, 100 µl of the starter culture was inoculated in 10 ml of fresh LB medium supplemented with 50 µg/ml kanamycin sulphate, and cells were grown for approximately 3 hours in 37 °C shaking incubator at 220 rpm. Cell growth was monitored

by measuring OD_{600} every 30 minutes. When OD_{600} had reached 0,75–0,8, the cell culture was divided into two equal aliquots. In one half of the cell culture expression of recombinant protein was induced by adding IPTG to the final concentration of 0,5 mM, and the second half of the cell culture was not induced by IPTG. The cells were grown for additional 4 hours in 37 °C shaking incubator at 220 rpm. After 4 hours OD_{600} of both cultures were measured, and 300 µl of the induced cell culture and the uninduced cell culture with the volume corresponding to the same amount of cells, were transferred to 1,5 ml tubes, and the cells were harvested by centrifugation at 5000x g at 22 °C for 10 minutes. Then, the cell pellet was resuspended in 100 µl of B-PER bacterial protein extraction reagent pre-mixed with 0,5 µl of 25 mg/ml lysozyme from hen egg white and 0,5 µl of 10 mg/ml RNase A from bovine pancreas, and the cells suspension was incubated at room temperature for 10 minutes to complete cell lysis. The incubation was followed by centrifugation at 15000x g for 5 minutes to separate insoluble fraction of protein with cell debris from soluble fraction of protein. 50 µl of supernatant was transferred to a new 1,5 ml tube and mixed with 12 µl of 5x SDS-PAGE sample buffer, and cell debris after removing remaining supernatant was resuspended in 50 µl of water pre-mixed with 12 µl of 5x SDS-PAGE sample buffer. Both samples were heated at 95 °C for 5 minutes, cooled down to room temperature, and 10 µl of each sample was separated by SDS-PAGE to estimate the amount of recombinant protein in soluble and insoluble fraction.

3.3.4.2 Large-scale production of recombinant proteins

A starter culture was established overnight from a single colony selected from the agar plate and grown in 10 ml of LB medium containing 50 µg/ml kanamycin sulphate in 37 °C shaking incubator at 220 rpm. The next day, 500 ml of fresh LB medium supplemented with 50 µg/ml kanamycin sulphate was inoculated with 5 ml of the starter culture, and cells were grown for approximately 3 hours in 37 °C shaking incubator at 220 rpm. Cell growth was monitored by measuring OD_{600} every 30 minutes. When OD_{600} had reached 0,75–0,8, expression of recombinant protein was induced by adding IPTG to the final concentration of 0,5 mM, and cells were grown for additional 4 hours in 37 °C shaking incubator at 220 rpm.

3.3.4.3 Cell harvesting and lysis

After induction cells were harvested by centrifugation at 3000x g at 4 °C for 20 minutes and stored at −20 °C. The next day, the cell pellet was thawed on ice and resuspended in 15 ml of LYS buffer with one tablet of EASYpack protease inhibitor cocktail pre-dissolved in that buffer. Then, 10 µl of 25 mg/ml lysozyme from hen egg white, 5 µl of 10 mg/ml RNase A from bovine pancreas and 10 µl of 1 U/µl RNase-free DNase I were added, and the cell suspension was stirred at room temperature for 30 minutes followed by sonication on ice at 1,5 V (18 cycles, 10 seconds on and 20 seconds off) to enhance cell lysis. The cell lysate was then centrifuged at 35000x g for 30 minutes at 4 °C to remove cell debris.

3.3.4.4 Anion-exchange chromatography

Recombinant proteins were purified using NGC Quest 10 Plus System, and all chromatographic steps were performed at 20 °C.

To perform anion-exchange chromatography, two 5 ml HiTrap Capto Q columns were connected in series and first equilibrated with 10 column volumes of IonA buffer. After equilibration, the crude lysate was loaded onto the columns at 2 ml/min flow rate, and the column was washed with 10 column volumes of IonA buffer at the same flow rate. Recombinant proteins were eluted from the column with 0–50 % linear gradient of IonB buffer. The elution was performed at 3 ml/min rate with 20 column volumes of the buffer. The fractions were collected and analysed by SDS-PAGE for the presence of recombinant protein.

3.3.4.5 Protein concentration

The fractions after anion-exchange chromatography, containing recombinant protein, were combined and concentrated by centrifugation at 5000x g at 4 °C using Amicon Ultra 4 ml centrifugal unit, MWCO 10000. After each 10 minutes of centrifugation, protein solution was pipetted up and down several times to reduce the probability of protein aggregation. Centrifugation was proceeded until the final volume of protein sample was reduced to 5,0 ml.

3.3.4.6 Immobilized metal affinity chromatography

The concentrated protein sample after anion-exchange chromatography was loaded onto 5 ml HisTrap HP column at 2 ml/min flow rate. The column was pre-equilibrated before loading with 5 column volumes of NiA buffer. The column was then washed with

3 column volumes of 97 % NiA buffer and 3 % NiB buffer at 3 ml/min flow rate. The gradient elution of recombinant proteins was run with 10 column volumes of NiB buffer (0–50 % gradient) at 3 ml/min flow rate. The fractions were collected and analysed by SDS-PAGE for the presence of recombinant protein.

3.3.4.7 Preparative size-exclusion chromatography

The fractions after immobilized metal affinity chromatography, containing recombinant protein, were combined and concentrated to 0,5 ml as described in section 3.3.4.5. Concentrated protein sample was first centrifuged at 20000x g for 20 min at 4 °C to remove potential precipitation. After centrifugation, the protein sample was loaded onto HiLoad 16/600 Superdex 200 pg column at 2 ml/min flow rate. The column was pre-equilibrated with 2 column volumes of GF buffer. Recombinant proteins were eluted from the column by isocratic elution with 1 column volume of GF buffer at 1 ml/min flow rate. The fractions with purified proteins were collected, and the purity and integrity of the proteins were confirmed by SDS-PAGE. The protein samples were dialyzed against DIAL buffer overnight at 4 °C and concentrated up to 200 µl by centrifugation at 5000x g at 4 °C using Amicon Ultra 4 ml centrifugal unit, MWCO 10000. The purified proteins were aliquoted, flash frozen in liquid nitrogen and stored at −80 °C for further analysis.

3.3.5 Biophysical characterization of proteins

3.3.5.1 Matrix-assisted laser desorption/ionization

The identities and molecular weights of the purified recombinant proteins were confirmed by matrix-assisted laser desorption/ionization (MALDI) using UltrafleXtreme MALDI-TOF/TOF mass spectrometer with sinapinic acid as the matrix. The MALDI spectra of proteins were collected at the mass spectrometry core facility at IOCB according to the standard procedure. First, the matrix solution was prepared by dissolving 10 mg of sinapinic acid in 1 ml of 50 % (w/v) acetonitrile, 49,9 % (w/v) water and 0,1 % (w/v) trifluoroacetic acid. Next, MALDI target plate was spotted with 1 µl of the matrix solution, and the drop of matrix solution was allowed to air dry. Then, 1 µl of protein sample solution in 50 % acetonitrile was placed onto the target plate, and after drying the target plate was inserted into the instrument. The spectra were collected with a frequency of 1 kHz and averaged from several different locations on the target plate. The resulting MALDI spectrum was automatically processed in Flexanalysis software.

3.3.5.2 Analytical size-exclusion chromatography

The oligomerization states of the produced proteins were determined by analytical size-exclusion chromatography using Superdex 75 10/300 GL column. The column was calibrated using Gel filtration low molecular weight calibration kit. First, the individual proteins standards with defined molecular weight were dissolved in PBS buffer to the final concentration of 20 mg/ml, mixed together and diluted with PBS buffer to the final volume of 100 µl (Tab. 3.5).

Table 3.5. Protein standards mixture for gel filtration calibration

Protein standard	Concentration (mg/ml)	Volume (µl)	Final concentration (mg/ml)
Conalbumin (75 kDa)	20	7,5	1,5
Ovalbumin (44 kDa)	20	20	4,0
Carbonic anhydrase (29 kDa)	20	7,5	1,5
RNase A (13,7 kDa)	20	15,0	3,0
Aprotinin (6,5 kDa)	20	7,5	1,5

The protein standard mixture was loaded onto Superdex 75 10/300 GL column at 0,5 ml/min flow rate. The column was pre-equilibrated with 2 column volumes of PBS buffer. Protein standards were eluted from the column by isocratic elution with 1 column volume of PBS buffer at 0,5 ml/min flow rate. The elution volumes of protein standards were determined and plotted against logarithms of their molecular weights to construct the calibration curve.

Then, 40 µg of purified recombinant proteins were diluted with PBS buffer to the final volume of 100 µl, and the sample was subjected to the same procedure as protein standards. The molecular weights of the recombinant proteins were calculated from the calibration curve.

3.3.5.3 Multi-angle dynamic light scattering

The size distribution of the protein samples was estimated using multi-angle dynamic light scattering (MADLS) technique. The protein samples were diluted in PBS buffer to the final concentration of 0,1 mg/ml and centrifuged at 25000x g for 30 minutes at 4 °C. To completely remove dust particles, the samples were immediately filtered using 0,22 µm Ultrafree-MC centrifugal filters. The measurements were performed at 20 °C in a 3×3 mm quartz cuvette with internal volume of 40 µl using Zetasizer Ultra. The results were processed using the original Zetasizer 6.2 Malvern Instruments software.

3.3.5.4 Circular dichroism spectroscopy

Circular dichroism spectra of proteins were collected on Jasco J-1500 circular dichroism spectrophotometer in the 195–260 nm spectral range using 1mm cylindrical quartz cell. The protein samples were diluted with PBS buffer to the final concentration of 0,25 mg/ml. The experimental setup was as follows: 0,05 nm step resolution, 5 nm/min scanning speed, 16 seconds response time, 1 nm spectral band width and 2 accumulations. After baseline correction, the spectra were expressed as molar ellipticity per residue θ ($deg \cdot cm^2 \cdot dmol^{-1}$).

3.3.5.5 Limited proteolysis

The recombinant proteins were diluted with Lys-C buffer to the final concentration of 1 mg/ml, and then reaction mixtures for proteolytic digestion were prepared by mixing 7 µg of recombinant protein and 56 µl of Lys-C buffer. After incubation at 37 °C for 10 minutes reaction was started by adding 7 µl of 5 ng/µl Lys-C endoproteinase from *Lysobacter enzymogenes*. The whole reaction mixture was then divided into 10 µl aliquots that were incubated at 37 °C for different times: 0; 2; 5; 10; 20 and 40 minutes. After incubation the reaction was stopped by adding 3 µl of 5x SDS-PAGE sample buffer followed by heating at 95 °C for 10 minutes. All samples were then separated by SDS-PAGE, and after imidazol-zinc reverse staining the areas of the bands corresponding to the intact protein were estimated using ImageJ program.

3.3.6 Enzyme activity analysis

3.3.6.1 Estimation of enzyme activity

Activities of the recombinant proteins during purification were tested by a coupling assay using ADP Quest Assay kit. This commercial kit is based on monitoring of ADP formation. ADP, formed during the phosphotransferase reaction or ATP hydrolysis, is utilized by pyruvate kinase to convert phosphoenolpyruvate to pyruvate [114]. The latter is oxidized by pyruvate oxidase to produce hydrogen peroxide, that is used by peroxidase to convert 10-acetyl-3,7-dihydroxyphenoxazine, a fluorescent dye precursor, into fluorescent resorufin. Thus, the growth of resorufin fluorescence intensity measured in equal time intervals corresponds to accumulation of ADP during reaction, from which the rate of phosphotransferase reaction or ATP hydrolysis can be calculated. ADP Quest Assay kit includes 225 µM ADP standard, ADP detection reagent A, ADP detection reagent B and ADP assay buffer (15 mM HEPES, 20 mM NaCl, 1 mM EGTA, 0,02 %

(w/v) Tween-20, 10 mM $MgCl_2$ and 0,1 mg/ml bovine-γ-globulin). The compositions of ADP detection reagent A and ADP detection reagent B are not specified by the manifacturer.

Phosphotransferase activity of the recombinant proteins was tested according to the manufacturer's instructions. Protein samples were diluted with ADP assay buffer so that the enzymatic reaction proceeded with zero-order kinetics and the increase in fluorescence caused by the formation of the fluorescent resorufin product was linear. The diluted protein sample was mixed with 5 µl of 4 mM dCoA in 96-well black microplate, and the volume of the reaction mixture was adjusted to 30 µl with ADP assay buffer. After 20 µl of ADP detection reagent A and 40 µl of ADP detection reagent B were added, the plate was incubated at 37 ℃ for 10 minutes. The reaction was initiated by adding 10 µl of 2 mM ATP, and the fluorescence intensity was measured at 37 ℃ in kinetic mode with 2 minutes intervals using CLARIOstar Plus microplate reader at excitation/emission wavelengths of 530/590 nm. All fluorescence intensity measurements were performed in triplicates. Additionally, ATPase activity of the purified proteins was measured by the same procedure, but dCoA was not added to the reaction mixture.

To estimate activity of the recombinant proteins, ADP calibration curve was established by measuring fluorescence intensities of the reaction mixtures with different concentrations of ADP (0; 2,82; 5,63; 11,25 and 22,5 µM) that were prepared by serial dilution of 225 µM ADP standard provided by the kit. The reaction mixtures with different ADP concentrations were incubated for 30 minutes at 37 °C, then the fluorescence intensity was measured at 37 °C in end-point mode as described above. The measured fluorescence intensities were plotted against ADP concentrations to obtain ADP calibration curve. The slope of the calibration curve, which corresponded to fluorescence intensity gained from 1 µM ADP, was used to express recombinant proteins activities as the amount of nascent ADP per minute ($\mu mol_{ADP} \cdot min^{-1} \cdot L^{-1}$ and $\mu mol_{ADP} \cdot min^{-1} \cdot mg^{-1}$).

3.3.6.2 Determination of kinetic parameters

Kinetic parameters were estimated using ADP Quest Assay kit. Enzyme assays were carried out using 80 ng (33 nM) of wild type protein and 500 ng (214 nM) of aromatic amino acid-lacking mutants of DPCK. To determine kinetic parameters for both dCoA and ATP, a set of reaction mixtures with different concentrations of one substrate and single saturating concentration of the other substrate for each recombinant protein were

prepared. Kinetic parameters for dCoA were determined at 200 μM ATP and 0; 3,13; 6,25; 12,5; 18,75; 25; 37,5; 50; 75; 100; 150 and 200 μM dCoA, kinetic parameters for ATP were determined at 200 μM dCoA and 0; 3,13; 6,25; 12,5; 18,75; 25; 37,5; 50; 75; 100; 150 and 200 μM ATP. All reactions were performed at 37 °C in ADP assay buffer in black microplate with 100 μl total volume and initiated by adding ATP. The fluorescence intensity was measured according to the procedure described in section 3.3.6.1. All fluorescence intensity measurements were performed in triplicates. The amount of ADP formed per minute over the first ten minutes of the reaction was taken as the initial velocity. The kinetic parameters were calculated using the non-linear least-squares regression based on the Michaelis–Menten equation. ATP conversion did not exceed 10 %. The experiments were repeated three times, and kinetic values are presented as the means ± SE.

3.3.6.3 High performance liquid chromatography

Formation of coenzyme A in phosphotranspherase reactions was detected by HPLC/MS analysis. For this purpose, 100 μl of reaction mixtures were prepared by mixing 80 ng (33 nM) of wild type protein or 500 ng (214 nM) of aromatic amino acid-lacking mutants, 100 μM dCoA and 100 μM ATP in HPLC buffer. The reaction mixture was incubated at 37 ℃ for 1 hour, then, reaction was stopped by adding 100 μl of acetonitrile. Precipitated recombinant protein was separated by centrifugation at 20000x g at 4 ℃ for 20 minutes. The reaction samples were then analysed using the Dionex Ultimate 3000RS HPLC system equipped with TSQ Quantiva MS detector. The electrospray ionization source was used for ionization in a positive mode. The HPLC solvent system consisted of 10 mM ammonium carbonate, pH 9,3 (A) and 97% acetonitrile (B). After injection of 1 μl of the reaction sample in 50 % B, elution was performed using the gradient of 15 % A and 85 % B for 3,5 minutes followed by an increase to 75 % A and 25 % B over 11,5 minutes and its continuation for 5 minutes with the ZIC-HILIC HPLC column (5 μm, 150 mm x 2,1 mm), at a flow rate of 0,13 ml/min. HPLC/MS analysis was kindly performed by RNDr. Anna Březinová at the Biocev mass spectrometry core facility.

4 Results

4.1 Design of aromatic amino acid lacking mutants

Aromatic amino acid residues play a key role in driving folding of globular proteins which is essential for their catalytic activity [115]. Therefore, aromatic amino acid residues (phenylalanine, tyrosine and tryptophan) in DPCK from *Aquifex aeolicus* were substituted by non-aromatic amino acid residues according to two different strategies that imply the best preservation of protein structural stability.

