

**Implications of molecular interactions for
protein structure, function and design**

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Markus Meier

aus Flaach (Zürich)

Basel, 2005

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von
PD Dr. Peter Burkhard, Prof. Dr. Olga Mayans, Prof. Dr. Ueli Aebi

Basel, den 08.06.2004

Prof. Dr. Marcel Tanner
Dekan

Table of contents

Table of contents

Acknowledgements

Summary

Chapter 1

Crystallization and preliminary X-ray diffraction analysis of the active core of human recombinant cystathionine β -synthase: an enzyme involved in vascular disease

Introduction

Material and methods

Cloning of truncated human CBS cDNA expression vector

Purification of recombinant human CBS Δ 414-551

Enzyme characterization

Results

References

Chapter 2

Structure of human cystathionine β -synthase: a unique pyridoxal 5'-phosphate-dependent heme protein

Introduction

Results and Discussion

Description of the structure

The active site

The heme binding site

Oxidoreductase active site motif

The regulatory domain

Structural characterization of selected CBS mutations

Conclusions and outlook

Material and Methods

Crystallisation and data collection

X-ray structure determination

Coordinates

Acknowledgements

References

Corrigendum

Chapter 3

Structural insights into mutations of cystathionine β -synthase

Introduction

Results and discussion

Mutations in the dimer interface

Mutations in the active site

Mutations in the heme binding site

Other mutations

Acknowledgements

References

Chapter 4

Introduction to coiled coils

1. Geometry of coiled coils
2. Assignment of the heptad core positions
 - 2.1. *SOCKET* algorithm
 - 2.2. *TWISTER* algorithm
3. Factors determining the stability of coiled coils
 - 3.3. Intrahelical interactions
 - 3.4. Interhelical interactions
4. Conclusion
5. Acknowledgements
6. References

Chapter 5

Design of a minimal protein oligomerization domain by a structural approach

Introduction

Results and discussion

Factors contributing to monomeric α -helix stability

Factors contributing to oligomeric α -helix stability

Biophysical properties of the peptide

Structure of the peptide

Conclusions

Material and methods

Synthetic peptides

Analytical ultracentrifugation

CD spectroscopy

Crystallization

Data collection, processing, and phasing

Model building and refinement

Accession number

Acknowledgements

References

Chapter 6

Removing an Interhelical Salt Bridge Abolishes Coiled-Coil Formation in a *de Novo* Designed Peptide

Introduction

Materials and methods

Analytical ultracentrifugation

CD spectroscopy

Crystallization

Data collection and processing

Structure determination and refinement

Results and discussion

Biophysical characterization of the peptide

Structure of the peptide

References

Chapter 7

Statistics of intrahelical salt bridges in coiled coils

1. Introduction

2. Materials and methods

2.5. Limitations of the *SOCKET* program

2.6. Limitations of *SBSCC*

2.7. Differences between *SOCKET* and *SBSCC*

2.8. Statistical methods

2.10. Scaling of amino acid frequencies

2.11. Preparation of the reference data set

2.12. Preparation of the coiled-coil datasets

3. Definitions

4. Results

4.1. Size of the datasets

4.2. Average coiled-coil length

4.3. Residues involved in ionic interactions

4.4. Exposure of the coiled coils to solvent

4.5. Intrahelical salt bridges

5. Discussion

5.1. Sequence bias in the coiled-coil datasets

5.2. Effect of solvent exposure

5.3. Helix dipole

5.4. Importance of side chain conformations for the formation of salt bridges

5.5. Probability of salt-bridge formation of the 3DR configuration

5.6. Comparison with earlier results

5.7. Dissecting the importance of intrahelical ionic interactions for helix stabilisation

5.8. Consequences for coiled-coil design

6. Conclusion

7. Future work

8. Acknowledgements

9. References

Curriculum vitae

List of publications

Acknowledgements

I would like to thank Peter Burkhard for his outstanding scientific and personal support and for being an excellent boss.

I am grateful to all people in the Jan Kraus' research group, especially to Jan Kraus, Jana Oliveriusova, Miroslav Janosik and Nina Frank for their fruitful collaboration.

Special thanks go to Sergei Strelkov and Jörg Stetefeld for many helpful discussions and advice.

I would like to thank all present and former members of Peter Burkhard's research group for their constant support and encouragement, especially Alexandra Meli, Arundhati Chattopadhyay and Senthil Kumar. The same gratitude goes to the people in the groups of Olga Mayans and Tilman Schirmer.

Then, I would like to thank Astrid Arion for her friendship and understanding.

Finally, I would like to thank my parents for all their support and love. Without them this project would not have been feasible.

Summary

Protein structures are kept in a delicate balance of stability by the interactions of the amino acid residues among themselves, the solvent and other molecules. On one hand they must be stable enough not to unfold, on the other hand they must be mobile enough to undergo structural changes if necessary. Only thus they are able to fulfil their various functions in living organisms, e.g. the catalytic function of an enzyme, protein-ligand recognition or the rapid reorganisation of the cytoskeleton.

In this work, we have investigated the contributions of such molecular interactions to protein structure in a functional enzyme, cystathionine β -synthase. We have further analysed the contribution of ionic interactions to the stability of various designed peptides which form coiled coils. Finally, we have collected a statistics of electrostatic interactions in naturally occurring coiled coils to find out which ionic interactions significantly contribute to a successful formation of coiled coils. The results have important implications for the design of coiled-coil proteins.

Cystathionine β -synthase is an enzyme of the transsulfuration pathway in eukaryotic cells which catalyses the condensation of serine and homocysteine to yield cystathionine in a pyridoxal 5'-phosphate-dependent β -replacement reaction. The human enzyme also contains heme as a second cofactor which is not required for catalysis. We have solved the structure of the catalytic domain of human cystathionine β -synthase by X-ray crystallography. This is the first protein structure containing a heme binding motif where the iron of the heme is coordinated by a histidine and a cysteine residue. We have also discovered an oxidoreductase active site motif on the surface which might play a role in enzyme regulation. There are more than 100 point mutations known in this enzyme which can cause homocysteinuria disease in humans, characterised by dislocated eye lenses, skeletal problems, vascular disease and mental retardation. We have mapped the mutations in the catalytic domain on the structure and were able to find an explanation for the harmful effect of some mutations by analysing the molecular interactions of the concerned residues.

Coiled coils are a simple and regular structural motif in proteins consisting of α -helices which coil around each other. They can form dimers, trimers, tetramers and pentamers depending on their amino acid sequence and the environment. The principles and factors

which lead to this specific fold can therefore be studied in detail. The stability of coiled coils is mainly achieved by the systematic packing of the side chains of the residues at the interface between the helices, called knobs-into-holes packing. We could show, however, that a complex network of inter- and intrahelical salt bridges also contributes significantly to coiled-coil stability by designing short peptides which form dimeric or trimeric coiled coils. The importance of the ionic interactions could be demonstrated by removing a single interhelical salt bridge which abolished the formation of the coiled coil. The peptides were characterised by circular dichroism, analytical ultracentrifugation and X-ray crystallography.

We have developed the computer program *SBSCC* to collect a statistics of intrahelical salt bridges in pure α -helices and coiled coils from the protein database. We have identified the salt-bridge configurations that have the highest probability to form the ionic interaction and which occur most frequently in α -helices and coiled coils. We have found interesting differences between α -helices, parallel and antiparallel 2-stranded coiled-coils with important implications for the coiled-coil design. We have found a positive correlation between the probabilities of different salt-bridge configurations to form the ionic interaction and their frequencies in α -helices and coiled coils. This indicates that nature relies indeed on ionic interactions to stabilise α -helices and coiled coils, an issue which was hitherto controversially discussed.