According to the first strategy, aromatic amino acids were substituted by leucine (leucine mutant, leuDPCK). The choice of leucine was based on the consensus between the chronological order of the amino acid appearance in the genetic code and the ranking of amino acids according to their increasing propensity to promote ordered protein structure (Fig. 4.1) [51,116]. The latter was derived by comparing frequencies of amino acids occurrences in structured and disordered regions of extant characterized proteins [116]. Aromatic amino acid residues have the highest frequency of occurrence in structured regions and appear to contribute the most to the emergence of ordered protein structure. Apart from aromatic amino acids, leucine, methionine and valine are the most order promoting non-aromatic amino acids, and among them leucine is considered to be the earliest addition to the genetic code (Fig. 4.1).

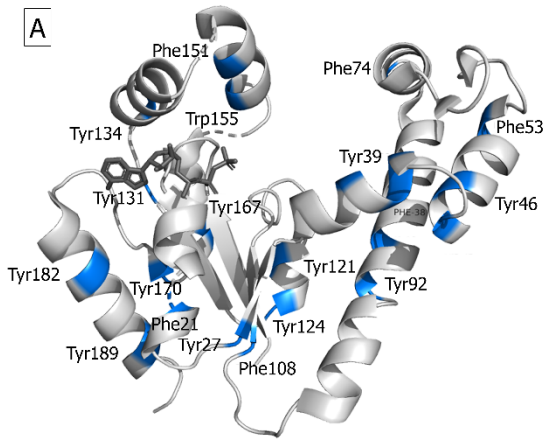
Figure 4.1. The correlation of canonical amino acids in various orders. (a) The 20 canonical amino acids are ranked according to the consensus chronological order of their incorporation into the genetic code [51]. (b) The 20 canonical amino acids are ranked according to their increasing propensity to promote ordered protein structure [116].

The second strategy was based on substitutions for aromatic amino acids residues evaluated by HotSpot Wizard 3.0 (hotspot wizard mutant, hswDPCK). HotSpot Wizard 3.0 is a bioinformatic tool utilized for the identification of hotspots in proteins to give improved protein stability, catalytic activity, substrate specificity and enantioselectivity [117]. The substitutions for aromatic amino acid residues were selected based on two criteria: occurrence at the corresponding positions in homologous proteins and probability of preservation of protein function. The non-aromatic amino acid residues

with the highest probability of protein function preservation and the highest occurrence at the corresponding positions in sequence homologs were considered to be the best substitutions.

The proposed substitutions of aromatic amino acid residues in DPCK from *Aquifex aeolicus* are shown in Fig. 4.2. The full sequences of wild type protein and two aromatic amino acid lacking mutants are presented in supplementary Fig. S1 (page 89).

A



B

Position	21	27	38	39	46	53	74
wtDPCK	Phe	Tyr	Phe	Tyr	Tyr	Phe	Phe
leuDPCK	Leu	Leu	Leu	Leu	Leu	Leu	Leu
hswDPCK	Leu	Pro	Val	Val	Leu	Val	Val

Position	92	108	121	124	131	134
wtDPCK	Tyr	Phe	Tyr	Tyr	Tyr	Tyr
leuDPCK	Leu	Leu	Leu	Leu	Leu	Leu
hswDPCK	Arg	Val	Asp	Val	Asp	Glu

Position	151	155	167	170	182	189
wtDPCK	Phe	Trp	Tyr	Tyr	Tyr	Tyr
leuDPCK	Leu	Leu	Leu	Leu	Leu	Leu
hswDPCK	Ala	Ile	Leu	Val	Arg	Leu

Figure 4.2. Three-dimensional X-ray structure of dephosphocoenzyme A kinase from *Aquifex aeolicus*. (A) Ribbon diagram of polypeptide fold with highlighted aromatic amino acid residues. (B) The proposed substitutions of aromatic amino acids in leucine mutant (leuDPCK) and hotspot wizard mutant (hswDPCK).

4.2 Genes and primers design

The genes encoding wild type protein and its aromatic amino acid-lacking mutants were synthesized commercially as GeneArt Strings by Thermo Fisher Scientific (USA). The genes were designed with ATG start codon at 5' end but without stop codon at 3' end, and their nucleotide sequences were optimized for expression in *Escherichia Coli* (supplementary Fig. S2, page 89).

Since all genes had identical 5' and 3' terminal nucleotide sequences, they were amplified using the same pair of gene-specific primers that contain restriction sites for *XhoI* (DPCK forward primer) and *NdeI* (DPCK reverse primer) to enable ligation of the synthetic genes into corresponding restriction sites of pET-24a (+) expression vector. Annealing sequences of forward and reverse primers were designed to have the same nearest neighbour melting temperature calculated using the Sequence Manipulation Suite online program.

4.3 PCR amplification and cloning

The temperatures of denaturation and elongation steps of PCR amplification by Pfu-X polymerase were set up according to the manufacturer's instructions. The annealing temperature (56 °C) was derived from the nearest neighbour melting temperature of primers annealing sequences. After PCR amplification the length and purity of products were estimated by horizontal agarose gel electrophoresis (Fig. 4.3).

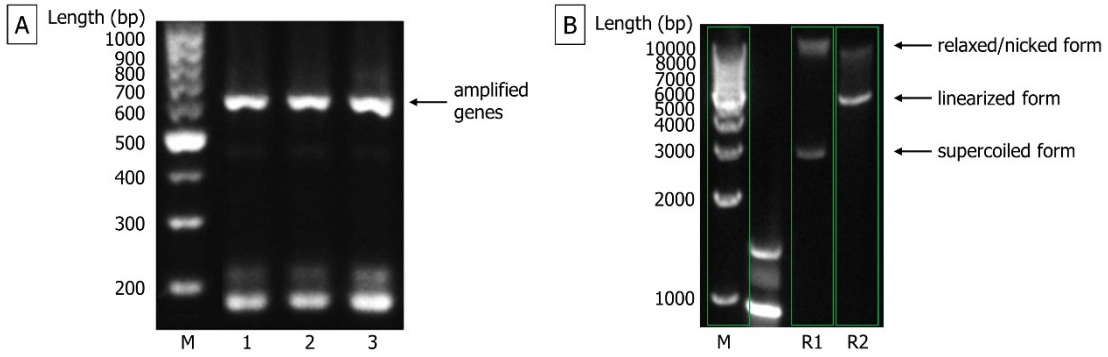


Figure 4.3. 1% agarose gels. (A) Gel after PCR amplification of *DPCK* genes: lane M – GelPilot 100 bp Plus Ladder (3 µl), lane 1 – reaction mixture after PCR amplification of *DPCK_WT* gene (1 µl), lane 2 – reaction mixture after PCR amplification of *DPCK_LEU* gene (1 µl), lane 3 – reaction mixture after PCR amplification of *DPCK_HSW* gene (1 µl). (B) Gel after restriction digestion of pET-24a (+) expression vector: lane M – GelPilot 1 kb Ladder (3 µl), lane R1 – pET-24a (+) expression vector before restriction digestion (100 ng), lane R2 – reaction mixture after restriction digestion of pET-24a (+) expression vector (2 µl).

Since aromatic amino acids substitutions did not change the total number of amino acids, and all amplified genes have the same length, PCR amplification was expected to yield the products of the uniform length of 630 bp. However, in the case of all genes, PCR amplification led to formation of several by-products. Effort was made to reduce the formation of by-products by optimization of PCR using different annealing temperatures (data not shown) but they did not lead to improvement, which might probably indicate impurities in the synthesized genes. The upper bands corresponding to the PCR products of the correct length (indicated by the arrow in Fig. 4.3) were extracted from the gel. After measuring DNA concentrations and purities, samples were used for subsequent cloning into pET-24a (+) expression vector.

The expression vector and amplified genes were cleaved using the same pair of restriction endonucleases, *NdeI* and *XhoI*. The results of pET-24a (+) expression vector restriction digestion are presented in Fig. 4.3. The band corresponding to the linearized plasmid was excised and extracted from the gel.

After restriction digestion the genes were inserted into the corresponding restriction sites of the pET-24a (+) expression vector using T4 DNA ligase. Before ligation, the plasmid DNA was preliminary dephosphorylated using Antarctic phosphatase to prevent self-ligation of the plasmid. The genes inserted into pET-24a (+) expression vector are expressed under the precise control of strong bacteriophage T7 promoter and must be initially cloned using a non-expression host that does not contain T7 RNA polymerase gene to avoid plasmid instability [118]. Thus, the ligation reaction mixture was used to transform OneShot TOP10 *Escherichia Coli* chemically competent cells for plasmid propagation. Successful transformants were selected using colony PCR. For each gene 2 cell colonies were picked from the LB agar plates and submitted to colony PCR to check for the presence of the insert. To minimize false positive results plasmid DNA released from cell colonies was amplified using insert-specific primer (DPCK forward primer), annealing to 5' end of the gene sequence, and backbone-specific reverse primer (T7 terminator primer), annealing to T7 terminator sequence, which flanks the 3' end of the gene sequence only if it was properly inserted in the plasmid backbone. After PCR amplification the length and purity of the products were estimated by horizontal agarose gel electrophoresis (Fig. 4.4).

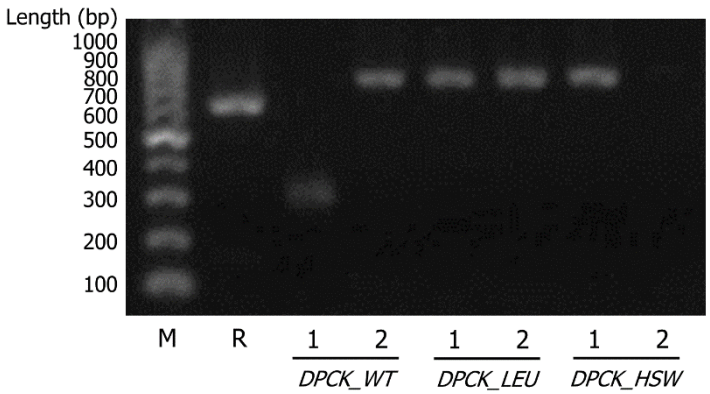


Figure 4.4. 1% agarose gel after colony PCR with OneShot TOP10 *Escherichia Coli* chemically competent cells transformed with pET-24a-DPCK gene ligation product: lane M – GelPilot 100 bp Plus Ladder (3 µl), lane R – *DPCK_WT* gene after PCR amplification (100 ng), lanes 1 and 2 – reaction mixtures after colony PCR of cell colonies №1 and №2, respectively (1 µl).

Colony PCR was expected to yield PCR products with the size of 700 bp, that can be observed in all cell colonies except colony №1 in the case of *DPCK_WT* gene and colony №2 in the case of *DPCK_HSW* gene. Positive clones, namely colony №2 for *DPCK_WT* gene and colonies №1 for *DPCK_LEU* and *DPCK_HSW* genes, were used to start overnight cell cultures. Plasmid DNA were then extracted from the cells and subjected to Sanger sequencing to confirm the correct sequence of the inserts.

4.4 Expression and solubility analysis

After the sequences of the inserts were confirmed by Sanger sequencing, plasmid DNA was used to transform BL21 (DE3) *Escherichia Coli* competent cells for recombinant protein production. Expression of recombinant proteins in BL21 (DE3) *Escherichia Coli* cell strain was first tested on a small scale. The results of expression and solubility analysis are presented in Fig. 4.5.

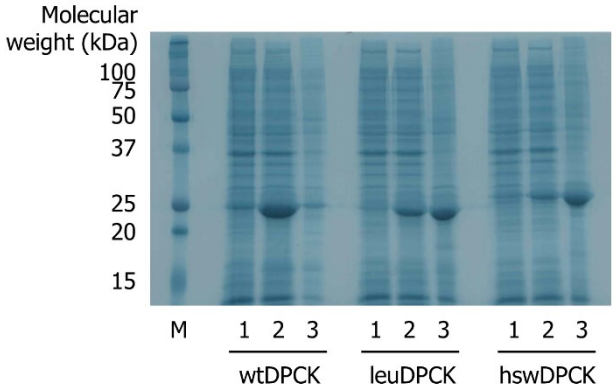


Figure 4.5. 14% SDS-polyacrylamide gels visualized by imidazole-zinc reverse staining of samples obtained after recombinant proteins expression: lane M – Precision Plus Protein Dual Xtra Prestained Protein Standards (3 µl), lane 1 – lysate of cell culture not induced by IPTG (10 µl), lanes 2 and 3 – soluble and insoluble fraction of lysate of cell culture induced by 0,5 mM IPTG for 4 hours at 37 °C, respectively (10 µl).

The recombinant proteins were expressed with a C-terminal fusion of 6xHis-tag. The expected molecular weights of the proteins lie in the narrow range of 23–24 kDa. Wild type protein demonstrates the highest level of expression entirely as a well-soluble protein, while both aromatic amino acid-lacking mutants have lower expression level and are mostly expressed as insoluble inclusion bodies. The established expression conditions (0,5 mM IPTG, 4 hours, 37 °C) were considered satisfactory to produce sufficient amounts of these recombinant proteins.

4.5 Purification of recombinant proteins

After establishing expression conditions, recombinant proteins were produced in a large volume of LB medium – 1 L for wild type protein due to its high expression level and high solubility and 2 L for both aromatic amino acid-lacking mutants. After harvesting cells and their lysis, C-terminal polyhistidine-tagged recombinant proteins were purified by three-step chromatographic purification procedure using sequentially anion-exchange chromatography, immobilized metal affinity chromatography and size-exclusion chromatography.

First, the recombinant proteins were purified by anion-exchange chromatography using HiTrap Capto Q column. The progress of purification was monitored by absorbance at 280 nm for the wild type protein. Since both mutants are deficient in aromatic amino acid residues that can absorb at 280 nm, the progress of their purification was monitored by absorbance at 215 nm (absorption wavelength for peptide bonds). The results of purification by anion-exchange chromatography were monitored by SDS-PAGE analysis of individual collected fractions of 1,0 ml (Fig. 4.6, page 56).

After anion-exchange chromatography fractions 110–126 for wtDPCK, 114–126 for leuDPCK and 108–123 for hswDPCK were concentrated to 5,0 ml and further purified by immobilized metal affinity chromatography using HisTrap HP column. The progress of purification was monitored by absorbance at 280 nm for the wild type protein and 215 nm for the aromatic amino acid-lacking mutants. During purification fractions of 1,0 ml from the sample injection, column wash and elution steps were collected and analysed by SDS-PAGE. (Fig. 4.7, page 57).

After immobilized metal affinity chromatography fractions 58–62 for wtDPCK, 51–56 for leuDPCK and 53–59 for hswDPCK were concentrated to 0,5 ml and further purified by size-exclusion chromatography using. The progress of purification was monitored by absorbance at 280 nm for the wild type protein and 215 nm for the aromatic amino acid-lacking mutants. The fractions of 1 ml corresponding to the elution peak were collected and analysed by SDS-PAGE (Fig. 4.8, page 58).

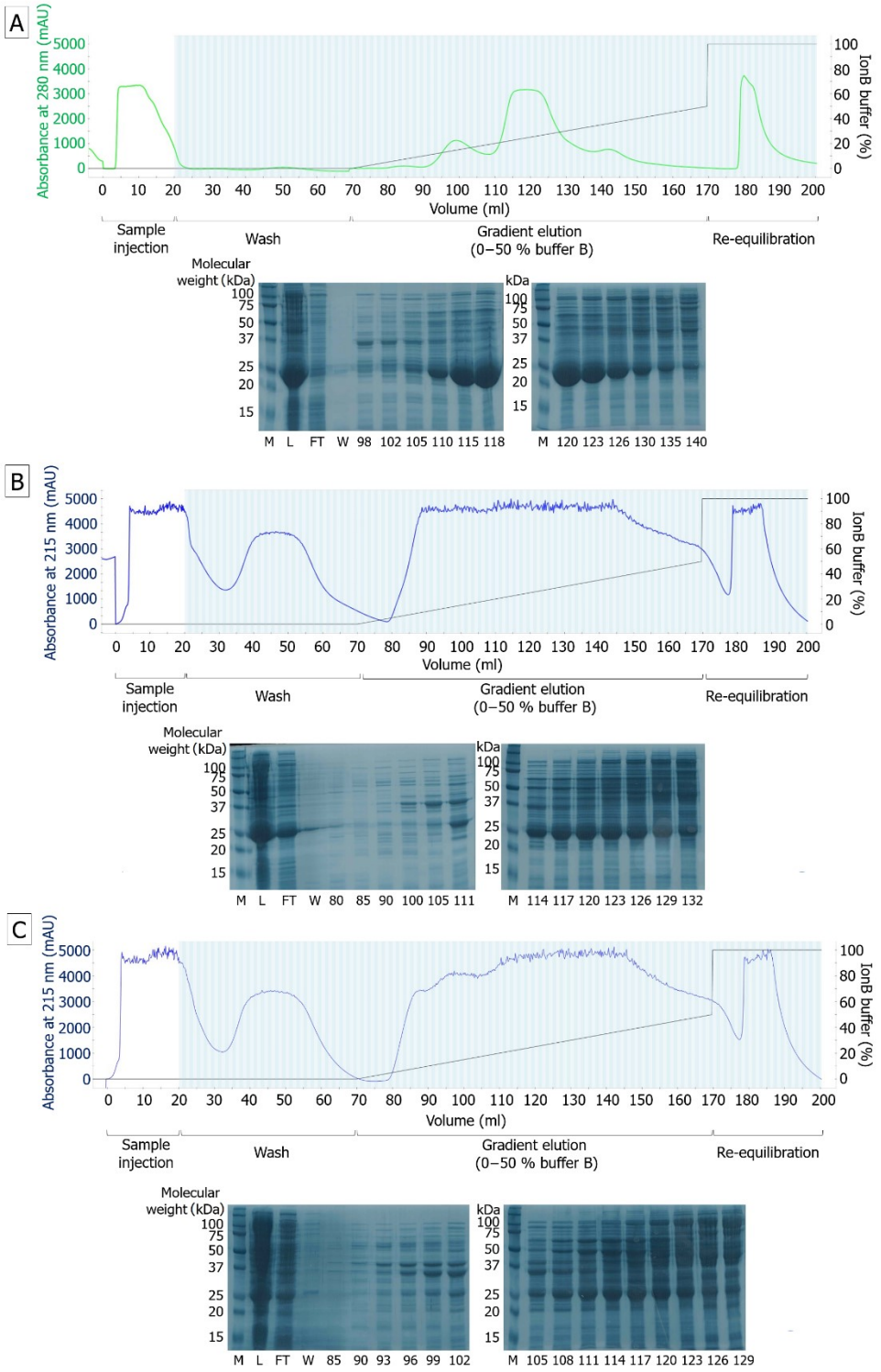


Figure 4.6. The results of recombinant proteins purification by anion-exchange chromatography. (A) wtDPCK, (B) leuDPCK and (C) hswDPCK. Chromatograms: green line – absorbance at 280 nm, blue line – absorbance at 215 nm, black line – percentage of IonB buffer. 14% SDS-polyacrylamide gels of individual fractions: lane M – Precision Plus Protein Dual Xtra Prestained Protein Standards (3 µl), lane L – sample before purification (3 µl), lane FT – combined fractions of the sample injection step (10 µl), lane W – combined fractions of the wash step (10 µl), lanes with numbers – individual fractions collected during the gradient elution step (10 µl). Purification was performed using HiTrap Capto Q column (IonA buffer: 20 mM Tris, 20 mM NaCl, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 8,0; IonB buffer: 20 mM Tris, 1 M NaCl, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 8,0).

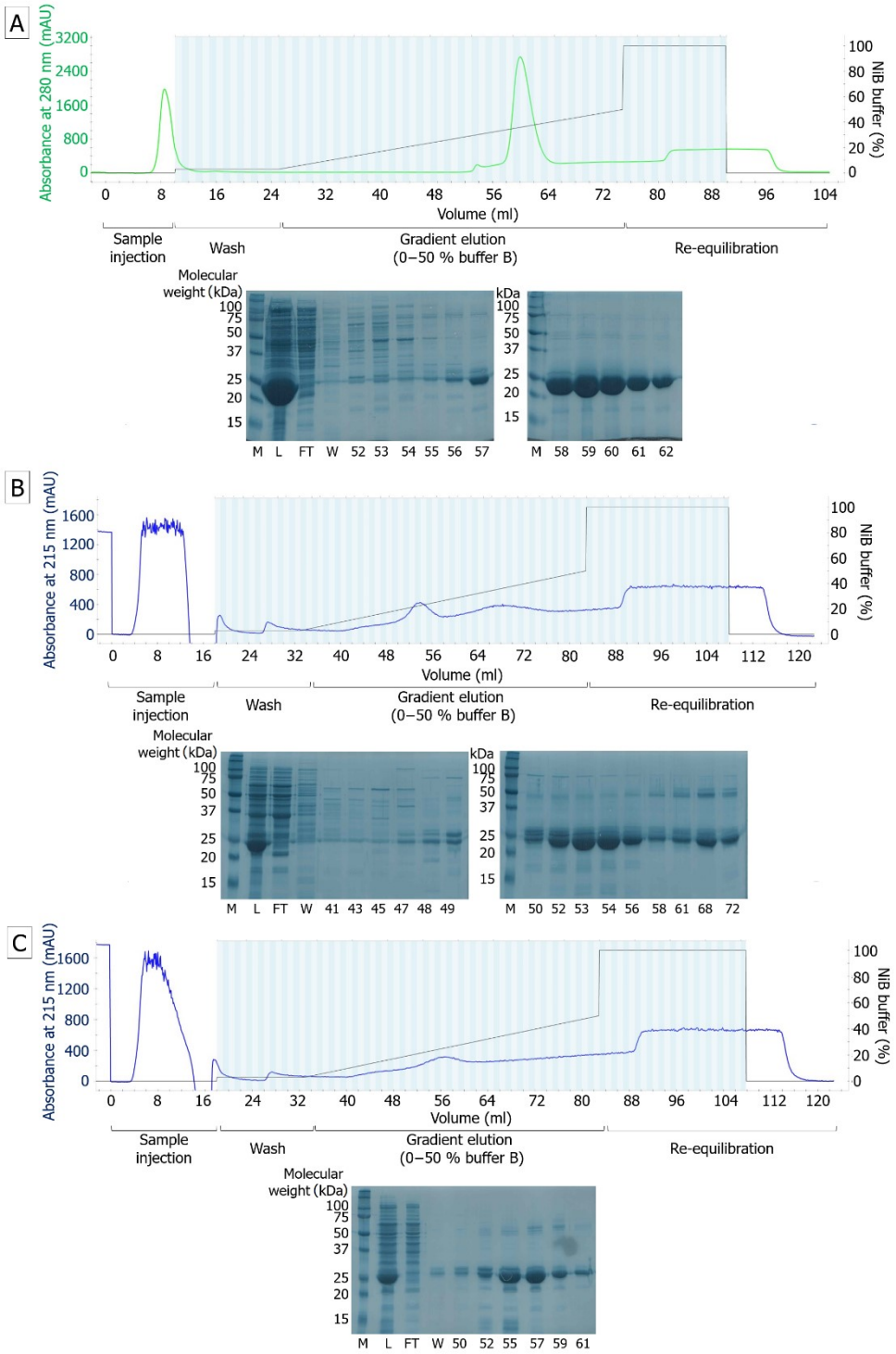


Figure 4.7. The results of recombinant proteins purification by immobilized metal affinity chromatography. (A) wtDPCK, (B) leuDPCK and (C) hswDPCK. Chromatograms: green line – absorbance at 280 nm, blue line – absorbance at 215 nm, black line – percentage of NiB buffer. 14% SDS-polyacrylamide gels of individual fractions: lane M – Precision Plus Protein Dual Xtra Prestained Protein Standards (3 µl), lane L – protein sample before purification (3 µl), lane FT – combined fractions of the sample injection step (10 µl), lane W – combined fractions of the wash step (10 µl), lanes with numbers – individual fractions collected during the gradient elution step (10 µl). Purification was performed using HisTrap HP column (NiA buffer: 20 mM Tris, 500 mM NaCl, 10 mM imidazole, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 7,6; NiB buffer: 20 mM Tris, 500 mM NaCl, 500 mM imidazole, 1 mM β-mercaptoethanol, 0,1 % (w/v) Tween-20, pH 7,6).

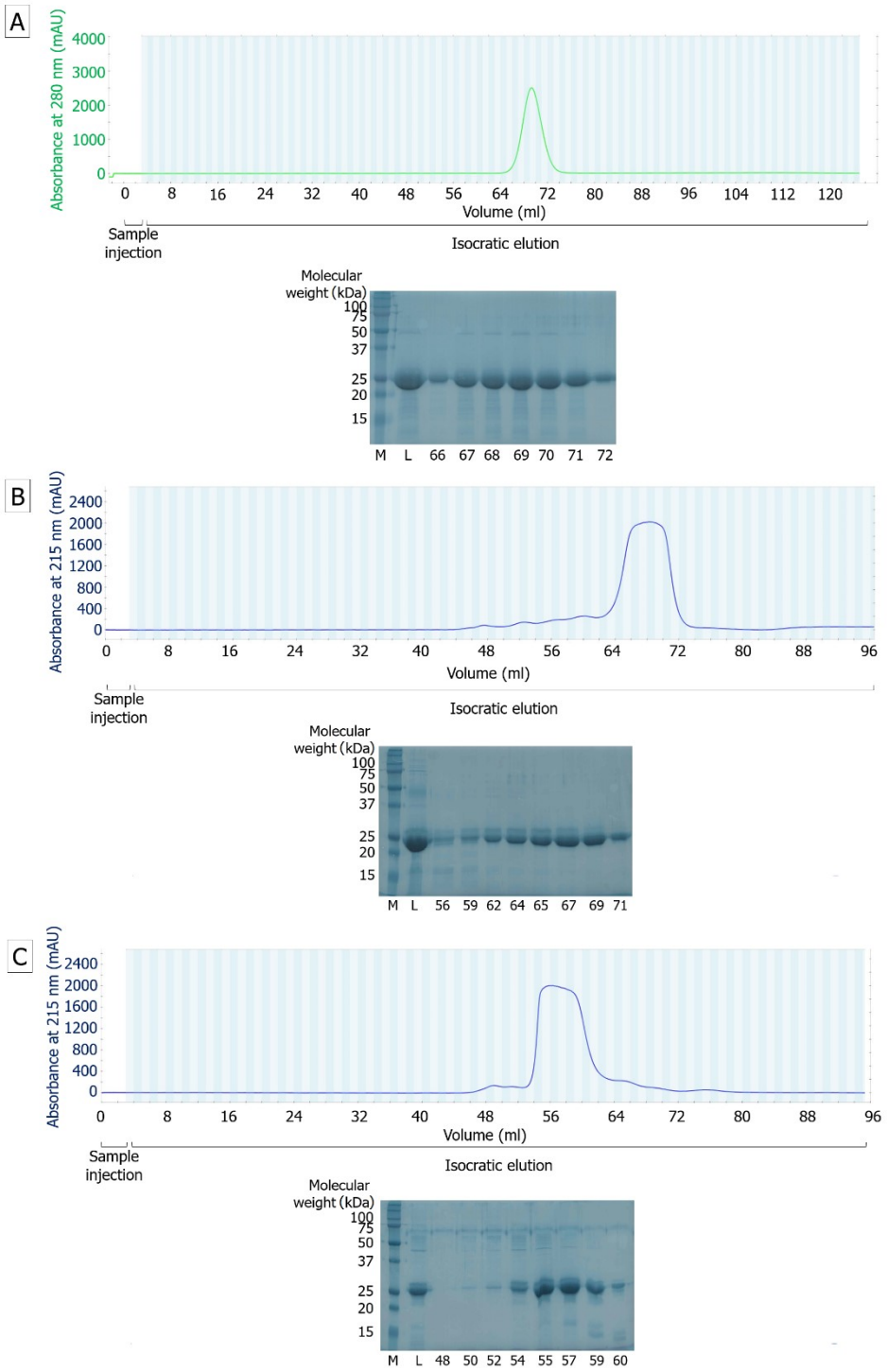


Figure 4.8. The results of recombinant proteins purification by size-exclusion chromatography. (A) wtDPCK, (B) leuDPCK and (C) hswDPCK. Chromatograms: green line – absorbance at 280 nm, blue line – absorbance at 215 nm. 14% SDS-polyacrylamide gels of individual fractions: lane M – Precision Plus Protein Dual Xtra Prestained Protein Standards (3 µl), lane L – protein sample before purification (1 µl for wtDPCK and 3 µl for leuDPCK and hswDPCK), lanes with numbers – individual fractions collected during the isocratic elution step (3 µl for wtDPCK and 10 µl for leuDPCK and hswDPCK). Purification was performed using HiLoad 16/600 Superdex 200 pg column (GF buffer: 50 mM Tris, 500 mM NaCl, 20 mM KCl, 10 mM $MgCl_2$, 0,5 mM DTT, pH 7,6).

The progress of the protein purification was monitored also by SDS-PAGE of samples after each chromatographic step (Fig. 4.9). All proteins were purified to one single band on 14% SDS-polyacrylamide gel. However, the formation of a slight upper band can be observed for aromatic amino acid-lacking mutants, which could indicate partial oxidation of cysteine residues in the mutants or some structural heterogeneity as sometimes observed for intrinsically disordered proteins (see the Discussion section for more detail). After size-exclusion chromatography, fractions 67–72 for wtDPCK, 63–71 for leuDPCK and 54–60 for hswDPCK were dialyzed overnight against a buffer containing strong reducing agent (TCEP). After dialysis they were concentrated to 10 mg/ml, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C.

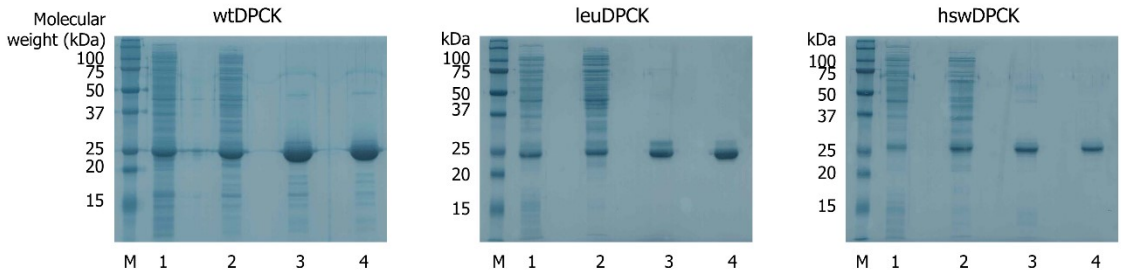


Figure 4.9. 14% SDS-polyacrylamide gels summarizing the protein purification course. lane M – Precision Plus Protein Dual Xtra Prestained Protein Standards (3 µl), lane 1 – protein sample after cell lysis, lane 2 – protein sample after anion-exchange chromatography, lane 3 – protein sample after immobilized metal affinity chromatography, lane 4 – protein sample after size-exclusion chromatography. The amount of protein in each lane is 1 µg as estimated by Bradford protein assay using Bio-Rad Protein Assay kit II.

To monitor the progress of purification, amount of protein and phosphotransferase activity of protein samples after each purification step were determined. Amount of protein was determined by Bradford protein assay using Bio-Rad Protein Assay kit II. The protein concentration in the samples after final purification step was additionally estimated by amino acid analysis (Tab. 4.1).

Table 4.1. Protein concentration values obtained for the samples after size-exclusion chromatography using Bradford protein assay and amino acid analysis

Protein	wtDPCK	leuDPCK	hswDPCK
Protein concentration estimated by Bradford protein assay (mg/ml)	13,8	10,2	9,8
Protein concentration estimated by amino acid analysis (mg/ml)	15,2	4,6	2,9

Comparison of protein concentration values obtained by Bradford protein assay and amino acid analysis revealed that Bradford protein assay gives inadequately high protein concentrations in the case of aromatic amino acid-lacking mutants and is not

reliable for determination of their amount (see the Discussion section for more detail). The progress of purification of the wild type protein is shown in Tab. 4.2.

Table 4.2. Purification table for recombinant dephosphocoenzyme A kinase from *Aquifex aeolicus*

Purification step	Total protein (mg)	Total activity ($\mu mol_{ADP} \cdot min^{-1}$)	Specific activity ($\mu mol_{ADP} \cdot min^{-1} \cdot mg^{-1}$)	Purification (fold)	Yield (%)
Crude lysate	182,44	15,77	0,086	1	100
Anion-exchange chromatography	36,96	12,95	0,35	4	82,1
Immobilized metal affinity chromatography	7,54	11,03	1,48	17,2	69,9
Size-exclusion chromatography	6,08	9,42	1,55	18	59,7

The yield of the wild type protein is 6,08 mg per 1 L *Escherichia coli* culture. The yields of aromatic amino acid-lacking mutants are significantly lower – 0,92 mg per 1 L *Escherichia coli* culture for leuDPCK and 0,81 mg per 1 L *Escherichia coli* culture for hswDPCK.

4.6 Quality assessment of purified proteins

Prior to characterization of enzymatic activity and secondary and tertiary structure, quality of recombinantly expressed and purified proteins was estimated by various techniques including MALDI, analytical size-exclusion chromatography and dynamic light scattering. First, the identities and molecular weights of the purified proteins were confirmed by MALDI (Fig. 4.10, page 61).