Chapter 1

Crystallization and preliminary X-ray diffraction analysis of the active core of human recombinant cystathionine β -synthase: an enzyme involved in vascular disease

Janosik, M., Meier, M., Kery, V., Oliveriusova, J., Burkhard, P. and Kraus, J.P. (2001) Crystallization and preliminary X-ray diffraction analysis of the active core of human recombinant cystathionine beta-synthase: an enzyme involved in vascular disease. *Acta crystallographica Sect D*, **57**, 289-291.

Initial conditions for the crystallisation of the Δ 414-551 CBS with additional 23 *N*-terminal amino acid residues were found by P. Burkhard. I refined these conditions which finally lead to crystals suitable for diffraction. The expression and purification of the CBS mutant described in this chapter was done by M. Janosik in the lab of Jan Kraus.

Crystallization and preliminary X-ray diffraction analysis of the active core of human recombinant cystathionine β -synthase: an enzyme involved in vascular disease

Miroslav Janosik,^a Markus Meier,^b Vladimir Kery,^a Jana Oliveriusova,^a Peter Burkhard^{b*} and Jan P. Kraus^a

^aDepartments of Pediatrics and Cellular and Structural Biology, University of Colorado School of Medicine, Denver, CO 80262, USA, and ^bM. E. Müller-Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056, Basel, Switzerland

Correspondence e-mail:
peter.burkhard@unibas.ch

Cystathionine β -synthase (CBS) is a unique heme enzyme that catalyzes a PLP-dependent condensation of serine and homocysteine to give cystathionine. Deficiency of CBS leads to homocystinuria, an autosomal recessively inherited disease of sulfur metabolism. A truncated form of CBS in which the C-terminal amino-acid residues have been deleted has been prepared. The truncated CBS subunits form a dimer, in contrast to the full-length subunits which form tetramers and higher oligomers. The truncated CBS yielded crystals diffracting to 2.6 Å which belong to space group $P3_1$ or $P3_2$. This is the first comprehensive structural investigation of a PLP and heme-containing enzyme.

Received 21 July 2000

Accepted 17 November 2000

1. Introduction

Cystathionine β -synthase (CBS, L-serine hydrolyase; E.C. 4.2.1.22) is the first enzyme of the transsulfuration pathway in which the toxic homocysteine is converted to cysteine. A deficiency of CBS activity is the most common cause of homocystinuria, an inherited metabolic disease characterized by dislocated eye lenses, skeletal problems, vascular disease and mental retardation (Mudd *et al.*, 1995). There have been now over 100 mutations described in this gene (Kraus *et al.*, 1999). Hyperhomocysteinemia, a condition characterized by small increases in plasma concentrations of homocysteine, represents an independent risk for vascular disease.

The human CBS is a homotetramer consisting of 63 kDa subunits (Skovby *et al.*, 1984). Each CBS subunit of 551 amino-acid residues binds two substrates (homocysteine and serine) and three additional ligands: pyridoxal 5'-phosphate (PLP), *S*-adenosylmethionine (AdoMet) and heme (Kery *et al.*, 1994). While the role of heme in CBS is unknown, catalysis by CBS can be explained solely by participation of PLP in the reaction mechanism (Kery *et al.*, 1999). In fact, yeast CBS catalyzes the same reaction and does not contain heme (Jhee *et al.*, 2000). Limited proteolysis of the full-length enzyme yields the 'active core' of CBS (amino-acid residues 40–413). The reduction in size is accompanied by a significant increase in the specific activity of the enzyme and a change from a tetramer to a dimer. The dimer is about twice as active as the tetramer. It binds both PLP and heme cofactors, but is no longer activated by AdoMet (Kery *et al.*, 1998). This 45 kDa active core is the portion of CBS most homologous with the evolutionarily related enzymes isolated from

plants or bacteria (Swaroop *et al.*, 1992; Kraus, 1994).

2. Material and methods

2.1. Cloning of truncated human CBS cDNA expression vector

In order to simplify expression and subsequent purification of the human CBS protein, we generated a human CBS expression construct that allowed us to express truncated human CBS in *Escherichia coli* as a fusion protein with glutathione *S*-transferase (GST). This fusion protein could subsequently be cleaved with Xa protease to generate human CBS active-core protein. This construct, designated pGEX-5X-1hCBS Δ 414–551, was prepared by a three-piece ligation. Firstly, two primers were designed: 337, sense 5'-GAT-CCCGAGCGAAACCCCGCAGGCGGA-AGTGGGGCC-3', and 338, antisense 5'-CCA-CTTCCGCTGCGGGGTTTCGCTCGG-3'. These two primers encode the first ten amino acids of human CBS and respect the codon preference of *E. coli*. Under favorable conditions, primers 337 and 338 will hybridize together to form a short double-stranded insert with sticky overhangs, which enables cloning of this insert as a *Bam*HI/*Apa*I fragment (Kraus *et al.*, 1988). Secondly, a portion of the human CBS cDNA coding for amino acids 11–413 was cut out of the previously described pAX5-HCBS construct (Bukovska *et al.*, 1994) using *Apa*I and *Sal*I restriction endonucleases. Thirdly, both annealed primers 337/338 and the gel-purified *Apa*I/*Sal*I fragment were cloned into *Bam*HI and *Sal*I sites of the pGEX-5X-1 expression vector (Pharmacia). Three of the *Bam*HI/*Apa*I cassettes were cloned into the final construct so that the CBS polypeptide

Table 1
Data-collection statistics.

	MAD	Native
Resolution (Å)	3.15	2.6
Unique reflections	83338	63997
Multiplicity	4.5	2.7
Completeness (%)		
Overall	94.4 (50.00–3.15 Å)	82.4 (50.00–2.60 Å)
Outermost shell	67.1 (3.26–3.15 Å)	63.9 (2.69–2.60 Å)
R_{sym} (%)		
Overall	12.2 (50.00–3.15 Å)	7.6 (50.00–2.60 Å)
Outermost shell	38.5 (3.26–3.15 Å)	30.9 (2.69–2.60 Å)

contains an extra 23 amino acids at its N-terminus. If only two residues separated the Xa cleavage site and the N-terminus of CBS, the fusion protein had no activity and did not bind to the glutathione-Sepharose column. The vector-insert junctions and the entire coding sequence were verified by DNA sequencing using the Thermo Sequenase Cy5.5 dye terminator cycle-sequencing kit and the Long-Read Tower System V 3.1 automated DNA sequencer (Visible Genetics, Toronto, Ontario, Canada).

2.2. Purification of recombinant human CBS Δ 414–551

E. coli XL1-Blue MRF' bearing the pGEX-5X-1-hCBS Δ 414–551 construct were grown in 81 batch cultures of NZCYM medium supplemented with 0.1 mg ml⁻¹ ampicillin, 0.3 mM δ -aminolevalinate (δ -ALA) and 0.001% (w/v) thiamine. Cells were grown to an OD₅₅₀ of ~0.4 and then induced with 0.1 mM IPTG for 14 h. After the incubation period, the cells were chilled on ice and all subsequent steps were performed at 277 K using pre-chilled solutions at the same temperature. Cells were collected by centrifugation and washed twice with 1×PBS buffer. Cell lysis was performed by osmotic shock according to a previously described method (Witholt *et al.*, 1976). After a 1 h incubation on ice and a 5 min heat shock at 310 K, crude extract fractions were separated by ultra-

centrifugation (96 000g at 283 K for 1 h). The crude extract was saturated with PLP (1 mM final concentration), diluted 1:1 with the washing buffer (1×PBS containing 0.3 M NaCl and 1 mM DTT). This fraction was loaded onto a GST Sepharose 4B column (BioRad, 2 × 15 cm) equilibrated with 10 bed volumes of the washing buffer. The column was washed with 8–10 volumes of the washing buffer. The GST Sepharose 4B resin with the bound GST-CBS fusion protein was transferred into a 50 ml Falcon tube and the buffer was changed to 1×TBS with 1 mM DTT and 2 mM CaCl₂. The GST–CBS fusion protein was cleaved by digestion with Xa protease, which was used at a final concentration of 7.5 μ g per milligram of fusion protein. The digest was performed at 277 K overnight. The mixture was centrifuged at 5000g for 5 min at 277 K. To further purify the cleaved CBS from any undigested product and from GST, we applied the supernatant to a secondary affinity column (GST Sepharose 4B) equilibrated with 1×PBS. The flowthrough fraction containing CBS was collected and concentrated using an Amicon 50 or Centricon 50-spin column. Prior to crystallization, the enzyme was stored in 20 mM HEPES pH 7.4 at 188 K.