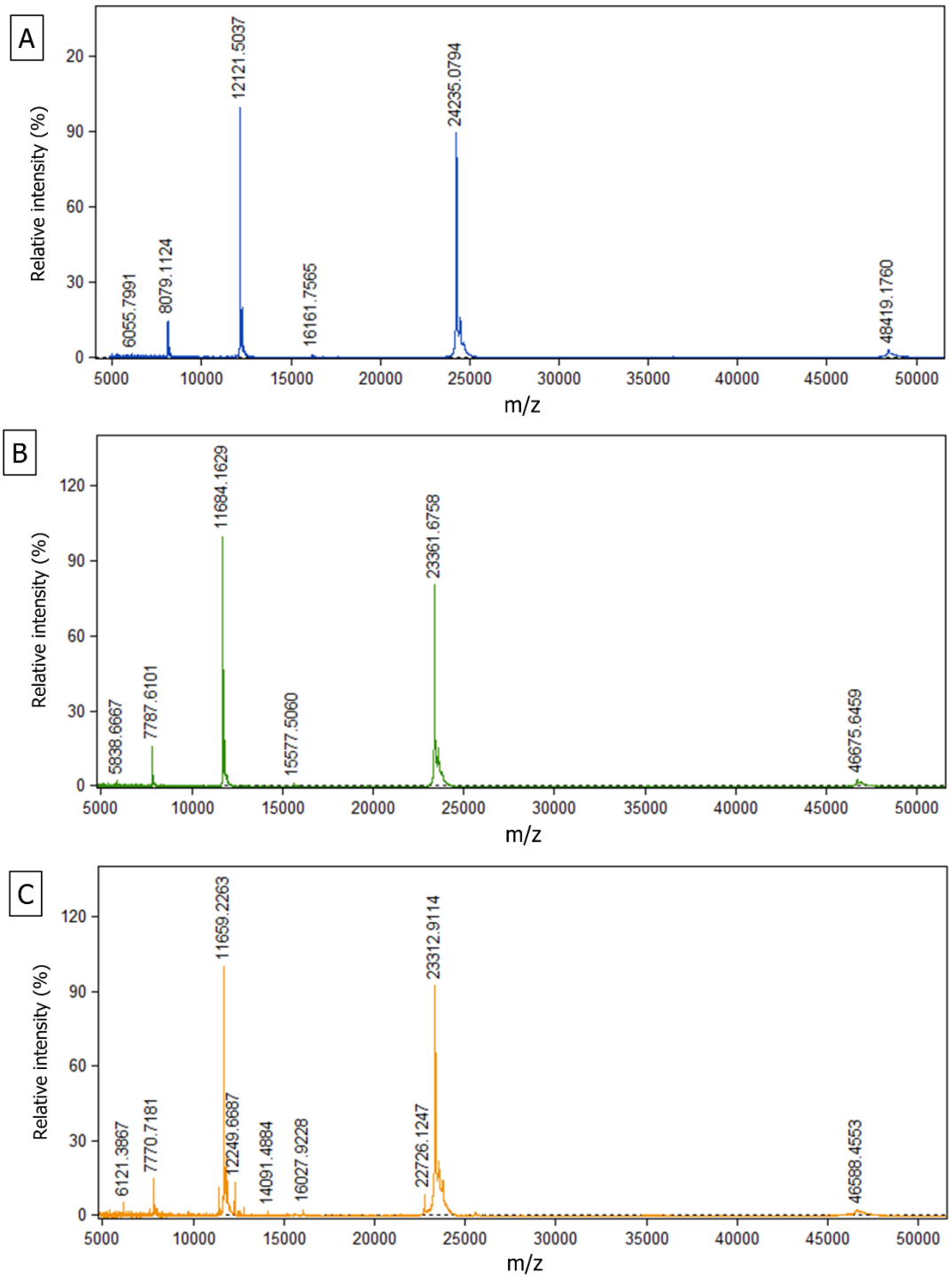


Figure 4.10. The MALDI-TOF mass spectra of recombinant proteins. (A) wtDPCK, (B) leuDPCK and (C) hswDPCK. The spectra were obtained using sinapinic acid as the matrix. Expected molecular weights are 24,24 kDa for wtDPCK, 23,36 kDa for leuDPCK and 23,31 kDa for hswDPCK.

The integrity and size homogeneity of the purified proteins were then evaluated by analytical size-exclusion chromatography using Superdex 75 Increase 10/300 GL column. The column was first calibrated using Gel filtration low molecular weight calibration kit with range of 6,5 to 75 kDa. The elution volumes of individual protein

standards in the range were used to plot the calibration curve for estimation of molecular weight of the recombinant proteins. The results of Superdex 75 Increase 10/300 GL column calibration are shown in Fig. 4.11.

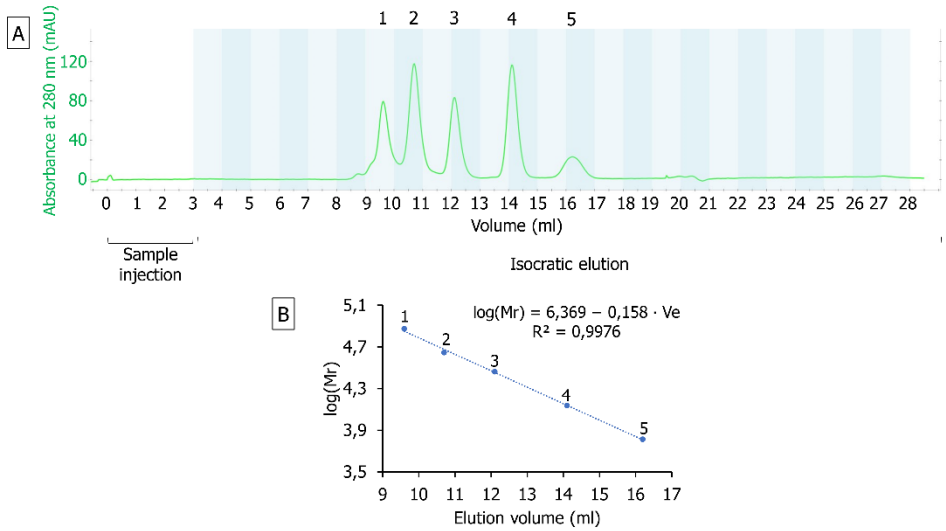


Figure 4.11. Superdex 75 Increase 10/300 GL column calibration. (A) chromatogram: green line – absorbance at 280 nm; (B) calibration curve: 1 – conalbumin (75 kDa), 2 – ovalbumin (44 kDa), 3 – carbonic anhydrase (29 kDa), 4 – RNase A (13,7 kDa) and 5 – aprotinin (6,5 kDa). Calibration of the column was performed in PBS buffer at a flow rate of 1 ml/min.

After calibration, 40 µg of each recombinant protein was loaded onto the column. The elution profiles of the proteins are shown in Fig. 4.12.

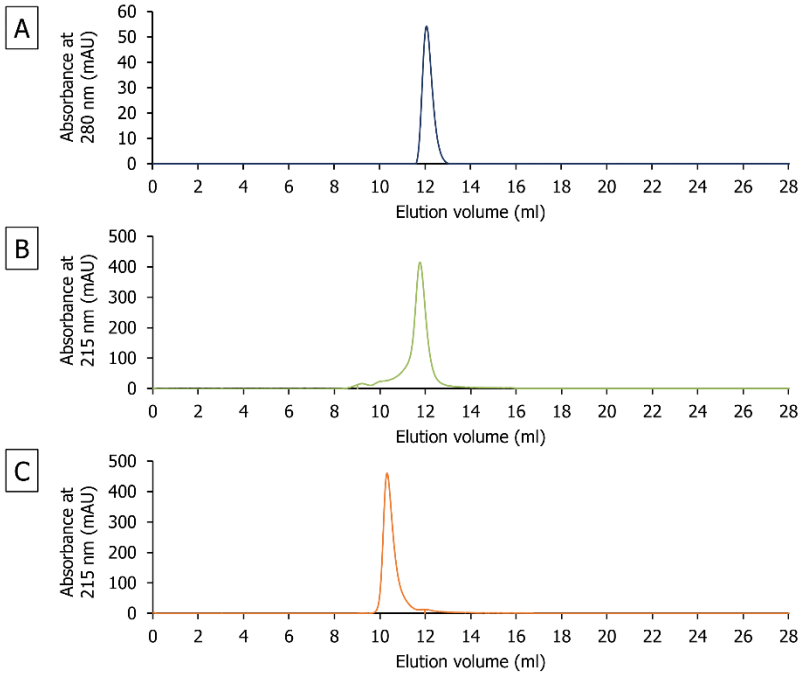


Figure 4.12. Analysis of recombinant proteins by size-exclusion chromatography. Protein elution was monitored by absorbance at 280 nm for wtDPCK (A) and 215 nm for leuDCPCK (B) and hswDPCK (C). Chromatography was performed using Superdex 75 Increase 10/300 GL column with PBS buffer at a flow rate of 1 ml/min.

All proteins were eluted as single, almost symmetrical peaks at 12,1; 11,8 and 10,2 ml for wtDPCK, leuDPCK and hswDPCK, respectively. Despite the similarities of their molecular weights, detectable difference in elution volume for hswDPCK is observed. Estimated molecular weights for wtDPCK, leuDPCK and hswDPCK are 29, 33 and 55 kDa, respectively, if globular structure of the proteins is expected. Regarding the oligomeric state, the purified proteins can be considered solely monomeric based on the good agreement between expected molecular weights and molecular weights estimated by analytical size-exclusion chromatography in the case of wtDPCK and leuDPCK. However, hswDPCK is characterized by approximately 2 times higher estimated molecular weight in comparison with wtDPCK.

To detect possible aggregation, protein samples in PBS buffer were characterized by MADLS technique, where size distribution and dispersity were analyzed (Fig. 4.13). Each protein sample was measured at 20 °C and with an equilibration time of 180 seconds to default temperature in between measurements. The data was collected and processed using Zetasizer 6.2 Malvern Instruments software, and for data processing the default General Purpose analysis model was used.

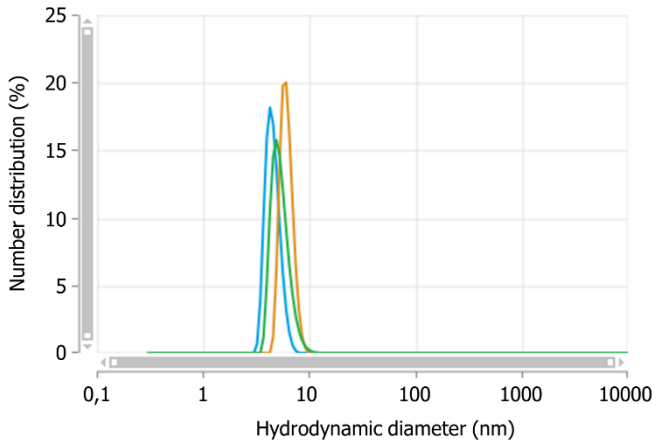


Figure 4.13. Multi-angle dynamic light scattering analysis of recombinant proteins. Representation of particle size distribution by number in function of particle size of recombinant proteins: blue line – wtDPCK, green line – leuDPCK and orange line – hswDPCK. The particle size distribution curves were obtained for 0,1 mg/ml protein samples in PBS buffer.

All proteins demonstrate high monodispersity, indicating no aggregation, with the mean hydrodynamic diameter of 4,2, 4,9 and 6,1 nm for wtDPCK, leuDPCK and hswDPCK, respectively.

4.7 Enzyme activity characterization

Activities of wild type protein and aromatic amino acid-lacking mutants were tested by a coupling assay using ADP Quest Assay kit. Dephosphocoenzyme A kinases are enzymes that catalyse the transfer of a phosphate group from ATP to dephosphocoenzyme A. They consist of three domains: the ATP-binding domain, the CoA domain and the lid domain which prevents non-specific transfer of phosphate group from ATP to water. Since aromatic amino acids are spread all over the sequence of DPCK from *Aquifex aeolicus* but do not constitute ATP binding motif, the P-loop, substitution of aromatic amino acids in wild type enzyme was expected to lead to drastic changes in the conformation of all domains including the CoA and the lid domains, which might result in the loss of specifity to dCoA and a different mechanism for enzymatic reaction. Thus, recombinant proteins were tested for both phosphotransferase (transfer of phosphate from ATP to dCoA) and ATPase (hydrolysis of ATP, transfer of phosphate from ATP to water) activities.

80 ng (32 nM) of wild type enzyme and 500 ng (214 nM) of aromatic amino acid-lacking mutants were used for preliminary characterization of their activities. Phosphotransferase activity was tested with 200 µM dCoA and 200 µM ATP, ATPase activity was tested only with 200 µM ATP. Since the fluorescence background increases with time, the reaction mixture consisting of 200 µM dCoA and 200 µM ATP but without protein was used as a negative control. All fluorescence intensity measurements were performed in triplicates. The results are presented in Fig. 4.14. To estimate the activities of proteins, ADP calibration curve was plotted by measuring fluorescence intensities of the reaction mixtures with different concentrations of ADP. The slope of the ADP calibration curve corresponds to the fluorescence intensity gained from 1 µM ADP. The slope of the negative control was subtracted from the ones corresponding to ATPase or phosphotransferase activities of each protein, and the resulting numbers were divided by the slope from ADP calibration curve to estimate the amount of ADP formed during reaction in one minute.

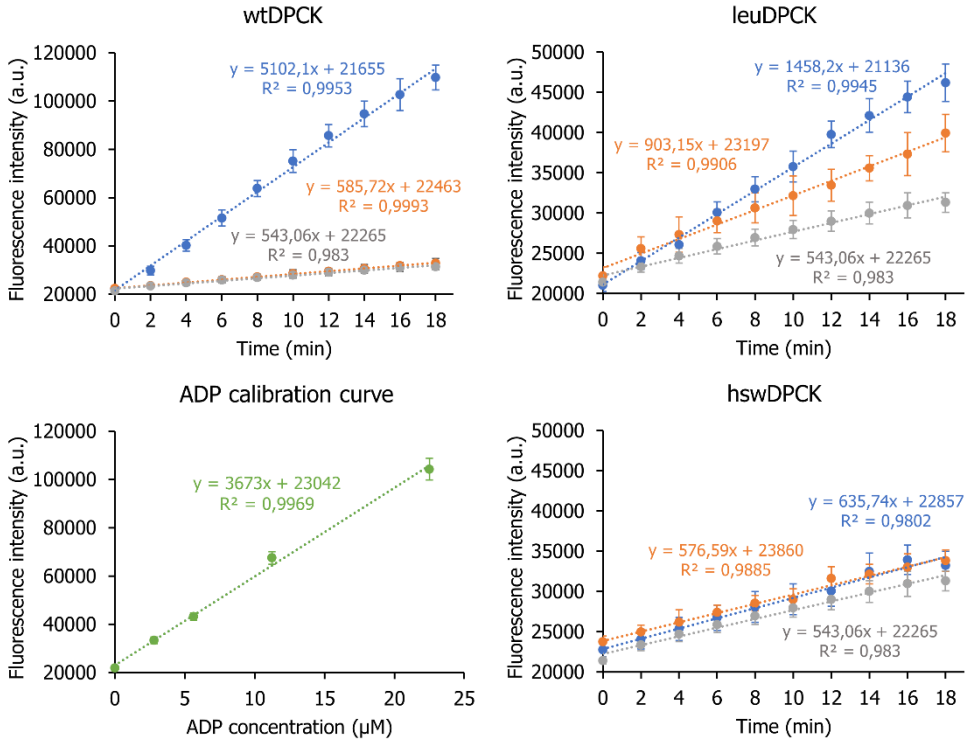


Figure 4.14. Estimation of phosphotransferase and ATPase activities of recombinant proteins. Fluorescence intensity was measured in different reaction mixtures: gray line – 200 μM dCoA + 200 μM ATP (negative control), orange line – protein + 200 μM ATP (ATPase activity test), blue line – protein + 200 μM dCoA + 200 μM ATP (phosphotransferase activity test). All measurements were performed at 37 °C in kinetic mode with 2 minutes intervals. ADP calibration curve (green line) was plotted by measuring fluorescence intensities from serially dilluted 225 μM ADP standard.

As can be seen, wtDPCK demonstrates solely phosphotransferase activity (1,24±0,058 $\mu mol_{ADP} \cdot min^{-1} \cdot L^{-1}$ or 1,55±0,073 $\mu mol_{ADP} \cdot min^{-1} \cdot mg^{-1}$), while leuDPCK gains the ability to hydrolyse ATP in the absence of the second substrate, dCoA, and demonstrates both phosphotransferase (0,25±0,013 $\mu mol_{ADP} \cdot min^{-1} \cdot L^{-1}$ or 0,05±0,003 $\mu mol_{ADP} \cdot min^{-1} \cdot mg^{-1}$) and ATPase (0,10±0,006 $\mu mol_{ADP} \cdot min^{-1} \cdot L^{-1}$ or 0,02±0,001 $\mu mol_{ADP} \cdot min^{-1} \cdot mg^{-1}$) activities. Neither phosphotransferase nor ATPase activities were detected for hswDPCK that was excluded from further enzymatic characterization.

Since the conversion of the substrates did not exceed 10 % and the fluorescence intensity grew linearly over the first 18 minutes of the reactions, the same amounts of wtDPCK and leuDPCK were used for determination of kinetic parameters. The amount of ADP formed per minute over the first ten minutes of the reaction was taken as the initial velocity. The initial velocities of the reactions were measured at different concentrations of one substrate and single saturating concentration of the other substrate. All fluorescence intensity measurements were performed in triplicates. The obtained values were then plotted as a function of the substrate concentration, and the probable

substrate inhibition and validity of obtained data were verified by constructing Hanes–Woolf plots (Fig. 4.15–16).

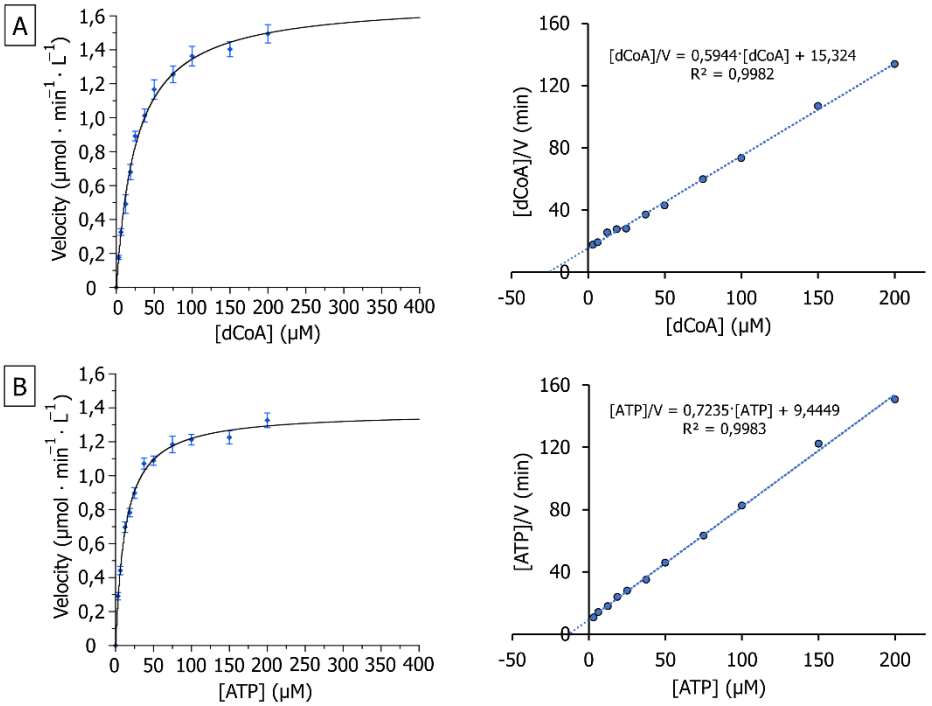


Figure 4.15. The kinetic curves and Hanes–Woolf plots of wtDPCK plotted for (A) 200 µM ATP and 0–200 µM dCoA and (B) 200 µM dCoA and 0–200 µM ATP.

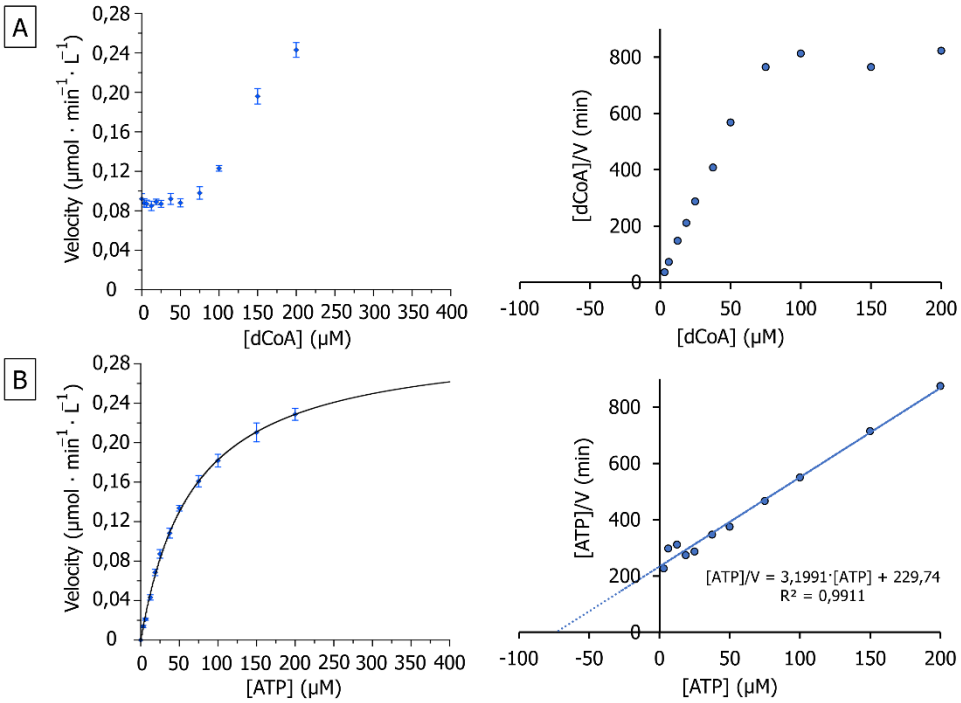


Figure 4.16. The kinetic curves and Hanes–Woolf plots of leuDPCK plotted for (A) 200 µM ATP and 0–200 µM dCoA and (B) 200 µM dCoA and 0–200 µM ATP.

As can be seen, wtDPCK obeys Michaelis-Menten kinetics with the high linearities of Hanes–Woolf plots for both substrates. LeuDPCK follows the Michaelis-Menten kinetics only for one of the substrates, ATP, however, the full saturation of leuDPCK with ATP cannot be reached due to the limited ATP tolerance of ADP Quest Assay Kit. The Michaelis constant and maximum velocity values were established by fitting the kinetic curves to Michaelis−Menten equation (Tab. 4.3).

Table 4.3. Kinetic parameters of wtDPCK and leuDPCK estimated for both substrates

Protein	Substrate	V_{max} (μmol · min^{-1} · L^{-1})	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} · M^{-1})
wtDPCK	dCoA	1,57 ± 0,17	24,3 ± 1,7	0,817 ± 0,088	33620 ± 3620
	ATP	1,41 ± 0,05	12,7 ± 0,3	0,730 ± 0,026	57480 ± 2050
leuDPCK	dCoA	–	–	–	–
	ATP	0,30 ± 0,01	65,9 ± 4,5	0,0234 ± 0,0008	355 ± 12

Since, leuDPCK is capable of ATP hydrolysis in the absence of the second substrate, dCoA, the kinetic curve of leuDPCK for different dCoA concentrations has the exponential shape, and the enzyme does not follow the Michaelis-Menten kinetics which is reflected in non-linearity of the Hanes–Woolf plot.

The formation of the second product, coenzyme A, in the phosphotransferase reactions of both wtDPCK and leuDPCK was verified by HPLC/MS analysis. For that purpose, the same amounts of proteins that were used for determination of kinetic parameters, were incubated with 100 μM ATP and 100 μM dCoA for 1 hour at 37 °C in HPLC buffer. The reaction mixture consisted only of 100 μM ATP and 100 μM dCoA was used as a blank control. The reactions were stopped by adding acetonitrile, and after removing the precipitated protein by centrifugation the reaction mixtures were analysed using Dionex Ultimate 3000RS HPLC system equipped with TSQ Quantiva MS detector. The results of HPLC/MC analysis are presented in Fig. 4.17 (page 68).

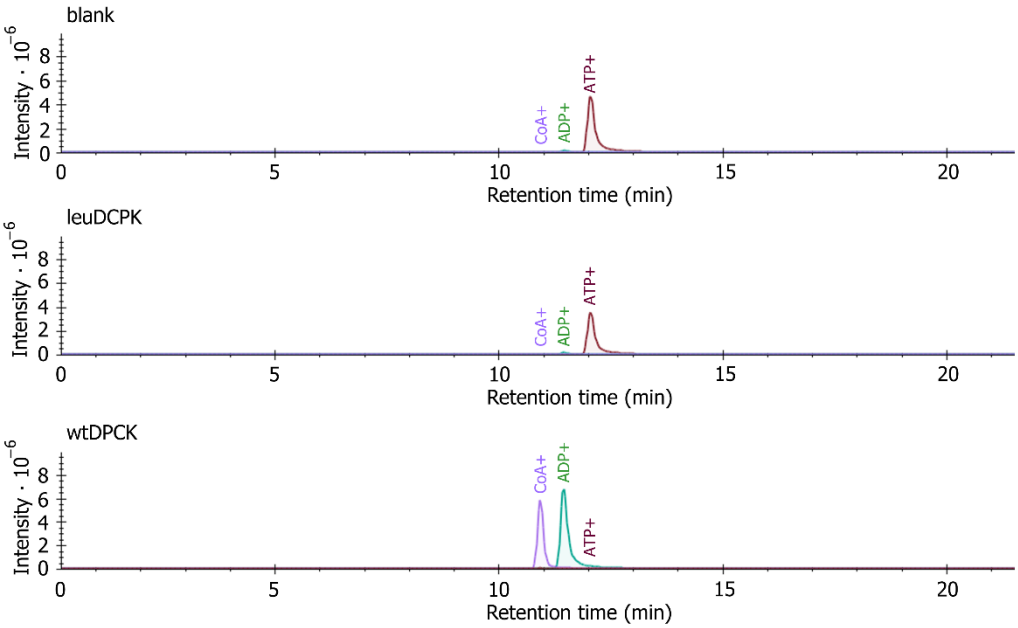


Figure 4.17. Detection of coenzyme A formation by HPLC/MS analysis. HPLC chromatograms of the reaction mixtures consisting of the protein (33 nM for wtDPCK and 214 nM for leuDPCK), 100 µM ATP and 100 µM dCoA. The reaction mixture consisting of only 100 µM ATP and 100 µM dCoA was used as a blank control. Peaks corresponding to dCoA are not shown in chromatograms due to their relatively high intensity.

All substances present in the reaction mixture (both substrates, dCoA and ATP, and both products, coenzyme A and ADP) were effectively separated from each other. Coenzyme A formed during the phosphotransferase reactions was eluted as a single peak at 10,9 minutes. Calculated areas of the peaks corresponding to coenzyme A are shown in the Tab. 4.4. As can be seen, approximately 100 times lower amount of coenzyme A was formed during phosphotransferase reaction catalysed by leuDPCK in comparison with wtDPCK.

Table 4.4. Estimated peak areas for coenzyme A formed in the phosphotransferase reactions of leuDPCK and wtDPCK

Sample	blank	leuDPCK	wtDPCK
Peak area for coenzyme A	$1{,}22 \cdot 10^4$	$4{,}06 \cdot 10^5$	$3{,}47 \cdot 10^7$

4.8 Structural characterization

The secondary structure of the recombinant proteins was characterized by CD spectroscopy. The far-UV CD spectra were collected for 0,25 mg/ml protein samples in PBS buffer at 20 °C in a cell of 1 mm pathlength. The spectra were truncated below 195 nm because of excessive noise at these wavelengths (Fig. 4.18, page 69).

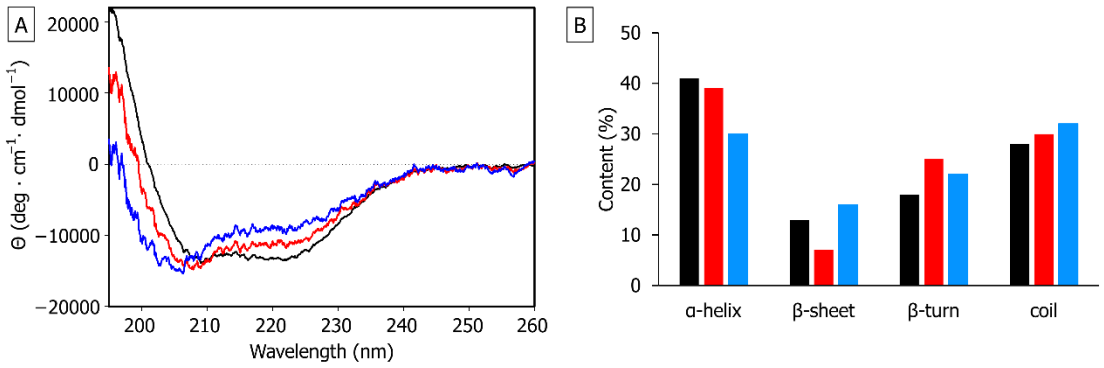


Figure 4.18. Characterization of secondary structure by CD spectroscopy. (A) Far-UV CD spectra collected for 0,25 mg/ml protein samples in PBS buffer: wtDPCK (black), leuDPCK (red) and hswDPCK (blue); (B) The secondary structure content estimated by numerical data analysis of CD spectra using CDNN software.

The CD spectrum of wild type protein has a pronounced minima at 208 and 222 nm thus indicating the predominantly α-helical character of the protein. Substitution of aromatic amino acid residues does not seem to have resulted in drastic changes in the secondary structure content of the protein as indicated by the high resemblance of CD spectra of wild type protein and its aromatic amino acid-lacking mutants. LeuDPCK content of α-helices is relatively conserved while there is a decrease in the β-sheet content and an increased content of β-turns. HswDPCK exhibits a significant decrease in the content of α-helices and increased contents of both β-sheets and β-turns. Partial loss of helicity for hswDPCK and the alterations in the content of β-sheets and β-turns for leuDPCK result in less pronounced and slightly shifted minima of their CD spectra.

The tertiary structure of wild type protein and its aromatic amino acid-lacking mutants was estimated by limited proteolysis. Limited proteolysis is a method of protein structure characterization in which a protein of interest is exposed to a protease for short limited periods of time [119]. Proteolytic cleavage of the protein occurs while its tertiary and quaternary structure are still preserved, and the initial cleavage sites are restricted to the regions of the protein that are accessible to the active site of the protease. The sites of limited proteolysis are usually characterized by enhanced backbone flexibility, thus indicating the regions of local unfolding in protein [120].

Endoproteinase Lys-C is a specific protease that cleaves proteins on the terminal C-side of lysine residues. The wild type protein and its aromatic amino acid-lacking mutants have the same content of lysine residues and are characterized by the same pattern of possible Lys-C cleavage sites distributed throughout the amino acid sequences of the proteins, which facilitates the relative comparison of their folding (Fig. 4.19).

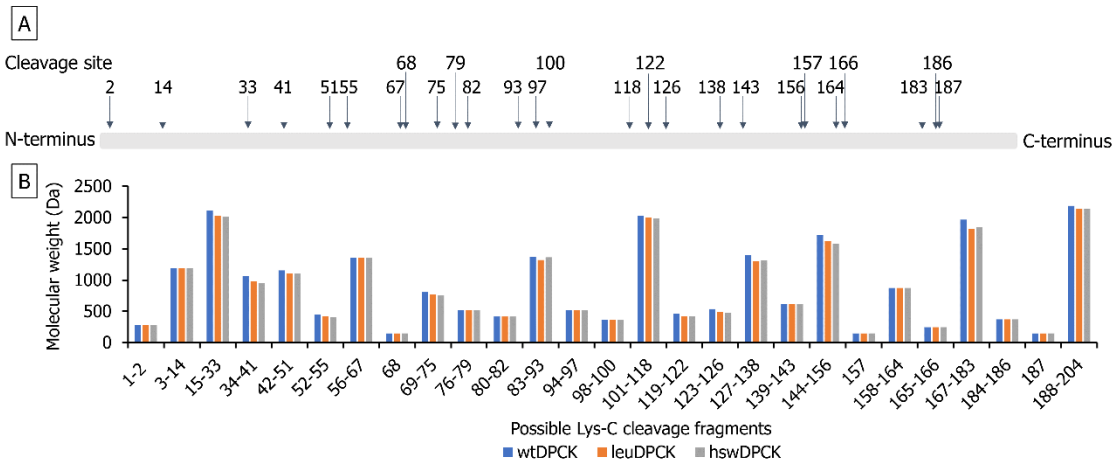


Figure 4.19. (A) Map of possible Lys-C cleavage sites for recombinant proteins and (B) Molecular weights of possible Lys-C cleavage fragments calculated using PeptideCutter ExPASy tool.

The recombinant proteins were exposed to Lys-C endoproteinase from *Lysobacter enzymogenes* (1:200 enzyme:substrate mass ratio) for serial time points, after which the protease was inactivated by heating, and the reaction mixtures were analysed by SDS-PAGE. The areas of the bands corresponding to the intact proteins were estimated from the 14% SDS-polyacrylamide gels using the ImageJ program and then expressed as the amount of uncleaved protein remaining after each time point (Fig. 4. 20).

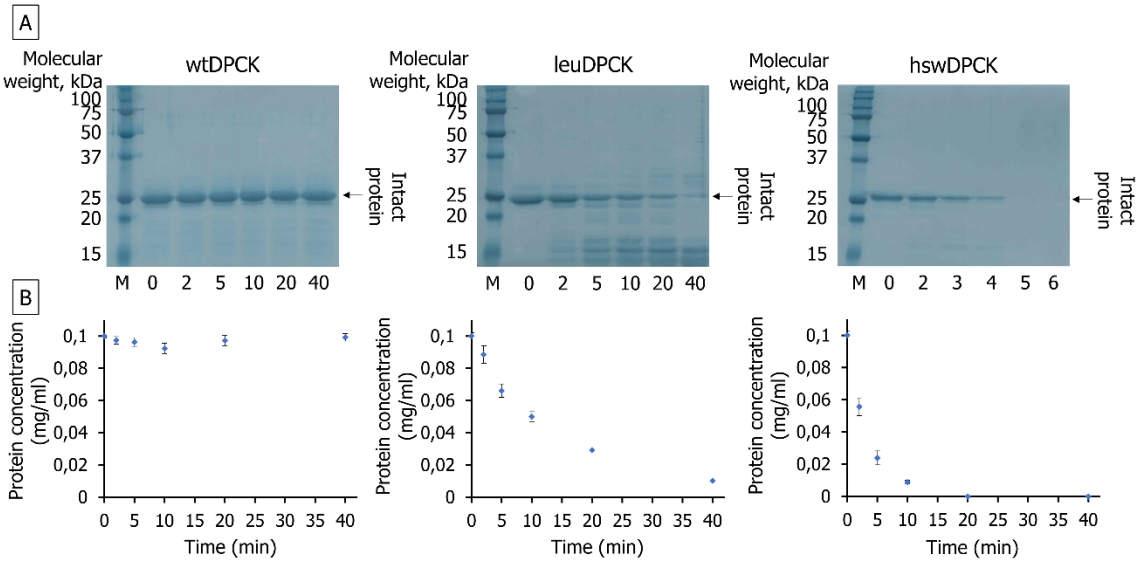


Figure 4.20. Limited proteolysis by Lys-C endoproteinase. (A) 14% SDS-polyacrylamide gels visualized by imidazole-zinc reverse staining after SDS-PAGE with 0,1 mg/ml protein samples in Lys-C buffer exposed to Lys-C endoproteinase for different times: lane M – Precision Plus Protein Dual Xtra Prestained Protein Standards (3 µl), lane 1 – 0 minutes (10 µl), lane 2 – 2 minutes (10 µl), lane 3 – 5 minutes (10 µl), lane 4 – 10 minutes (10 µl), lane 5 – 20 minutes (10 µl), lane 6 – 40 minutes (10 µl). (B) Graphs representing the amount of the uncleaved protein remaining at each time point.