2.3. Enzyme characterization

All enzyme preparations were further characterized for protein concentration (Lowry *et al.*, 1951), CBS specific activity (Kraus, 1987) and cofactor saturation. The PLP saturation of the purified enzyme was determined by a previously published fluorimetric method (Adams, 1979). The heme saturation was established using the pyridine–hemochromogen method (Morrison & Horie, 1965). For crystallization purposes we used only preparations with a PLP and heme content over 90%.

3. Results

The CBS tetramer has a strong tendency to aggregate, making physical studies of the enzyme very difficult. We have recently expressed and purified to near-homogeneity recombinant human CBS comprising amino-acid residues 1–413. This enzyme, missing ~140 C-terminal residues and similar to the abovementioned active core, forms dimers and does not exhibit the aggregating properties of the full-length enzyme. Using this enzyme, we were able to obtain the first diffracting crystals (Fig. 1).

The recombinant CBS enzyme was purified as described above and concentrated to 26 mg ml⁻¹ in 20 mM HEPES pH 7.4. Small crystals of CBS were obtained in sparse-matrix screens (Hampton Research, Laguna Hills, CA; Jancarik & Kim, 1991) using the vapor-diffusion hanging-drop method (McPherson, 1982) with 4 μ l drops consisting of 2 μ l protein and 2 μ l mother liquor equilibrated against 1 ml of reservoir solution at room temperature. Subsequent optimization resulted in the following crystallization procedure: CBS was mixed with an equal volume of reservoir solution containing 30% PEG 1000, 80 mM HEPES pH 7.5 and 0.4 mM FeCl₃. Crystals appeared

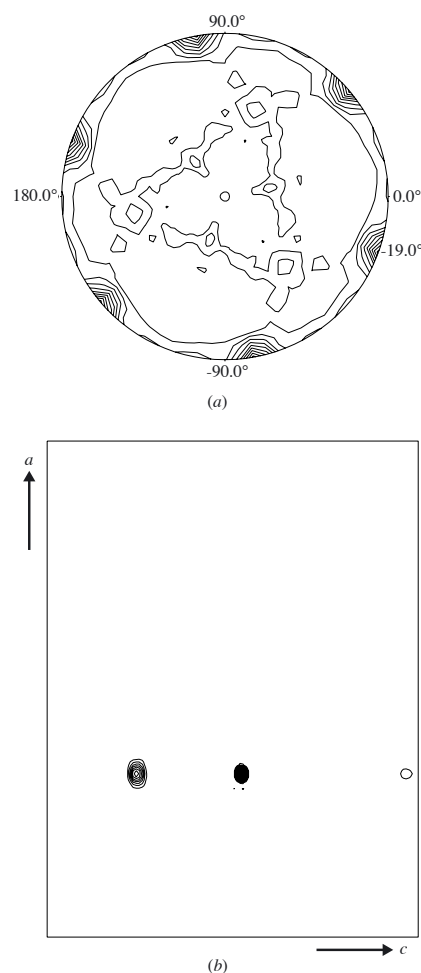


Figure 2
(a) Plot of the self-rotation function of cystathionine β -synthase at $\kappa = 180^\circ$. The orientation of the dimer twofold axes are perpendicular to the crystallographic threefold axis indicating that all dimers in the asymmetric unit have the same orientation. (b) Plot of the section $b = 1/3$ of the native Patterson function of cystathionine β -synthase. The fact that there is more than one native Patterson peak indicates that there are more than two molecules with the same orientation in the crystal packing.

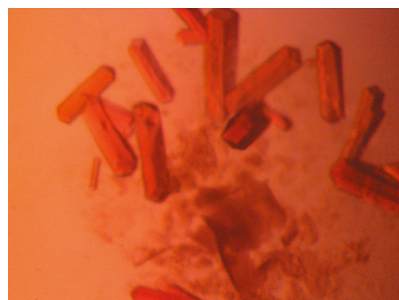


Figure 1
Crystals of human cystathionine β -synthase.

within 7 d and grew to a final size of $\sim 0.5 \times 0.2 \times 0.2$ mm after 30 d. The addition of FeCl_3 helped to prevent the formation of small satellite crystals on the surface of the crystals.

The X-ray diffraction data of the native crystal were collected at beamline BM1A of the SNBL at the ESRF in Grenoble at a temperature of 100 K. Furthermore, a MAD data set at three wavelengths around the absorption edge of the iron ion ($\lambda = 1.74$ Å) was collected at the BW7A beamline (EMBL, DESY Hamburg). All data were processed and integrated using *DENZO* and scaled using *SCALEPACK* (Otwinowski & Minor, 1997; Table 1). The crystals belong to a trigonal space group, with unit-cell parameters $a = b = 144.46$, $c = 108.21$ Å; systematic absences ($00l$ except for $l = 3n$) indicate that the space group is either $P3_1$ or $P3_2$. Given the dimeric nature of the truncated CBS enzyme, the crystals contain either two or three dimers per asymmetric unit, corresponding to a solvent content of 64 or 46% and a calculated Matthews volume V_M of 3.4 or 2.3 Å³ Da⁻¹, respectively, with both values being within the range typically observed in protein crystals (Matthews, 1968).

Analysis of the self-rotation function (Fig. 2*a*) showed strong peaks at $\kappa = 180^\circ$ corresponding to an orientation of the dimer twofold axis perpendicular to the crystallographic threefold axis. The presence of these strong peaks indicates that all dimers in the asymmetric unit have the same orientation. This is confirmed by the native Patterson

function (Fig. 2*b*), which shows two strong peaks. The fact that there is more than one native Patterson peak indicates that there are more than two molecules with the same orientation in the crystal packing. Hence, the crystal contains three dimers in the asymmetric unit, all having the same orientation, namely, with their twofold axis being perpendicular to the threefold crystallographic axis. The quality of the anomalous data sets is rather low (high R_{sym}) owing to the high mosaicity of the crystals (Table 1) and the anomalous Patterson map does not show any peaks. Using the preliminary phase information obtained from molecular replacement with the structure of *O*-acetylserine sulphydrylase from *Salmonella typhimurium* (Burkhard *et al.*, 1998), the Fe atoms could unambiguously be identified. Combining the phase information obtained from molecular replacement and MAD significantly improved the quality of the electron density (M. Meier, manuscript in preparation).