Wild type protein and their mutants lacking aromatic amino acids demonstrate different susceptibility to proteolytic cleavage by Lys-C endoproteinase, which indicates

significant differences in their folding. Wild type protein shows high resistance to proteolytic degradation during the whole time of limited proteolysis. In contrast, both aromatic amino acid-lacking mutants are gradually degraded by Lys-C endoproteinase, and the amounts of the intact leuDPCK and hswDPCK decrease exponentially with time. For leuDPCK the relatively large cleavage fragments with the approximate size of 15 kDa appear during proteolysis, while in the case of hswDPCK no large cleavage fragments are detected.

For quantitative evaluation the rate constants of proteolysis were determined by monitoring the disappearance of intact proteins during proteolytic cleavage by SDS-PAGE. Assuming the pseudo-first order kinetics, the natural logarithms of the intact proteins amounts were plotted against the time, and the plots were fitted with a first-order rate equation

$$\ln [protein] = \ln[protein]_0 - k_p \cdot t, \qquad (4.1)$$

where $[protein]$ is the concentration of the intact protein in the reaction mixture at definite time point $t = t$ (mg/ml), $[protein]_0$ is the concentration of the intact protein at $t = 0$ (mg/ml), t is time (min) and k_p is the rate constant of proteolysis (min^{-1}). The resulted plots are illustrated in Fig. 4.21.

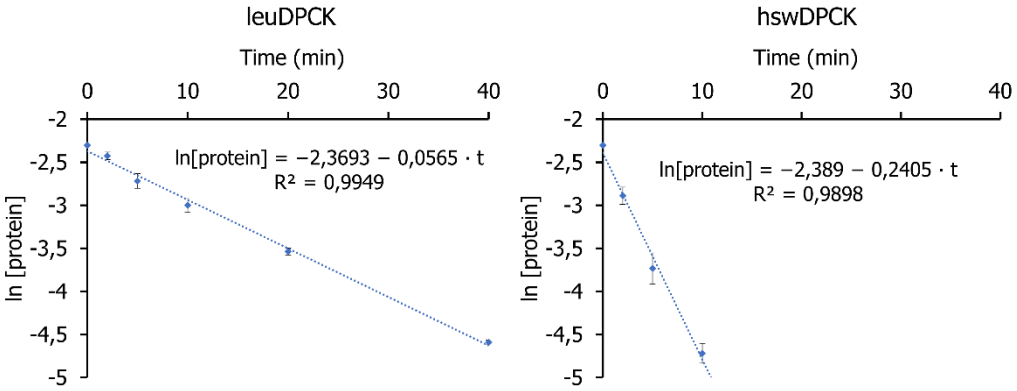


Figure 4.21. Determination of the rate constants of proteolysis from the plots of the natural logarithms of intact protein concentrations versus time.

The estimated rate constants for proteolytic cleavage of leuDPCK and hswDPCK by Lys-C endoproteinase are 0,057 and 0,241 min^{-1}, respectively.

5 Discussion

The major aim of this work was to evaluate the significance of aromatic amino acids for the structure and function of dephosphocoenzyme A kinase. The PDB database was initially searched for solved structures of this enzyme from different organisms. 17 different solved structures of DPCKs from 9 prokaryotic organisms are available in the PDB database. Most of them have been solved for pathogenic microorganisms since the coenzyme A biosynthetic pathway is a popular target for development of new antibacterial agents due to the remarkable differences between the bacterial and the mammalian enzymes involved in this pathway [86]. However, in the retrospective of origin of life the choice of the target dephosphocoenzyme A kinase was narrowed down to those that come from hyperthermophilic organisms: *Aquifex aeolicus* (PDB code 2IF2), *Thermotoga maritima (*PDB code 2GRJ) and *Thermus thermophilus* (PDB code 1UF9). Hyperthermophiles are microorganisms that optimally grow at temperatures between 80 and 110 °C [121]. They are represented only by archaeal and bacterial species that have been found in geothermally heated environments and some of them are considered to be the earliest known ancestors of all extant living organisms on Earth [122]. Therefore, enzymes isolated from hyperthermophilic organisms represent good model systems for understanding enzyme evolution [121].

The content of this work was preceded by recombinant expression, isolation and purification of dephosphocoenzyme A kinases from three above-mentioned hyperthermophilic organisms. An initial test of their expression and solubility and preliminary estimation of their activity which were not included in this work led to selection of DPCK from *Aquifex aeolicus.*

Dephosphocoenzyme A kinase from *Aquifex aeolicus* is a 204-amino-acid protein that contains 19 aromatic amino acid residues: six phenylalanine residues, twelve tyrosine residues and one tryptophan residue. They are spread all over the protein sequence but do not constitute ATP-binding motif, the P-loop. The sequence alignment with previously identified and characterized DPCKs (data not shown) demonstrates that only three of them (Phe21, Phe53 and Phe74) are highly conserved across bacterial DPCKs.

In this thesis, aromatic amino acid residues in DPCK from *Aquifex aeolicus* were substituted by non-aromatic canonical amino acids according to two strategies that imply the best preservation of protein structural stability. According to the first strategy, aromatic amino acids were substituted by leucine, which is among the most order-

promoting amino acids and appears to be one of the earliest additions to the genetic code [51,116]. The second mutant was designed based on the best substitutions for aromatic amino acid residues evaluated by HotSpot Wizard 3.0 web server. Thus, both aromatic amino acid-lacking mutants have 9,3 % of the total protein sequence substituted.

Even though the three-dimensional structure of wild type DPCK from *Aquifex aeolicus* was solved, the recombinant expression, isolation and purification protocols have not been described in literature. Therefore, their establishment was the first aim of this work. Synthetic genes encoding wild type protein and both mutants were successfully cloned and expressed in *Escherichia coli* to produce recombinant proteins with C-terminal polyhistidine-tag. All proteins demonstrated satisfactory expression levels in *Escherichia coli* BL21 (DE3) strain after induction with 0,5 mM IPTG at 37 °C for 4 hours, and were further produced in amounts sufficient for further characterization and purified to homogeneity using a three-step chromatographic purification procedure: anion-exchange chromatography, immobilized metal affinity chromatography and size-exclusion chromatography. Upon purification of the mutant proteins to homogeneity, a slight upper band was observed above the band corresponding to the protein on SDS-PAGE unlike in the case of the wild type protein (Fig. 4.9, page 59). This was unlikely caused by proteolytic cleavage since MALDI showed one specific protein product of the right molecular wright. More probably, the two variants with slightly different electrophoretic mobility were caused by the structural differences (described below in more detail) that could also lead to bigger solvent exposure of some residues and their modifications (such as cysteine oxidation). Reducing agents (β-mercaptoethanol and DTT) were used in relatively low concentrations during purification because of the low tolerance of HisTrap HP column used for immobilized metal affinity chromatography step. Prior to the further characterization, protein samples after size-exclusion chromatography were dialyzed overnight against the buffer with a strong reducing agent (TCEP), concentrated and stored at −80 °C. Both aromatic amino acid-lacking mutants were produced with significantly lower yields than wild type protein which besides their lower expression level and solubility was caused by significant loss of both mutants during anion-exchange chromatography step (Fig. 4.6, page 56). Despite the higher expression level of leuDPCK in comparison with hswDPCK, both mutants were purified with approximately the same yield.

The identities of the recombinantly produced and purified proteins were confirmed by estimating their molecular weights using MALDI. MADLS measurements demonstrated high monodispersity of the protein samples indicating the successful implementation of the purification protocol. Estimation of size homogeneity by size-exclusion chromatography confirmed the solely monomeric state of all proteins in solution, even though the wild type protein was crystallized in a trimeric form. The same behaviour was observed for DPCK from *Escherichia coli* [98]. The sequence alignment confirms the presence of Tyr121 and Lys122 residues in the sequence of DPCK from *Aquifex aeolicus* that are involved in trimerization of DPCK from *Escherichia coli* in the presence of sulphate ions.

The secondary structure of all purified proteins was evaluated by CD spectroscopy. The far-UV CD spectra of wild type protein and both mutants are characterized by high resemblance that indicates good preservation of the secondary structure after substitution of aromatic amino acid residues (Fig. 4.18, page 69). Wild type protein exhibits profound minima between 205 and 225 nm in the far-UV CD spectrum which is charasteric of proteins with predominantly α-helical structure, while leuDPCK and hswDPCK have moderately less pronounced or shifted minima reflecting partial loss of secondary structure content. The far-UV CD spectrum of leuDPCK appears to be more similar to the spectrum of the wild type protein than that of hswDPCK, which indicates higher structural resemblance of leuDPCK to the wild type protein.

Aromatic amino acids residues in proteins have characteristic π-π* absorption bands that generally dominate along with disulfides and prosthetic groups in the near-UV region between 250 and 300 nm [123]. However, in certain cases the shape and the magnitude of the far-UV CD spectrum of a protein can be greatly affected by contributions of aromatic side chains [124]. Far-UV CD spectra of proteins are mostly affected by local interactions between tyrosyl side chains and neighboring backbone amides, which in certain extreme cases result in the unusual CD pattern with a positive CD band between 225 and 230 nm [125]. Thus, the broad aromatic CD bands might interfere with structural interpretation of CD data and prevent accurate comparison of the secondary structure content between wild type protein and aromatic amino acid-lacking mutants. The removal of aromatic amino acid residues from the protein of study also caused that monitoring of protein purification and determination of protein concentration had to be adjusted to allow for direct comparison.

The absence of aromatic amino acid residues in mutants strictly limits the set of methods for characterization of their tertiary structure, since several fast and simple approaches such as intrinsic tryptophan fluorescence spectroscopy and near-UV CD spectroscopy rely on unique spectroscopic properties of aromatic amino acid side chains. Thus, limited proteolysis was chosen as the preliminary method for estimation of changes in tertiary structure caused by aromatic amino acids substitutions. Limited proteolysis is a simple biochemical approach that is based on the fact that proteolysis of a protein substrate can occur only if the polypeptide chain can bind and adapt to the specific stereochemistry of the active site of a protease [120]. Therefore, proteins are cleaved at specific chain fragments that are accessible for binding to the active site of the protease and simultaneously have increased backbone flexibility. Folded proteins are rather resistant to proteolysis due to their native rigid structure, whereas unfolded proteins are degraded much faster [126]. Thus, limited proteolysis can be used to probe protein conformation and identify partly folded states of proteins.

Limited proteolysis by Lys-C endoproteinase reveals significant loss of protein tertiary structure after substitution of aromatic amino acid residues. Wild type protein is highly resistant to proteolytic cleavage by Lys-C endoproteinase indicating that it is well-folded. In contrast, both aromatic amino acid-lacking mutants are characterized by high susceptibility to proteolytic cleavage. Analysis of cleavage fragments by SDS-PAGE and estimation of the rate constants of proteolysis indicates that leuDPCK is characterized by better preservation of tertiary structure than hswDPCK and might retain a partially native hydrophobic core since proteolytic cleavage by Lys-C endoproteinase yields a series of relatively large cleavage fragments with the approximate size of 15 kDa as estimated by SDS-PAGE (Fig. 4.20, page 70). This observation is consistent with high proximity of elution volumes of both proteins observed during analytical size-exclusion chromatography (Fig. 4.12, page 62). On the contrary, hswDPCK is degraded by Lys-C endoproteinase 4,2 times faster than leuDPCK, as estimated by comparison of the rate constants of proteolysis, and proteolysis seems to occur at all posible cleavage sites since no large cleavage fragments were detected even at the earliest time points (Fig. 4.20, page 70). However, a more detailed mass-spectrometric analysis of cleavage fragments is required to confirm the full exposure of the whole polypeptide chain to Lys-C proteinase due to the lack of fix or ordered three-dimensional structure in the case of hswDPCK. A more pronounced distortion of the protein native conformation in the case of hswDPCK is also confirmed by the results of analytical-size exclusion chromatography

and MADLS measurements. Both analyses indicate apparently higher hydrodynamic radius of hswDPCK in contrast to wild type protein and leuDPCK leading to higher mean particle size estimated by MADLS (Fig. 4.13, page 63) and significant shift of the protein elution peak to higher molecular weights during analytical size-exclusion chromatography (Fig. 4.12, page 62). Hydrodynamic radius of a protein molecule is dependant not only on its molecular weight but also the shape, therefore, the elution volumes of the proteins are determined by their Stokes radii rather than molecular weights [127].

The drastic changes in protein tertiary structure caused by aromatic amino acid substitutions are reflected in significant decrease or even loss of catalytic activity of aromatic amino acid-lacking mutants. Phosphotransferase and ATPase activities of wild type protein and both mutants were characterized using a commercial kit relying on fluorometric detection of ADP, one of the reaction products. Wild type DPCK from *Aquifex aeolicus* exhibits solely phosphotransferase activity. Initial tests demonstrated partial maintenance of catalytic activity in the case of leuDPCK and a complete loss of catalytic function in the case of hswDPCK. These results are consistent with the observation of better preservation of protein native conformation for leuDPCK. Enzyme assays with varying concentrations of substrates revealed that wild type DPCK from *Aquifex aeolicus* follows Michaelis-Menten kinetics (Fig. 4.15, page 66). Comparison of estimated kinetic parameters with those of previously identified and characterized DPCKs indicates that DPCK from *Aquifex aeolicus* shows higher affinities toward both substrates than DPCKs from *Mycobacterium tuberculosis* and *Escherichia Coli* (24,3 ± 1,7; 34,9 and 740 μM for dCoA and 12,7 ± 0,3; 56,8 and 140 μM for ATP, respectively) and exhibits turnover numbers and catalytic efficiencies similar to those of DPCK from *Entamoeba histolytica*. LeuDPCK shows lower affinity towards ATP (65,9 ± 4,5 and 12,7 ± 0,7 μM for leuDPCK and wtDPCK, respectively) and catalytic efficiency (355 ± 12 and 57480 ± 2050 $s^{-1} \cdot M^{-1}$ for leuDPCK and wtDPCK, respectively), mainly as a result of a decreased turnover number (0,0234 ± 0,0008 and 0,730 ± 0,026 s^{-1} for leuDPCK and wtDPCK, respectively). LeuDPCK follows the Michaelis−Menten kinetics only for varying concentrations of ATP (fig. 4.16, page 66). Substitutions of aromatic amino acids might have led to drastic changes in conformations of the lid and the CoA domains of DPCK that made it impossible for the enzyme to adopt the closed conformation in which the lid domain shields the active site and prevents transfer of phosphate group from ATP to water [105]. While varying concentration of dCoA, the estimated reaction rate does not

change at concentrations of dCoA lower than 50 μM, however, a significant increase of the reaction rate can be observed for higher dCoA concentrations, which might indicate non-essential activation (fig. 4.16, page 66). Thus, leuDPCK can be classified as an ATPase activated by dCoA rather than a kinase. However, leuDPCK is still capable of less efficient transfer of phosphate group from ATP to dCoA, as confirmed by HPLC/MS analysis (Tab. 4.4, page 68).

In conclusion, the results of this thesis provide an experimetal exemplary evidence in support of the idea that catalytically active polypeptides could have emerged at the very early stages of life evolution even before the completion of the standard genetic code.

6 Conclusion

- Wild type protein and two aromatic amino acid-lacking mutants of dephosphocoenzyme A kinase from *Aquifex aeolicus* were successfully expressed in *Escherischia coli* BL21(DE3) and purified by a three-step chromatographic procedure using anion-exchange chromatography, immobilized metal affinity chromatography and size exclusion chromatography.
- Proteins purified to homogeneity were characterized by MALDI, MADLS and analytical size-exclusion chromatography.
- The effect of substitution of aromatic amino acids on the secondary and tertiary structure of wild type protein was estimated by far-UV CD spectroscopy and limited proteolysis. Substitution of aromatic amino acids leads to good preservation of secondary structure of wild type protein, but results in drastic changes of its tertiary structure.
- The effect of substitution of aromatic amino acids on catalytic activity of wild type protein was determined. Characterization of enzyme activity shows partial preservation of catalytic activity in one of the aromatic amino acid-lacking mutant.

7 References

1. Fournier G. P.; Alm E. J.: Ancestral Reconstruction of a Pre-LUCA Aminoacyl-tRNA Synthetase Ancestor Supports the Late Addition of Trp to the Genetic Code. *J. Mol. Evol.* **80**:3−4, 171−85 (2015).

2. Miller S. L.: The prebiotic synthesis of organic compounds as a step toward the origin of life. In: *Major Events in the History of Life*. Schopf J. W. (Eds.). Boston: Jones & Bartlett Publishers, 1992, p. 1−28. ISBN 0867202688.

3. Bada J. L.; Bigham C.; Miller S. L.: Impact melting of frozen oceans on the early Earth: implications for the origin of life. *Proc. Natl. Acad. Sci. USA*. **91**:4, 1248−50 (1994).

4. Shock E. L.; Amend J. P.; Zolotov M. Y.: The early Earth vs the origin of life. In: *Origin of the Earth and Moon*. Canup R. M.; Righter K. (Eds.). Tucson: University of Arizona Press, 2000, p. 527–43. ISBN 0816520739.

5. Wächtershäuser G.: Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev*. **52**:4, 452–84 (1988).

6. Morowitz H. J.; J D Kostelnik J. D.; Yang J.; Cody G. D.: The Origin of Intermediary Metabolism. *Proc. Natl. Acad. Sci. USA*. **97**:14, 7704−8 (2000).

7. Cleaves II H. J.: The origin of the biologically coded amino acids. *J. Teor. Biol.* **263**:4, 490−8 (2010).

8. Miller S. L.: A production of Amino Acids Under Possible Primitive Earth Conditions. *Science*. **117**:3046, 528−9 (1953).

9. Miller S. L.; Urey H. C.: Organic Compound Synthesis on the Primitive Earth. *Science*. **130**:3370, 245–51 (1959).

10. Miller S. L.: The mechanism of synthesis of amino acids by electric discharges. *Biochim. Biophys. Acta*. **23**, 480−9 (1957).

11. Oró, J.; Kamat, S. S.: Amino acid Synthesis From Hydrogen Cyanide Under Possible Primitive Earth Conditions. *Nature*. **190**, 442−3 (1961).

12. Harada K.: Formation of Amino-acids by Thermal Decomposition of Formamide – Oligomerization of Hydrogen Cyanide. *Nature*. **214**, 479–80 (1967).

13. Ferris J. P.; Joshi P. C; Edelson E. H.; Lawless J. G.: HCN: A plausible source of purines, pyrimidines and amino acids on the primitive earth. *J. Mol. Evol.* **11**, 293–311 (1978).

14. Taillades J.; Beuzelin I.; Garrel L.; Tabacik V.; Bied C.; Commeyras A.: N-Carbamoyl-α-Amino Acids Rather than Free α-Amino Acids Formation in the Primitive

Hydrosphere: A Novel Proposal for the Emergence of Prebiotic Peptides. *Orig. Life Evol. Biosph*. **28**, 61–77 (1998).

15. Friedmann N.; Miller S. L.: Phenylalanine and Tyrosine Synthesis Under Primitive Earth Conditions. *Science*. **166**:3906, 766−7 (1969).

16. Sagan C.; Khare B. N.: Long-wavelength ultraviolet photoproduction of amino acids on the primitive Earth. *Science*. **173**:3995, 417−20 (1971).

17. Van Trump J. E.; Miller S. L.: Prebiotic Synthesis of Methionine. *Science.* **178**:4063, 859−60 (1972).

18. Vázquez-Salazar A.; Becerra A.; Lazcano A.: Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines. *PLoS One*. **13**:4, article number: e0196349, 1−22 (2018).

19. Anders E.; Grevesse N.: Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta*. **53**:1, 197−214 (1989).

20. Lazcano A.; Miller S. L.: The Origin and Early Evolution of Life: Prebiotic Chemistry, the pre-RNA World, and Time. *Cell*. **85**:6, 793−8 (1996).

21. McSween Jr. H. Y., Lauretta D. S., Leshin L. A.: Recent Advances in Meteoritics and Cosmochemistry. In: *Meteorites and early Solar system II*. Lauretta D. S., Leshin L. A. (Eds.). Tucson: University of Arizona Press, 2006, p. 53–66. ISBN 0816525625.

22. Miller S. L., Urey H. C., Oró J.: Origin of Organic Compounds on the Primitive Earth and in Meteorites. *J. Mol. Evol*. **9**:1, 59−72 (1976).

23. Botta O.; Bada J. L.: Extraterrestrial organic compounds in meteorites. *Surv. Geophys*. **23**. 411−67 (2002).

24. Sephton M. A.: Organic Compounds in Carbonaceous Meteorites. *Nat. Prod. Rep.* **19**:3, 292−311 (2002).

25. Pizzarello S.: The Chemistry of Life's Origin: A Carbonaceous Meteorite Perspective *Acc. Chem. Res.* **39**:4, 231–7 (2006).

26. Ring D.; Wolman Y., Friedmann N.; Miller S. L.: Prebiotic Synthesis of Hydrophobic and Protein Amino Acids. *Proc. Natl. Acad. Sci. USA*. **69**:3, 765−8 (1972).

27. Wolman Y.; Haverland W. J.; Miller S. L.: Nonprotein amino acids from spark discharges and their comparison with the Murchison meteorite amino acids. *Proc. Natl. Acad. Sci. USA*. **69**:4, 809−11 (1972).

28. Peltzer E. T; Bada J. L.; Schlesinger G.; Miller S. L.: The Chemical Conditions on the Parent Body of the Murchison Meteorite: Some Conclusions Based on Amino, Hydroxy and Dicarboxylic Acids. *Adv. Space Res*. **4**:12, 69−74 (1984).

29. Lu Y.; Freeland S. J.: A Quantitative Investigation of the Chemical Space Surrounding Amino Acid Alphabet Formation. *J. Theor. Biol*. **250**:2, 349−61 (2008).

30. Sengupta S.; Higgs P. G.: Pathways of Genetic Code Evolution in Ancient and Modern Organisms. *J. Mol. Evol*. **80**:5−6, 229−43 (2015).

31. Ilardo M.; Meringer M.; Freeland S. J.; Rasulev B.; Cleaves II H. J.: Extraordinarily adaptive properties of the genetically encoded amino acids. *Sci. Rep*. **5**, article number: 9414, 1−6 (2015).

32. Wong J. T.-F.: The evolution of a universal genetic code. *Proc. Natl. Acad. Sci. USA*. **73**:7, 2336−40 (1976).

33. Trifonov E. N.: The Triplet Code From First Principles. *J. Biomol. Struct. Dyn*. **22**:1, 1−11 (2004).

34. Miller S. L.; Orgel L. E.: *The Origins of Life on the Earth*. New Jersey, Prentice-Hall, 1974. ISBN 9780136420743.

35. White R. H.: Hydrolytic Stability of Biomolecules at High Temperatures and Its Implication for Life at 250 Degrees C. *Nature*. **310**, 430−2 (1984).

36. Cleaves H. J.; Miller S. L. Oceanic protection of prebiotic organic compounds from UV radiation. *Proc. Natl. Acad. Sci. USA*. **95**:13, 7260–3 (1998).

37. Canuto V. M.; Levine J. S.; Augustsson T. R.; Imhoff C. L.; Giampapa M. S.: The young sun and the atmosphere and photochemistry of the early earth. *Nature*. **305**, 281–6 (1983).

38. Wong J. T.: Evolution and mutation of the amino acid code. In: *Dynamics of Biochemical Systems*. Ricard J.; Cornish-Bowden A. (Eds.). New York: Plenum Press, 1984, p. 247–57. ISBN 9780306418303.

39. Cockell C. S.; Airo A.: On the Plausibility of a UV Transparent Biochemistry. *Orig. Life Evol. Biosph*. **32**:3, 255−74 (2002).

40. Di Mauro E.; Dunker A. K.; Trifonov E. N.: Disorder to Order, Nonlife to Life: In the Beginning There Was a Mistake. In: *Genesis - In The Beginning. Cellular Origin, Life in Extreme Habitats and Astrobiology, vol. 22*. Seckbach J. (Eds.). Dordrecht: Springer, 2012, p. 415−35. ISBN 9789400729407.

41. Granold M.; Hajieva P.; Toşa M. I.; Irimie F. D.; Moosmann B.: Modern diversification of the amino acid repertoire driven by oxygen. *Proc. Natl. Acad. Sci. USA.* **115**:1, 41−6 (2017).

42. Pearson R. G.: Absolute electronegativity and hardness correlated with molecular orbital theory. *Proc. Natl. Acad. Sci. USA.* **83**:22, 8440–1 (1986).

43. Aihara J.: Reduced HOMO−LUMO Gap as an Index of Kinetic Stability for Polycyclic Aromatic Hydrocarbons. *J. Phys. Chem. A*. **103**:37, 7487–95 (1999).

44. Wong J. T.-F.: Coevolution Theory of the Genetic Code at Age Thirty. *Bioessays*. **27**:4, 416−25 (2005).

45. Koonin E. V.; Novozhilov A. S.: Origin and evolution of the genetic code: the universal enigma. *IUBMB Life*. **61**:2, 99−111 (2009).

46. Murphy L. R.; Wallqvist A.; Levy R. M.: Simplified Amino Acid Alphabets for Protein Fold Recognition and Implications for Folding. *Protein Eng*. **13**:3, 149−52 (2000).

47. Akanuma S.; Kigawa T.; Yokoyama S.: Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc. Natl. Acad. Sci. USA.* **99**:21, 13549−53 (2002).

48. Longo L. M.; Blaber M.: Protein Design at the Interface of the Pre-Biotic and Biotic Worlds. *Arch. Biochem. Biophys*. **526**:1, 16−21 (2012).

49. Weber A. L.; Miller S. L.: Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol*. **17**:5, 273–84 (1981).

50. Higgs P. G; Pudritz R. E.: A Thermodynamic Basis for Prebiotic Amino Acid Synthesis and the Nature of the First Genetic Code. *Astrobiology*. **9**:5, 483−90 (2009).

51. Trifonov E. N.: Tracing Life back to elements. *Phys. Life Rev*. **5**:2, 121−32 (2008).

52. Osawa S.; Jukes T. H: Codon reassignment (codon capture) in evolution. *J. Mol. Evol*. **28**:4, 271−8 (1989).

53. Woese C. R.; Olsen G. J.; Ibba M.; Söll D.: Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process. *Microbiol. Mol. Biol. Rev*. **64**:1, 202−36 (2000).

54. Caetano-Anollés G.; Kim K. M.; Caetano-Anollés D.: The Phylogenomic Roots of Modern Biochemistry: Origins of Proteins, Cofactors and Protein Biosynthesis. *J. Mol. Evol*. **74**:1−2, 1−34 (2012).

55. Merkl R.; Sterner R.: Reconstruction of ancestral enzymes. *Perspect. Sci*. **9**, 17−23 (2016).

56. Brooks D. J.; Fresco J. R.: Increased Frequency of Cysteine, Tyrosine, and Phenylalanine Residues Since the Last Universal Ancestor. *Mol. Cell. Proteomics*. **1**:2, 125−31 (2002).

57. Fournier G. P.; Gogarten J. P.: Signature of a Primitive Genetic Code in Ancient Protein Lineages. *J. Mol. Evol.* **65**:4, 425−36 (2007).

58. Fournier G. P.; Gogarten J. P.: Rooting the Ribosomal Tree of Life. *Mol. Biol. Evol*. **27**:8, 1792−801 (2010).

59. Cavalcanti A. R. O.; Leite E. S.; Neto B. B.; Ferreira R.: On the Classes of aminoacyl-tRNA Synthetases, Amino Acids and the Genetic Code. *Orig. Life Evol. Biosph*. **34**:4, 407−20 (2004).

60. Klipcan L.; Safro M.: Amino Acid Biogenesis, Evolution of the Genetic Code and aminoacyl-tRNA Synthetases. *J. Theor. Biol*. **228**:3, 389−96 (2004).

61. Yang X.-L.; Otero F. J.; Skene R. J.; McRee D. E.; Schimmel P.; Pouplana L. R. Crystal Structures That Suggest Late Development of Genetic Code Components for Differentiating Aromatic Side Chains. *Proc. Natl. Acad. Sci. USA*. **100**:26, 15376−80 (2003).

62. Cech T. R.: The RNA Worlds in Context. *Cold Spring Harb. Perspect. Biol*. **4**:7, 1−5 (2012).

63. Higgs P. G; Lehman N.: The RNA World: Molecular Cooperation at the Origins of Life. *Nat. Rev. Genet*. **16**:1, 7−17 (2015).

64. DeGuzman V.; Vercoutere W.; Shenasa H.; Deamer D.: Generation of Oligonucleotides Under Hydrothermal Conditions by Non-Enzymatic Polymerization. *J. Mol. Evol*. **78**:5, 251−62 (2014).

65. Stairs S.; Nikmal A.; Bučar D.-K.; Zheng S.-L.; Szostak J.-W.; Powner M. W.: Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. *Nat. Commun*. **8**:13, article number: 15270, 1−12 (2017).

66. Bregestovski P. D.: "RNA World", a highly improbable scenario of the origin and early evolution of life on Earth. *J. Evol. Biochem. Physiol*. **51**:1, 72−84 (2015).

67. Joyce G. F.: The Antiquity of RNA-based Evolution. *Nature.* **418**:6894, 214−21 (2002).

68. Hartman H.; Smith T. F.: Origin of the Genetic Code Is Found at the Transition between a Thioester World of Peptides and the Phosphoester World of Polynucleotides. *Life*. **9**:3, article number: 69, 1−17 (2019).

69. Baltscheffsky H. and Baltscheffsky M.: Molecular origin and evolution of early biologic energy conversion. In: *Early Life on Earth, Nobel Symposium No. 84.* S. Bengtson (Eds.). New York, Columbia University Press, 1994, p. 81−90. ISBN 0231080883.

70. Schwartz A. W.: Phosphorus in Prebiotic Chemistry. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **361**:1474, 1743−9 (2006).

71. Tian T.; Chu X.-Y.; Yang Y.; Zhang X.; Liu Y.-M.; Gao J.; Ma B.-G.; Zhang H.-Y.: Phosphates as Energy Sources to Expand Metabolic Networks. *Life*. **9**:2, article number: 43, 1−12 (2019).

72. Bada J. L.: New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem. Soc. Rev*. **42**:5, 2186−96 (2013).

73. Parker E. T.; Cleaves H. J.; Dworkin J. P.; Glavin D. P.; Callahan M.; Aubrey A.; Lazcano A.; Bada J. L.: Primordial synthesis of amines and amino acids in a 1958 Miller H_2S-rich spark discharge experiment. *Proc. Natl. Acad. Sci. USA*. **108**:14, 5526−31 (2011).

74. Miller S. L.; Schlesinger G.: Prebiotic Syntheses of Vitamin Coenzymes: I. Cysteamine and 2-mercaptoethanesulfonic Acid (Coenzyme M). *J. Mol. Evol*. **36**:4, 302−7 (1993).

75. Miller S. L.; Schlesinger G.: Prebiotic syntheses of vitamin coenzymes: II. Pantoic acid, pantothenic acid, and the composition of coenzyme A. *J. Mol. Evol*. **36**:4, 308−14 (1993).

76. Keefe A. D.; Newton G. L.; Miller S. L.: A possible prebiotic synthesis of pantetheine, a precursor to coenzyme A. *Nature*. **373**:6516, 683−5 (1995).

77. Benner S. A; Ellington A. D.; Tauer A.: Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. USA*. **86**:18, 7054−8 (1989).

78. Genschel U.: Coenzyme A Biosynthesis: Reconstruction of the Pathway in Archaea and an Evolutionary Scenario Based on Comparative Genomics. *Mol. Biol. Evol*. **21**:7, 1242−51 (2004).

79. Lee C. H.; Chen, A. F.: Immobilized coenzymes and derivatives. In: *The Pyridine Nucleotide Coenzymes*. Everse J. (Eds.); Anderson B. (Eds.); You K. (Eds.), New York, Academic Press, 1982, p. 189−218. ISBN 9780122447501.

80. Jakubowski H.: Homocysteine Editing, Thioester Chemistry, Coenzyme A, and the Origin of Coded Peptide Synthesis. *Life*. **7**:1, article number: 6, 1−26 (2017).

81. Jakubowski H. Aminoacylation of coenzyme A and pantetheine by aminoacyl-tRNA synthetases: possible link between noncoded and coded peptide synthesis. *Biochemistry*. **37**:15, 5147−53 (1998).

82. Caetano-Anollés D.; Caetano-Anollés G.: Piecemeal Buildup of the Genetic Code, Ribosomes, and Genomes From Primordial tRNA Building Blocks. *Life*. **6**:4, article number: 43, 1−24 (2016).

83. Leonardi R.; Zhang Y.-M.; Rock C. O.; Jackowski S.: Coenzyme A: Back in action. *Prog. Lipid Res*. **44**:2−3, 125−53 (2005).