References

- Adams, E. (1979). *Methods Enzymol.* **62**, 407–410.
- Bukovska, G., Kery, V. & Kraus, J. P. (1994). *Protein Expr. Purif.* **5**, 442–448.
- Burkhard, P., Jagannatha Rao, G. S., Hohenester, E., Schnackerz, K. D., Cook, P. F & Jansonius, J. N. (1998). *J. Mol. Biol.* **283**, 121–133.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Jhee, K. H., McPhie, P. & Miles, E. W. (2000). *J. Biol. Chem.* **275**, 11541–4.
- Kery, V., Bukovska, G. & Kraus, J. P. (1994). *J. Biol. Chem.* **269**, 25283–25288.
- Kery, V., Poneleit, L. & Kraus, J. P. (1998). *Arch. Biochem. Biophys.* **355**, 222–232.
- Kery, V., Poneleit, L., Meyer, J. D., Manning, M. C. & Kraus, J. P. (1999). *Biochemistry*, **38**, 2716–2724.
- Kraus, J. P. (1987). *Methods Enzymol.* **143**, 388–394.
- Kraus, J. P. (1994). *J. Inher. Metab. Dis.* **17**, 383–390.
- Kraus, J. P., Janosik, M., Kozich, V., Mandell, R., Shih, V., Sperandio, M. P., Sebastio, G., de Franchis, R., Andria, G., Kluijtmans, L. A., Blom, H., Boers, G. H., Gordon, R. B., Kamoun, P., Tsai, M. Y., Kruger, W. D., Koch, H. G., Ohura, T. & Gaustadnes, M. (1999). *Hum. Mutat.* **13**, 362–375.
- Kraus, J. P., Novotny, J., Kalousek, F., Swaroop, M. & Rosenberg, L. E. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 8905–8909.
- Lowry, O. H., Rosenbrough, M. J., Farr, A. L. & Randall, R. J. (1951). *J. Biol. Chem.* **193**, 265–275.
- McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*, pp. 94–96. New York: John Wiley & Sons.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Morrison, M. & Horie, S. (1965). *Anal. Biochem.* **12**, 77–82.
- Mudd, S. H., Levy, H. L. & Skovby, F. (1995). In *The Metabolic and Molecular Bases of Inherited Disease*, edited by C. R. Scriver, A. Beaudet, W. S. Sly & D. Valle, pp. 1279–1327. New York: McGraw-Hill Inc.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **277**, 307–325.
- Skovby, F., Kraus, J. P. & Rosenberg, L. E. (1984). *J. Biol. Chem.* **259**, 583–587.
- Swaroop, M., Bradley, K., Ohura, T., Tahara, T., Roper, M. D., Rosenberg, L. E. & Kraus, J. P. (1992). *J. Biol. Chem.* **267**, 11455–11461.
- Witholt, B., Boekhout, M., Brock, M., Kingma, J., Heerikhuizen, H. V. & Leij, L. D. (1976). *Anal. Biochem.* **74**, 160–170.

Chapter 2

Structure of human cystathionine β -synthase: a unique pyridoxal 5'-phosphate-dependent heme protein

Meier, M., Janosik, M., Kery, V., Kraus, J.P. and Burkhard, P. (2001) Structure of human cystathionine beta-synthase: a unique pyridoxal 5'- phosphate-dependent heme protein. *Embo J*, **20**, 3910-3916.

The expression and purification of the CBS mutant described in this chapter was done by M. Janosik in the lab of Jan Kraus. Data collection at the *ESRF* and data processing were performed by P. Burkhard. All other crystallographic work and the interpretation of results is my responsibility.

Structure of human cystathionine β -synthase: a unique pyridoxal 5'-phosphate-dependent heme protein

Markus Meier, Miroslav Janosik¹,
Vladimir Kery¹, Jan P. Kraus¹ and
Peter Burkhard²

M.E.Müller Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland and

¹Departments of Pediatrics and Cellular and Structural Biology, University of Colorado School of Medicine, Denver, CO 80262, USA

²Corresponding author

e-mail: Peter.Burkhard@unibas.ch

Cystathionine β -synthase (CBS) is a unique heme-containing enzyme that catalyzes a pyridoxal 5'-phosphate (PLP)-dependent condensation of serine and homocysteine to give cystathionine. Deficiency of CBS leads to homocystinuria, an inherited disease of sulfur metabolism characterized by increased levels of the toxic metabolite homocysteine. Here we present the X-ray crystal structure of a truncated form of the enzyme. CBS shares the same fold with *O*-acetylserine sulfhydrylase but it contains an additional N-terminal heme binding site. This heme binding motif together with a spatially adjacent oxidoreductase active site motif could explain the regulation of its enzyme activity by redox changes.

Keywords: cystathionine β -synthase/cysteine biosynthesis/heme protein/pyridoxal 5'-phosphate/X-ray crystal structure

Introduction

Cystathionine β -synthase (CBS, L-serine hydrolyase, EC 4.2.1.22) is the first enzyme of the transsulfuration pathway in which the potentially toxic homocysteine is converted to cysteine (Figure 1). Deficiency of CBS activity is the most common cause of homocystinuria, an inherited metabolic disease characterized by dislocated eye lenses, skeletal problems, vascular disease and mental retardation (Mudd *et al.*, 2001). There have now been >100 mutations described in this gene (Kraus *et al.*, 1999). Hyperhomocysteinemia, a condition characterized by small increases in plasma concentrations of homocysteine, represents an independent risk for vascular disease (Yap *et al.*, 2000).

The human CBS is a homotetramer consisting of 63 kDa subunits, which binds two cofactors, pyridoxal 5'-phosphate (PLP) and heme (Skovby *et al.*, 1984; Kery *et al.*, 1994). Each CBS subunit of 551 amino acid residues binds two substrates (homocysteine and serine) and is further regulated by *S*-adenosyl-L-methionine (AdoMet) (Kery *et al.*, 1994). While the role of heme in CBS is unknown, catalysis by CBS can be explained solely by participation of PLP in the reaction mechanism (Kery *et al.*, 1999). In fact, yeast CBS catalyzes the same reaction but does not

contain heme (Jhee *et al.*, 2000; Maclean *et al.*, 2000). Limited proteolysis of the full-length enzyme yields the 'active core' of CBS (amino acid residues 40–413). The reduction in size is accompanied by a significant increase in the specific activity of the enzyme and a change in its oligomerization state. The active core enzyme is about twice as active as the full-length enzyme and forms dimers instead of tetramers. It binds both PLP and heme cofactors, but is no longer activated by AdoMet (Kery *et al.*, 1998). This 45 kDa active core is the portion of CBS most homologous with the related enzymes in plants and bacteria, *O*-acetylserine sulfhydrylase (OASS) and *O*-acetyl-L-serine(thiol)lyase (OASTL) (Swaroop *et al.*, 1992; Kraus, 1994). The C-terminal regulatory domain that is missing in the active core enzyme contains a recently identified protein folding motif called 'CBS domain', which is also found in inosine 5'-monophosphate dehydrogenase, chloride channels and several other proteins in various organisms (Bateman, 1997).

Since the CBS tetramer of the full-length enzyme has a strong tendency to aggregate, physical studies are very difficult. We have recently crystallized recombinant human CBS comprising the amino acid residues 1–413 (Janosik *et al.*, 2001). This truncated enzyme is similar to the above-mentioned active core in that the ~140 C-terminal residues including the 'CBS domain' are missing. It is about twice as active as the wild-type CBS, forms dimers, and does not exhibit the aggregating properties of the full-length enzyme. We have now solved the X-ray structure of the truncated form of CBS by combining phase information obtained from molecular replacement (MR) and multiple anomalous dispersion (MAD). The crystals belong to the trigonal space group $P3_1$ with cell parameters $a = b = 144.46 \text{ \AA}$, $c = 108.21 \text{ \AA}$, and contain three dimers per asymmetric unit (Table I).

Results and discussion

Description of the structure

The fold of the truncated human CBS enzyme belongs to the β -family of vitamin B₆ enzymes (Alexander *et al.*, 1994) and resembles the fold of OASS from *Salmonella typhimurium* (Figure 2) (Burkhard *et al.*, 1998). Three other PLP enzymes with known structure share the same fold type: tryptophan synthase (Hyde *et al.*, 1988), threonine deaminase (TD) (Gallagher *et al.*, 1998) and aminocyclopropane deaminase (Yao *et al.*, 2000). A least square superposition of the C α -positions of the structurally conserved parts between CBS and OASS yields a root mean square deviation (r.m.s.d.) of only 1.32 Å, while the differences between the structures are mainly located in the loop regions (residues 95–104, 282–298 and 359–369). The monomer is composed of 11 α -helices, seven short 3₁₀ helices and two β -sheets consisting of four (in the

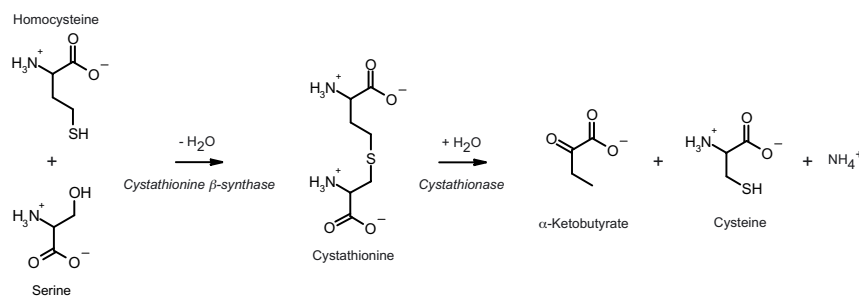


Fig. 1. Transsulfuration pathway.