84. Duncan D.; Auclair K.: The Coenzyme A Biosynthetic Pathway: A New Tool for Prodrug Bioactivation. *Arch. Biochem. Biophys*. **672**, article number: 108069, 1−14 (2019).

85. Hart R. J.; Abraham A.; Aly A. S. I.: Genetic Characterization of Coenzyme A Biosynthesis Reveals Essential Distinctive Functions during Malaria Parasite Development in Blood and Mosquito. *Front. Cell Infect. Microbiol*. **7**, article number: 260, 1−9 (2017).

86. Balibar C. J.; Hollis-Symynkywicz M. F.; Tao J.: Pantethine rescues phospho-pantothenoylcysteine synthetase and phosphopantothenoylcysteine decarboxylase deficiency in *Escherichia coli* but not in *Pseudomonas aeruginosa*. *J. Bacteriol*. **193**:13, 3304−12 (2011).

87. Webb M. E.; Smith A. G.: Pantothenate Biosynthesis in Higher Plants. *Adv. Bot. Res*. **58**, 203−255 (2011).

88. Ernst D. C.; Downs D. M. The STM4195 Gene Product (PanS) Transports Coenzyme A Precursors in *Salmonella enterica*. *J. Bacteriol*. **197**:8, 1368−1377 (2015).

89. Kotnik, M.; Anderluh P. S.; Prezelj A.: Development of novel inhibitors targeting intracellular steps of peptidoglycan biosynthesis. *Curr. Pharm. Des*. **13**:22, 2283−309 (2007).

90. Beveridge T. J.: Ultrastructure, Chemistry, and Function of the Bacterial Wall. *Int. Rev. Cytol*. **72**, 229−317 (1981).

91. Anderson M. S.; Bulawa C. E.; Raetz C. R.: The Biosynthesis of Gram-Negative Endotoxin. Formation of Lipid A Precursors From UDP-GlcNAc in Extracts of *Escherichia Coli*. *J. Biol. Chem*. **260**:29, 15536−41 (1985).

92. Harrington C. R.; Baddiley J.: Biosynthesis of Wall Teichoic Acids in Staphylococcus Aureus H, Micrococcus Varians and Bacillus Subtilis W23. Involvement of Lipid

Intermediates Containing the Disaccharide N-acetylmannosaminyl N-acetylglucosamine. *Eur. J. Biochem*. **153**:3, 639−45 (1985).

93. Ramos-Aires J.; Plésiat P.; Kocjancic-Curty L.; Köhler T.: Selection of an Antibiotic-Hypersusceptible Mutant of Pseudomonas Aeruginosa: Identification of the GlmR Transcriptional Regulator. *Antimicrob. Agents Chemother*. **48**:3, 843−51 (2004).

94. Yang K.; Eyobo Y.; Brand L. A.; Martynowski D.; Tomchick D.; Strauss E.; Zhang H.: Crystal Structure of a Type III Pantothenate Kinase: Insight into the Mechanism of an Essential Coenzyme A Biosynthetic Enzyme Universally Distributed in Bacteria. *J. Bacteriol*. **188**:15, 5532−40 (2006).

95. Strauss E.; Kinsland C.; Ge Y.; McLafferty F. W.; Begley T. P.: Phospho-pantothenoylcysteine synthetase from *Escherichia coli*. Identification and characterization of the last unidentified coenzyme A biosynthetic enzyme in bacteria. *J. Biol. Chem*. **276**:17, 13513−16 (2001).

96. Strauss E.; Mander L.; Liu H.-W.: Coenzyme A biosynthesis and enzymology. In: *Comprehensive Natural Products II Chemistry and Biology*. Liu H.-W. (Eds.); Mander L (Eds.). Oxford, Elsevier, 2010, p. 351−410. ISBN 9780080453828.

97. Daugherty M.; Polanuyer B.; Farrell M.; Scholle M.; Lykidis A.; Crécy-Lagard V. D.; Osterman A.: Complete Reconstitution of the Human Coenzyme A Biosynthetic Pathway via Comparative Genomics. *J. Biol. Chem*. **277**:24, 21431−9 (2002).

98. O'Toole N.; Barbosa J. A. R. G.; Li Y.; Hung L.-W.; Matte A.; Cygler M.: Crystal Structure of a Trimeric Form of Dephosphocoenzyme A Kinase From *Escherichia Coli*. *Protein Sci.* **12**:2, 327−36 (2003).

99. Shimosaka T.; Makarova K. S.; Koonin E. V.; Atomi H.: Identification of Dephospho-Coenzyme A (Dephospho-CoA) Kinase in *Thermococcus Kodakarensis* and Elucidation of the Entire CoA Biosynthesis Pathway in Archaea. *mBio*. **10**:4, article number: e01146-19, 1−14 (2019).

100. Mishra P. K.; Park P. K.; Drueckhammer D. G.: Identification of *yacE* (*coaE*) as the Structural Gene for Dephosphocoenzyme A Kinase in *Escherichia coli* K-12. *J. Bacteriol.* **183**:9, 2774–8 (2001).

101. Walia G.; Kumar P.; Surolia A.: The Role of UPF0157 in the Folding of *M. Tuberculosis* Dephosphocoenzyme A Kinase and the Regulation of the Latter by CTP. *PLoS One*. **4**:10, article number: e7645, 1−16 (2009).

102. Nurkanto A.; Jeelani G.; Yamamoto T.; Hishiki T.; Naito Y.; Suematsu M.; Hashimoto T.; Nozaki T.: Biochemical, Metabolomic, and Genetic Analyses of Dephospho Coenzyme A Kinase Involved in Coenzyme A Biosynthesis in the Human

Enteric Parasite *Entamoeba histolytica*. *Front. Microbiol*. **9**, article number: 2902, 1−18 (2018).

103. Aghajanian S.; Worrall D. M.: Identification and characterization of the gene encoding the human phosphopantetheine adenylyltransferase and dephospho-CoA kinase bifunctional enzyme (CoA synthase). *Biochem. J.* **365**, 13−8 (2002).

104. Wadler C.; Cronan J. E.: Dephospho-CoA Kinase Provides a Rapid and Sensitive Radiochemical Assay for Coenzyme A and Its Thioesters. *Anal. Biochem*. **368**:1, 17−23 (2007).

105. Obmolova G.; Teplyakov A.; Bonander N.; Eisenstein E.; Howard A. J.; Gilliland G. L.: Crystal Structure of Dephospho-Coenzyme A Kinase From *Haemophilus Influenzae*. *J. Struct. Biol*. **136**:2, 119−25 (2001).

106. Seto A.; Murayama K.; Toyama M.; Ebihara A.; Nakagawa N.; Kuramitsu S.; Shirouzu M.; Yokoyama S.: ATP-induced structural change of dephosphocoenzyme A kinase from Thermus thermophilus HB8. *Proteins*. **58**:1, 235−42 (2005).

107. Gong X.; Chen X.; Yu D.; Zhang N.; Zhu Z.; Niu L.; Mao Y.; Ge H.: Crystal structure of Legionella pneumophila dephospho-CoA kinase reveals a non-canonical conformation of P-loop. *J. Struct. Biol*. **188**:3, 233−39 (2014).

108. Cheek S.; Zhang H.; Grishin N. V.: Sequence and Structure Classification of Kinases. *J. Mol. Biol*. **320**:4, 855−81 (2002).

109. Via A.; Ferrè F.; Brannetti B.; Valencia A.; Helmer-Citterich M.: Three-dimensional view of the surface motif associated with the P-loop structure: *cis* and *trans* cases of convergent evolution. *J. Mol. Biol*. **303**:4, 455−65 (2000).

110. Walia G.; Surolia A. Insights Into the Regulatory Characteristics of the Mycobacterial Dephosphocoenzyme A Kinase: Implications for the Universal CoA Biosynthesis Pathway. PLoS One. 6:6, article number: e21390, 1−12 (2011).

111. Lin C.-Y.; Wang V.; Shui H.-A.; Juang R.-H.; Hour A.-L.; Chen P.-S.; Huang H.-M.; Wu S.-Y.; Lee J.-C.; Tsai T.-L.; Chen H.-M.: A comprehensive evaluation of imidazole-zinc reverse stain for current proteomic researches. *Proteomics*. **9**:3, 696−709 (2009).

112. Fernandez-Patron C.; Castellanos-Serra L.; Rodriguez P.: Reverse Staining of Sodium Dodecyl Sulfate Polyacrylamide Gels by Imidazole-Zinc Salts: Sensitive Detection of Unmodified Proteins. *Biotechniques*. **12**:4, 564−73 (1992).

113. Bradford M. M.: A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem*. **72**:1−2, 248−54 (1976).

114. “ADP Accumulation Assay Platform” [online]. Available from URL: <https://www.discoverx.com/technologies-platforms/fluorescence-intensity-technology/adp-accumulation-assay-platform> [cit. 12.04.20].

115. Dill K. A.: Dominant forces in protein folding. *Biochemistry*. **29**:31, 7133−55 (1990).

116. Campen A.; Williams R. M.; Brown C. J.; Meng J.; Uversky V. N.; Dunker A. K.: TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder. *Protein Pept. Lett*. **15**:9, 956–63 (2008).

117. Sumbalova L.; Stourac J.; Martinek T.; Bednar D.; Damborsky J.: HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Res*. **46**:W1, W356–62 (2018).

118. “pET System Manual, 8th edition” [online]. Available from URL: <https://research.fhcrc.org/content/dam/stripe/hahn/methods/biochem/pet.pdf> [cit. 12.04.20].

119. Petrotchenko E. V.; Borchers C. H.: Modern Mass Spectrometry-Based Structural Proteomics. *Adv. Protein Chem. Str*. **95**, 193–213 (2014).

120. Fontana A.; Polverino de Laureto P.; Spolaore B.; Frare E.; Picotti P.; Zambonin M.: Probing protein structure by limited proteolysis. *Acta Biochim. Pol*. **51**:2, 299−321 (2004).

121. Vieille C.; Zeikus G. J.: Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev*. **65**:1, 1−43 (2001).

122. Stetter K. O.: Hyperthermophiles in the history of life. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **361**:1474, 1837–43 (2006).

123. Woody R. W. Circular dichroism. *Methods Enzymol*. **246**, 34−71 (1995).

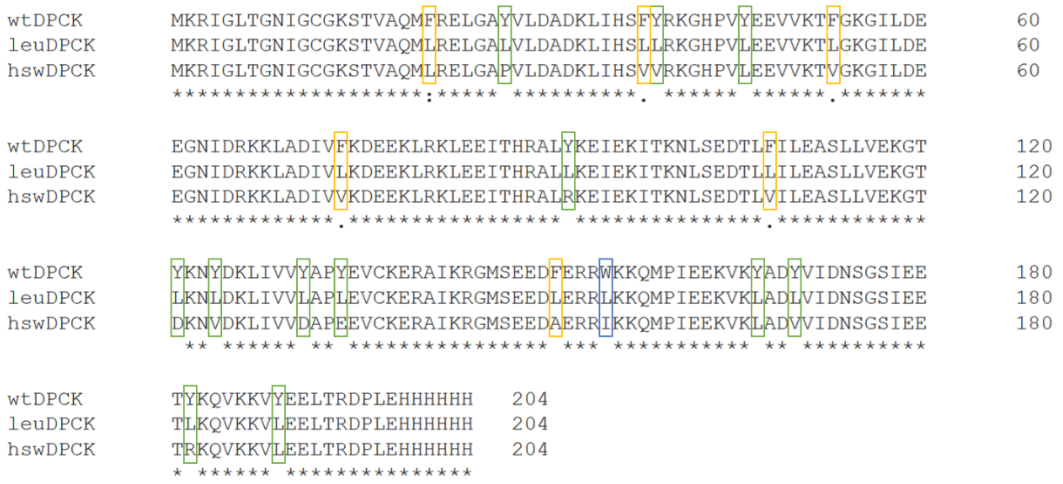
124. Krittanai C.; Johnson W. C.: Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains. *Anal. Biochem*. **253**:1, 57−64 (1997).

125. Woody R. W.: Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins. *Biopolymers*. **17**:6, 1451–67 (1978).

126. Fontana A.; Polverino de Laureto P.; De Filippis V.; Scaramella E.; Zambonin M.: Probing the partly folded states of proteins by limited proteolysis. *Fold. Des*. **2**:2, R17−26 (1997).

127. Ackers G. K.: Analytical gel chromatography of proteins. *Adv. Protein Chem*. **24**, 343−446 (1970).

Supplementary materials



Supplementary figure S1. Sequence alignment of wild type protein and aromatic amino acid-lacking mutants of dephosphocoenzyme A kinase from *Aquifex aeolicus*. Aromatic amino acids are highlighted in sequences: green – tyrosine, yellow – phenylalanine and blue – tryptophan. Alignment was performed using Clustal Omega program.

>DPCK_WT

5′-ATGAAACGTATCGGTCTGACCGGTAACATCGGTTGCGGTAAATCTACCGTTGCGCAGATGTTCCGTGAACTGGGTGCGTACGT
TCTGGACGCGGACAAACTGATCCACTCTTTCTACCGTAAAGGTCACCCGGTTTACGAAGAAGTTGTTAAAACCTTCGGTAAAGGTA
TCCTGGACGAAGAAGGTAACATCGACCGTAAAAAACTGGCGGACATCGTTTTCAAAGACGAAGAAAAACTGCGTAAACTGGAAGAA
ATCACCCACCGTGCGCTGTACAAAGAAATCGAAAAAATCACCAAAAACCTGTCTGAAGACACCCTGTTCATCCTGGAAGCGTCTCT
GCTGGTTGAAAAAGGTACCTACAAAAACTACGACAAACTGATCGTTGTTTACGCGCCGTACGAAGTTTGCAAAGAACGTGCGATCA
AACGTGGTATGTCTGAAGAAGACTTCGAACGTCGTTGGAAAAAACAGATGCCGATCGAAGAAAAAGTTAAATACGCGGACTACGTT
ATCGACAACTCTGGTTCTATCGAAGAAACCTACAAACAGGTTAAAAAAGTTTACGAAGAACTGACCCGTGACCCGCTGGAA-3′

>DPCK_LEU

5′-ATGAAACGTATCGGTCTGACCGGTAACATCGGTTGCGGTAAATCTACCGTTGCGCAGATGCTGCGTGAACTGGGTGCGCTGGT
TCTGGACGCGGACAAACTGATCCACTCTCTGCTGCGTAAAGGTCACCCGGTTCTGGAAGAAGTTGTTAAAACCCTGGGTAAAGGTA
TCCTGGACGAAGAAGGTAACATCGACCGTAAAAAACTGGCGGACATCGTTCTGAAAGACGAAGAAAAACTGCGTAAACTGGAAGAA
ATCACCCACCGTGCGCTGCTGAAAGAAATCGAAAAAATCACCAAAAACCTGTCTGAAGACACCCTGCTGATCCTGGAAGCGTCTCT
GCTGGTTGAAAAAGGTACCCTGAAAAACCTGGACAAACTGATCGTTGTTCTGGCGCCGCTGGAAGTTTGCAAAGAACGTGCGATCA
AACGTGGTATGTCTGAAGAAGACCTGGAACGTCGTCTGAAAAAACAGATGCCGATCGAAGAAAAAGTTAAACTGGCGGACCTGGTT
ATCGACAACTCTGGTTCTATCGAAGAAACCCTGAAACAGGTTAAAAAAGTTCTGGAAGAACTGACCCGTGACCCGCTGGAA-3′

>DPCK_HSW

5′-ATGAAACGTATCGGTCTGACCGGTAACATCGGTTGCGGTAAATCTACCGTTGCGCAGATGCTGCGTGAACTGGGTGCGCCGGT
TCTGGACGCGGACAAACTGATCCACTCTGTTGTTCGTAAAGGTCACCCGGTTCTGGAAGAAGTTGTTAAAACCGTTGGTAAAGGTA
TCCTGGACGAAGAAGGTAACATCGACCGTAAAAAACTGGCGGACATCGTTGTTAAAGACGAAGAAAAACTGCGTAAACTGGAAGAA
ATCACCCACCGTGCGCTGCGTAAAGAAATCGAAAAAATCACCAAAAACCTGTCTGAAGACACCCTGGTTATCCTGGAAGCGTCTCT
GCTGGTTGAAAAAGGTACCGACAAAAACGTTGACAAACTGATCGTTGTTGACGCGCCGGAAGAAGTTTGCAAAGAACGTGCGATCA
AACGTGGTATGTCTGAAGAAGACGCGGAACGTCGTATCAAAAAACAGATGCCGATCGAAGAAAAAGTTAAACTGGCGGACGTTGTT
ATCGACAACTCTGGTTCTATCGAAGAAACCCGTAAACAGGTTAAAAAAGTTCTGGAAGAACTGACCCGTGACCCGCTGGAA-3′

Supplementary figure S2. Nucleotides sequences of the synthetic genes. *DPCK_WT* – gene encoding wild type dephosphocoenzyme A kinase from Aquifex aeolicus; *DPCK_LEU* – gene encoding mutant, in which aromatic amino acids were substituted by leucines; *DPCK_HSW* – gene encoding mutant, in which aromatic amino acids were substituted according to HotSpot Wizard 3.0.