Table I. Data statistics

Data set	Native	λ_1	λ_2	λ_3
Space group	$P3_1$		$P3_1$	
a, b (Å)	144.46		144.52	
c (Å)	108.21		108.16	
Wavelength	0.9711	1.7411	1.7424	1.7300
Resolution range (Å)	50–2.6	50–3.1	50–3.15	50–3.15
Unique reflections	63 997	83 338	82 233	84 316
Overall completeness (outermost shell) (%)	82.4 (63.9)	94.4 (67.1)	93.8 (60.1)	91.3 (68.6)
R_{merge}^a (outermost shell)	7.6 (30.9)	12.2 (38.5)	15.0 (44.3)	16.9 (48.7)
Phasing power ^b (acentric)	0.585	1.045	0.716	0.633
FOM ^c			0.24658	
R factor ^d (R_{free})	25.7 (29.6)			
R.m.s.d. from ideality				
bonds (Å)	0.008			
angles (°)	1.42			
Average B factors (Å ²)				
protein	37.6			
solvent	24.1			
PLP	17.4			
heme	34.1			

$$^a R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i |I_i - \langle I \rangle|}{\sum_{\text{hkl}} \sum_i I_i}$$

$$^b \text{Phasing power} = (\text{FH}/\text{lack of closure}).$$

$$^c \text{FOM} = \text{figure of merit} [(\cos\phi)^2 + (\sin\phi)^2]^{1/2}$$

$$^d R \text{ factor} = \frac{\sum_{\text{hkl}} |F_{\text{obs}} - F_{\text{calc}}|}{\sum_{\text{hkl}} F_{\text{obs}}}$$

N-terminal domain) and six strands (in the C-terminal domain), respectively (Figure 2A). The additional β -strand 1 interacts with strand 2 of the C-terminal sheet of the other monomer in the dimer in a parallel manner. The α -helices 1 and 11, β -strand 1 and the N-terminal heme binding site of CBS are missing in OASS, while in turn OASS has an additional α -helix that is not present in CBS (helix 8 in Burkhard *et al.*, 1998). The heme binding motif lacks any secondary structure with the exception of a short 3_{10} helix. It has been shown that OASS, and to a lesser extent tryptophan synthase, undergoes a large conformational change from the open, uncomplexed to the closed, complexed conformation upon ligand binding. This conformational change involves a rigid body rotation of the so-called moveable domain (Schneider *et al.*, 1998; Burkhard *et al.*, 1999, 2000). The C_{α} -backbone of CBS in this region (residues 186–222) more closely resembles the open conformation of OASS.

The dimer interface is mainly hydrophobic in character and is composed of the side chains of the residues Ile76, Leu77, Ile80, Thr87, Val90, Ile92, Ile152, Leu156, Val160, Val180, Ala183, Leu184, Ile339, Ala340, Leu344, Leu345, Val378, Met382 and Leu386. The central part of the dimer interface is formed by the residues Phe111 and Phe112 close to the 2-fold dimer axis,

thus Phe112 of monomer A interacts with Phe112 of monomer B and vice versa. But polar interactions also contribute to the dimer interactions. On the other hand, the guanidinium group of Arg379 is completely buried within the core of the dimer interface but is not involved in any polar interactions between monomers, but rather forms hydrogen bonds to Gly115 and Asn380 of the same monomer.

The active site

The coenzyme PLP is deeply buried in a cleft between the N- and C-terminal domains, and the active site is accessible only via a narrow channel. The cofactor is linked to the ϵ -amino group of Lys119 via a Schiff base linkage forming the so-called 'internal aldimine' (Figure 3A) (Christen and Metzler, 1985; Kery *et al.*, 1999). The nitrogen of the pyridine ring forms a hydrogen bond to the O_{γ} of Ser349 similar to the other enzymes of the β -family of PLP enzymes, OASS, tryptophan synthase and TD (Gallagher *et al.*, 1998). Another hydrogen bond is formed between the 3' hydroxyl group of PLP and the N_{82} of Asn149. This residue is coplanar with the pyridine ring and thus allows the expected ring tilt upon transaldimination. The phosphate binding loop is located between β -strand 8 and α -helix 8 and is composed of the residues

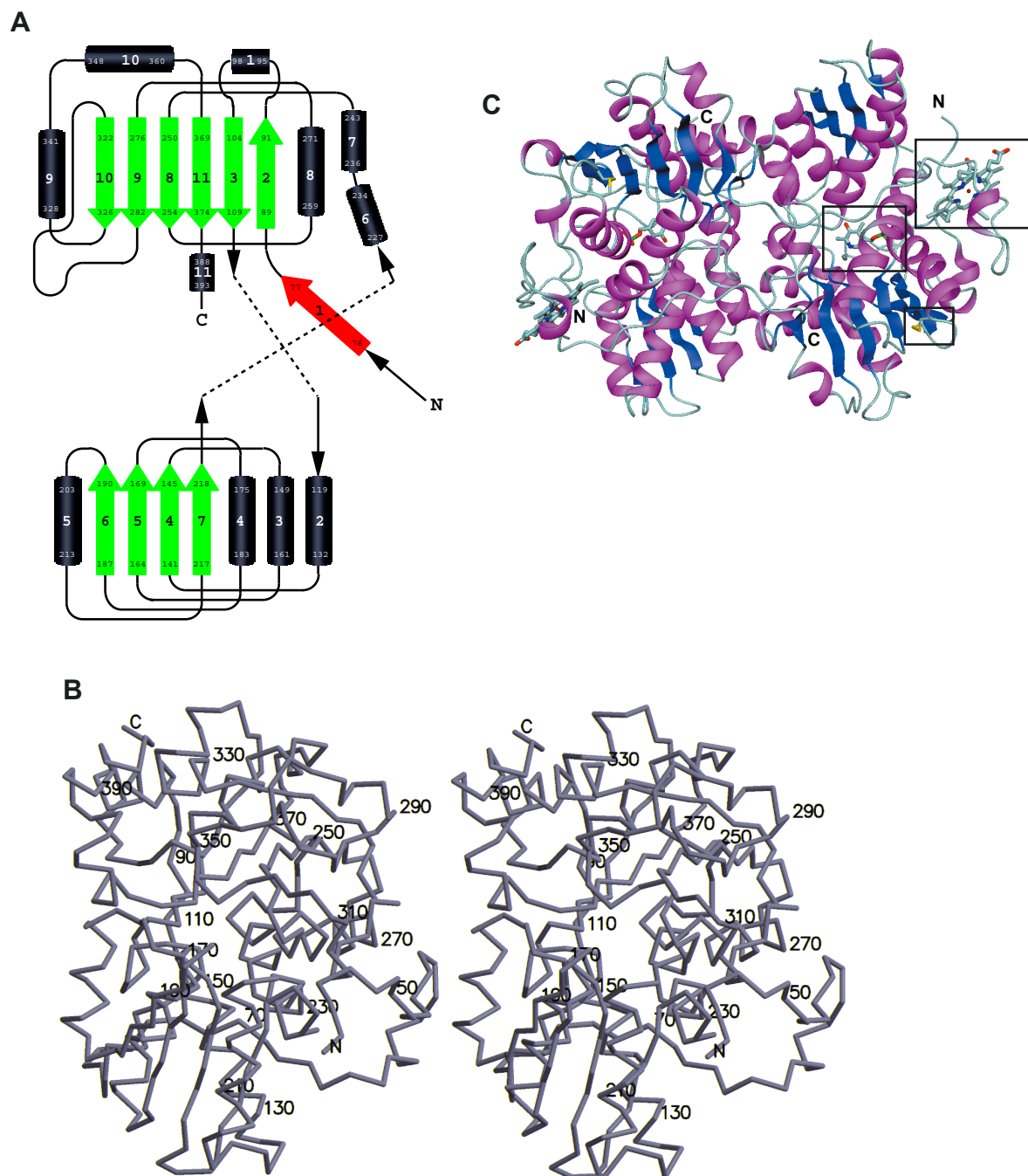


Fig. 2. Overall structure of truncated CBS. **(A)** Topology of the fold in CBS. Above, the C-terminal domain (with the first two strands of the β -sheet formed by the N-terminal residues); below, the N-terminal domain. Both domains are of the type α/β and contain a central β -sheet surrounded by several α -helices. Strand 1 (red) is part of the C-terminal β -sheet of the other monomer in the dimer. **(B)** Stereo drawing showing the overall fold of CBS with every twentieth residue labeled. **(C)** Schematic representation of the tertiary fold of a dimer of CBS. The central β -sheets are colored in blue and the surrounding α -helices are colored in magenta. In ball-and-stick representation and marked by a black box are the active site PLP, the heme and the oxidoreductase motif. The view is down the non-crystallographic 2-fold axis, which relates the two subunits of the dimer.

Gly256, Thr257, Gly258, Gly259 and Thr260. These residues form an extended hydrogen bonding network with the phosphate moiety of PLP, thus anchoring the cofactor to the protein matrix. In addition, the positive pole of the helix dipole from α -helix 8 compensates for the negative charge of the phosphate group.

The conformation of the residues surrounding the cofactor is highly conserved between CBS and OASS (Figure 3A). In the two monomers of all three dimers the

asparagine loop (residues 146–149; cf. Burkhard *et al.*, 1999) adopts two slightly different conformations, indicating its flexibility and ability to bind the carboxylate group of the substrate by a local conformational change. In OASS this conformational change includes atom movements of >7 Å upon substrate binding. Preliminary X-ray data suggest that in CBS the substrate serine binds to the active site in a similar way to the substrate analog methionine in OASS (data not shown). Residues Tyr223

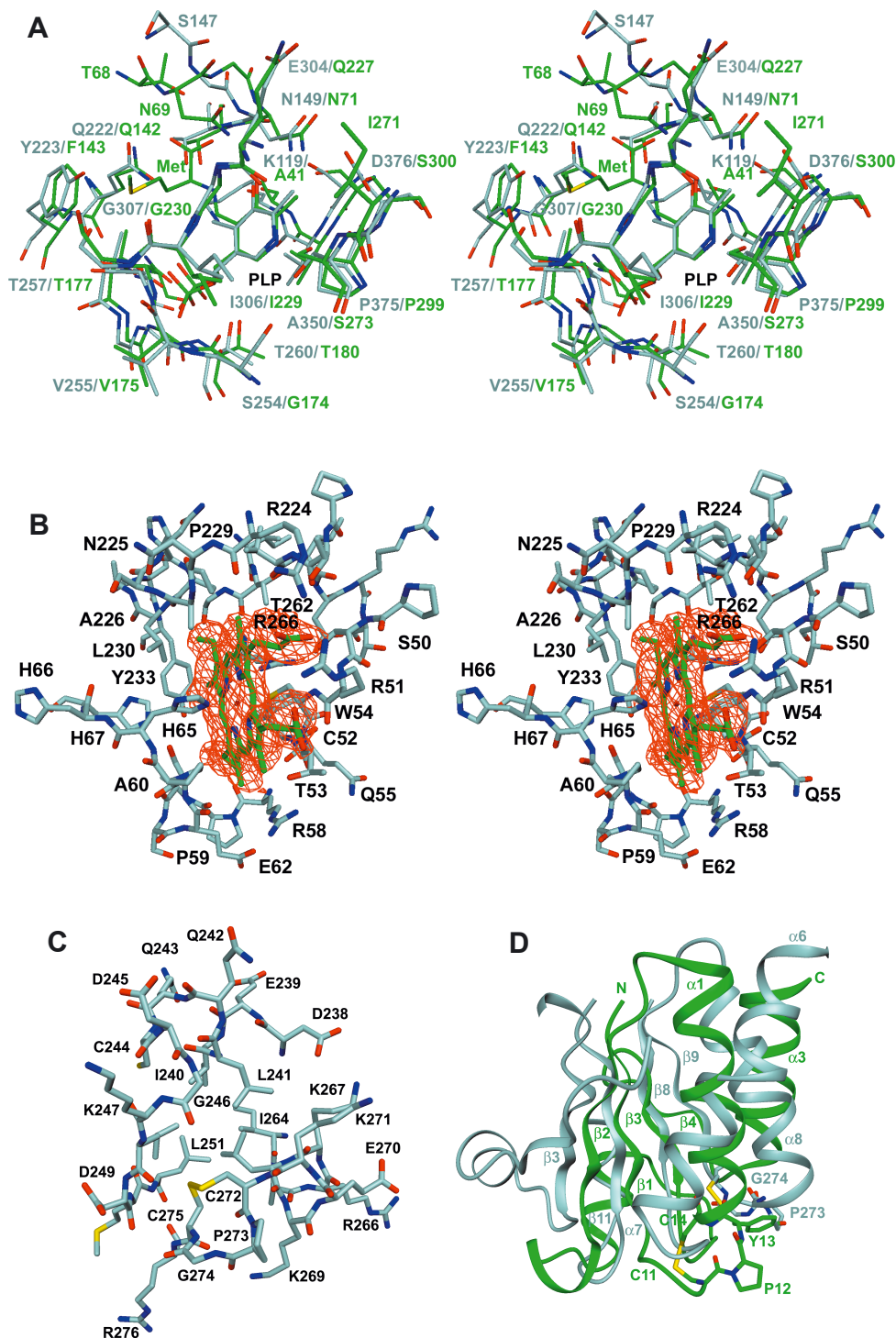


Fig. 3. Structural details of CBS. **(A)** The active site region of CBS (gray) in a superposition with the active site of OASS (green). The sequences as well as the structure of the two enzymes are very similar. A superposition of the 25 structurally most conserved residues yields an r.m.s.d. of 1.6 Å of their C α positions. The substrate analog of OASS methionine indicates the probable binding mode of the first substrate serine and also the region where the second substrate homocysteine is expected to bind. **(B)** The heme binding site of CBS with heme in green and the surrounding residues in gray. The two residues His65 and Cys52 are the ligands to the heme iron (dark red). The difference density for the cofactor is shown in red contoured at 3.5 σ . **(C)** The oxidoreductase motif in ball-and-stick representation and **(D)** in a superposition with the structure of glutaredoxin. The structure of CBS is in gray, the one of glutaredoxin in green. The overall topology is very similar, but the active site motif in CBS is switched to the other helix compared with glutaredoxin.

and Gly307 are probably the key residues for substrate specificity, as they are spatially adjacent to the substrate binding site.

The heme binding site

All dimers of the asymmetric unit contain two heme molecules that are located at distal ends of the dimers with

Table II. Sequence comparison

Human	248 LDMLVASVGTGGTITGIARKLKEK CPGC RIIIGVDPEGSILAEP 290
Rabbit	248 LDMLVASAGTGGTITGIARKLKEK CPGC QIIIGVDPEGSILAEP 290
Rat	245 VDMLVASAGTGGTITGIARKLKEK CPGC KIIIGVDPEGSILAEP 287
<i>Dictyostelium</i>	202 IDMIVCTAGTGGTITGIARKIKER LPNC IVVGVDPHGSILAQP 244
Yeast	188 LRAVVAGAGTGGTISGISKYLKEQ NDKI QIVGADPFSGSILAQP 230

the orientation of their ring planes normal to the protein surface (Figure 2C). The heme is bound in a hydrophobic pocket formed by residues 50–67, α -helices 6 and 8 and a loop preceding β -strand 10 (Figure 3B). The sulfhydryl group of Cys52 and the N_{ε2} atom of His65 axially coordinate the iron in the heme. This result confirms an earlier finding that the 5th and 6th coordination positions of the heme are a thiolate and a nitrogenous group, respectively (Omura *et al.*, 1984). Unlike the globins or cytochrome *c* peroxidase, the N_{δ1} of His65 is solvent accessible and lacking any hydrogen bonding partner from protein residues. The sulfur atom of Cys52 is deprotonized (Ojha *et al.*, 2000) and forms additional polar interactions with the side chain of Arg266 and the main chain nitrogen of Trp54. The heme carboxylate groups are involved in ionic interactions with Arg51 and Arg224 and are partially solvent accessible. This is in contrast to heme enzymes like cytochrome *c* peroxidase and cytochrome P450cam, where the heme is completely buried in an internal cavity of the protein (Poulos, 1987). Since the iron ion is ligated from both sides by protein residues this makes an enzymatic role of the heme unlikely. The situation is more similar to the *c*-type cytochromes, which are involved in electron transfer.

Since the heme is not covalently attached to the protein it can be reversibly released under reducing conditions from CBS crystals in the presence of carbon monoxide (CO) (Bruno *et al.*, 2001). Similarly, the heme can be fully dissociated from the reduced enzyme in CO-saturated solution. Under these conditions the enzyme retains ~20% of original activity. In contrast to the reversible removal of heme from crystals, the heme cannot be introduced back in solution (J.P.Kraus, unpublished results). Under oxidizing conditions, the heme cannot be released, probably because CO does not bind to heme in its ferric state. It is likely that CO displaces one of the axial heme ligands, followed by a local unfolding of the N-terminal residues leading to a release of the prosthetic group. The displaced ligand is probably the cysteine, because the absorption spectrum of CBS treated with CO is similar to the spectra of other CO–heme–imidazole protein complexes (Taoka *et al.*, 1999). It has been suggested that the redox state of the heme iron influences the catalytic rate in the full-length enzyme (Taoka *et al.*, 1998). We find that the redox state of the heme has no impact on the CBS activity (J.P.Kraus, unpublished results) and that the heme itself is not required for catalysis (Bruno *et al.*, 2001). Further evidence that the heme moiety is not involved in the catalytic steps comes from a CBS enzyme in which the first 70 residues were deleted, including the heme binding residues Cys52 and His65. This enzyme retains ~25% of wild-type CBS activity (J.P.Kraus, unpublished results).

Oxidoreductase active site motif

The loop between α -helix 8 and β -strand 9 harbors a motif similar to the active site of disulfide oxidoreductases. The consensus sequence of this motif contains two cysteines that are linked by two residues, one of which is a proline (Table II). The two cysteines of oxidoreductases are involved in various redox reactions in the cell, because they can be reversibly oxidized and/or reduced by switching between a disulfide and dithiol form during the catalytic process. In CBS this motif consists of the sequence CPGC (residues 272–275, Figure 3C) and forms a β -turn. The two cysteines are oxidized and form a disulfide bridge. The disulfide bridge is in the same right-handed hook conformation as those in disulfide oxidoreductases and is located on the surface of the protein and hence is solvent accessible. The same two cysteines, however, are not solvent accessible in the full-length enzyme (J.P.Kraus, unpublished results).

The structure of glutaredoxin from *Escherichia coli* can be roughly superimposed onto CBS (Figure 3D). In this superposition α -helix 1 of glutaredoxin fits onto α -helix 7 of CBS and α -helix 3 onto α -helix 8 of CBS, but in opposite directions. The four strands of the central β -sheet of glutaredoxin fit nicely onto strands 3, 8, 9 and 11 of the C-terminal domain of CBS. In this alignment the CXXC motifs are at similar positions but on the adjacent α -helices; in CBS the motif is found at a location corresponding to the N-terminal end of α -helix 3 of glutaredoxin, while in glutaredoxin it is located at the N-terminal end of helix 1. Thus, the motifs in the two proteins are in a similar environment with respect to their three-dimensional structure.

It is striking that CBS contains yet another motif that might be involved in redox reactions apart from the heme moiety. Moreover, as shown in Table II, this CPGC motif of CBS is also present in the sequences of the other mammalian CBS enzymes, which also contain the heme group, but is absent in the more distant species that are lacking the heme.

The regulatory domain

Full-length CBS contains a C-terminal regulatory domain of ~140 residues, including the so-called ‘CBS domain’ (Bateman, 1997) of 53 residues. The C-terminal domain of CBS contains an autoinhibitory region that gets displaced from the active site upon binding of the allosteric activator AdoMet (J.P.Kraus, unpublished results).

The catalytic domain of TD has the same fold as CBS and also contains a C-terminal regulatory domain. A superposition of both structures suggests that the regulatory domains of TD and CBS are located at similar positions. The surface of CBS corresponding to the interface between the catalytic and regulatory domains

of TD is largely hydrophobic, which is characteristic for most protein–protein interaction regions.

The fact that truncated CBS forms dimers rather than tetramers or higher order oligomers suggests that the regulatory domain is involved in tetramer formation. This idea is further supported by the fact that two CBS domains can associate e.g. as in the structure of inosine-5'-monophosphate dehydrogenase (Zhang *et al.*, 1999). In TD, however, the regulatory domain is not involved in tetramerization; thus, the mechanism leading to higher order oligomers appears to be different in the two proteins.

Structural characterization of selected CBS mutations

Presently there are >100 CBS mutations known that lead to more or less severe phenotypes in the patient. Five mutations are located close to or at the heme binding site (R58W, H65R, R224H, A226T and R266G/K), eight mutations affect the active site region of CBS and binding of the cofactor PLP (G148R, N228K, T257M, G259S, E302K, G305R, G307S and T353M) and six mutations are located at the dimer interface (P88S, A114V, G116R, I152M, E176K and V180A). Twelve mutated residues are on the surface of the protein and exposed to solvent. Most of these mutations are expected to affect enzyme activity by a general destabilization of the protein structure, which is probably the case for one of the most frequent mutations in patients, I278T. Together with the other frequent mutation, G307S, they represent ~40% of all mutant alleles. This second mutation (G307S), which confers a severe phenotype, probably influences binding of the second substrate homocysteine, as homocysteine is expected to bind to the protein in this region (Figure 3A). This is in agreement with the finding that patients with this mutation are not B6-responsive, because cofactor binding is not expected to be affected by this mutation.

Also the effects of some other mutations can be nicely explained by the crystal structure of CBS. Even though the mutation T257M is only found in a single patient, its effect on substrate and cofactor binding can easily be envisaged. Thr257 is hydrogen bonded to the phosphate group of the cofactor. The bulkier side chain of methionine would then occlude parts of the substrate binding pocket leading to a reduced enzyme activity. This mutation is not B6-responsive even though Thr257 is directly involved in cofactor binding. This is probably because an additional water molecule could easily replace the lost hydrogen bond to the phosphate group.

Because His65 is one of the two ligands of the heme iron, the mutation H65R would be expected to cause loss of heme binding. It has been shown that heme-free enzyme is still enzymatically active (Bruno *et al.*, 2001); however, this mutation shows a severe phenotype (L.S.Chen, unpublished). This is difficult to explain by its PLP-dependent catalytic activity alone and points to the importance of heme during folding of CBS (Kery *et al.*, 1994; Shan *et al.*, 2001). Furthermore, recent work has shown that CBS subunits aggregate and are found in inclusion bodies when CBS is expressed in the absence of heme (J.P.Kraus, unpublished results).

The mutation G148R is located in the asparagine loop, which undergoes a large conformational change upon ligand binding in OASS (Burkhard *et al.*, 1999). This loop

forms an extended hydrogen bonding network to the carboxylate moiety of the ligand in OASS and is very likely to do the same in CBS. The homologous residue in OASS is also a glycine and seems to be important for the flexibility of this loop. It has, in both the open and the closed conformation of OASS, ϕ and ψ angles that are only allowed for a glycine residue. Even though a mutation to arginine would not interfere sterically with the substrate itself, the reduced flexibility of the asparagine loop would most probably reduce ligand binding, because the hydrogen bonding network to the carboxylate moiety of the ligand cannot be properly established. It is interesting to note that among the mutations discussed above, A226T, R224H, I278T and T353M, but not G307S, are functionally suppressed in a truncated CBS missing the last ~140 residues. In addition, the common I278T mutation is suppressed by any of seven secondary missense mutations in the C-terminal domain of CBS (Shan *et al.*, 2001).

Conclusions and outlook

Taken together, the crystal structure of the truncated form of CBS gives the first insights into structural details of this unique PLP-dependent heme enzyme. Since deficiency of CBS leads to homocystinuria, an inherited metabolic disease, structural insight into its enzymatic mechanism and its regulation are crucial for a better understanding of its physiological role. The identification of the residues involved in heme binding explains some of the previously described mutations leading to homocystinuria. This heme binding motif together with a spatially adjacent oxidoreductase active site motif could explain the regulation of its enzyme activity by redox changes. This intriguing hypothesis will now be investigated in more detail. Furthermore, CBS is also regulated by AdoMet through the interaction of this allosteric activator with the regulatory domain. To understand better the complicated regulatory mechanism of its enzymatic activity, we are trying to achieve further structural information of the full-length enzyme.

Materials and methods

Crystallization and data collection

Expression of the truncated CBS protein and subsequent crystallization were carried out as described (Janosik *et al.*, 2001). The X-ray diffraction data of the native crystal were collected at the beam-line BM1A of the SNBL at the ESRF in Grenoble at a temperature of 100 K. Furthermore, a MAD data set at three wavelengths around the absorption edge of the iron ion ($\lambda = 1.74 \text{ \AA}$) was collected at the BW7A beam-line (EMBL, DESY Hamburg) (Janosik *et al.*, 2001). The crystals belong to the trigonal space group $P3_1$ with unit cell dimensions $a = b = 144.46 \text{ \AA}$, $c = 108.21 \text{ \AA}$; $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$ (Table I). The crystals contain three dimers per asymmetric unit corresponding to a solvent content of 46%.

X-ray structure determination

The structure was solved by combining phase information from MR and MAD of the heme iron. For molecular replacement a polyserine model of OASS was used as search model. Loops that were not conserved in a three-dimensional alignment between the structures of OASS (open and closed form), tryptophan synthase and TD were removed. The calculations were made with the program AMoRe (Navaza, 1994) and yielded a unique single solution in the rotation search, since all dimers have almost the same orientation with respect to each other. The direction of local 2-fold axes was found to be perpendicular to the crystallographic 3-fold axis in agreement with the self-rotation function (Janosik *et al.*, 2001).

The phase information obtained by MR was sufficient to trace those parts of the protein that were homologous to the search model and to identify the positions of the heme irons. However, the N-terminal residues

were completely missing in these electron density maps. For this reason the MAD data set was collected. The phases obtained with this method were combined with the MR phases. The quality of the maps resulting from the combined phases was sufficient to be interpreted so that the missing parts of the model could be built.

The electron density maps were averaged over all six monomers A–F and density modified (CCP4, 1994). During refinement, which was performed with the program CNS (Brünger *et al.*, 1998), NCS symmetry restraints were applied on all monomers, using monomer A as reference. The model was built using the program O (Jones *et al.*, 1991).

The final model comprises a total of 2090 amino acids, corresponding to ~80% of all amino acids, six PLP, six heme and 154 water molecules. The final *R* factor and free *R* factors are rather high (25.7 and 29.6%, respectively). This can be explained by the rather poor quality of the crystals (high mosaicity) and by the fact that 20% of the protein residues were found to be disordered. These residues are probably disordered due to the truncation of the enzyme and are mainly located at the N- and C-terminus of the truncated protein. We are now addressing this problem by the design of more appropriate CBS truncation mutants lacking these disordered parts of the protein and hope to obtain better quality crystals using such improved deletion mutants.

Coordinates

The coordinates have been deposited in the Protein Data Bank (accession code 1JBQ).

Acknowledgements

We are grateful to J.N.Jansonius and U.Aebi for careful reading of the manuscript and for helpful discussions. This work was supported by grants of the Swiss National Science Foundation, by the M.E.Müller Foundation and the Canton Basel-Stadt, and by an NIH grant to J.P.K.

References

- Alexander,F.W., Sandmeier,E., Mehta,P.K. and Christen,P. (1994) Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific α , β and γ families. *Eur. J. Biochem.*, **219**, 953–960.
- Bateman,A. (1997) The structure of a domain common to archaeobacteria and the homocystinuria disease protein. *Trends Biochem. Sci.*, **22**, 12–13.
- Brünger,A.T. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Bruno,S., Schiavetti,F., Burkhard,P., Kraus,J.P., Janosik,M. and Mozzarelli,A. (2001) Functional properties of the active core of human cystathionine β -synthase crystals. *J. Biol. Chem.*, **276**, 16–19.
- Burkhard,P., Rao,G.S., Hohenester,E., Schnackerz,K.D., Cook,P.F. and Jansonius,J.N. (1998) Three-dimensional structure of *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.*, **283**, 121–133.
- Burkhard,P., Tai,C.H., Ristroph,C.M., Cook,P.F. and Jansonius,J.N. (1999) Ligand binding induces a large conformational change in *O*-acetylserine sulfhydrylase from *Salmonella typhimurium*. *J. Mol. Biol.*, **291**, 941–953.
- Burkhard,P., Tai,C.H., Jansonius,J.N. and Cook,P.F. (2000) Identification of an allosteric anion-binding site on *O*-acetylserine sulfhydrylase: structure of the enzyme with chloride bound. *J. Mol. Biol.*, **303**, 279–286.
- Christen,P. and Metzler,D.E. (1985) *Transaminases*. John Wiley & Sons, New York, NY.
- Collaborative Computing Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Gallagher,D.T., Gilliland,G.L., Xiao,G., Zondlo,J., Fisher,K.E., Chinchilla,D. and Eisenstein,E. (1998) Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. *Structure*, **6**, 465–475.
- Hyde,C.C., Ahmed,S.A., Padlan,E.A., Miles,E.W. and Davies,D.R. (1988) Three-dimensional structure of the tryptophan synthase $\alpha\beta\gamma$ multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.*, **263**, 17857–17871.
- Janosik,M., Meier,M., Kery,V., Oliveriusova,J., Burkhard,P. and Kraus,J.P. (2001) Crystallization and preliminary X-ray diffraction analysis of the active core of human recombinant cystathionine β -synthase: an enzyme involved in vascular disease. *Acta Crystallogr. D*, **57**, 289–291.
- Jhee,K.H., McPhie,P. and Miles,E.W. (2000) Domain architecture of the heme-independent yeast cystathionine β -synthase provides insights into mechanisms of catalysis and regulation. *Biochemistry*, **39**, 10548–10556.
- Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
- Kery,V., Bukovska,G. and Kraus,J.P. (1994) Transsulfuration depends on heme in addition to pyridoxal 5'-phosphate. Cystathionine β -synthase is a heme protein. *J. Biol. Chem.*, **269**, 25283–25288.
- Kery,V., Poneleit,L. and Kraus,J.P. (1998) Trypsin cleavage of human cystathionine β -synthase into an evolutionarily conserved active core: structural and functional consequences. *Arch. Biochem. Biophys.*, **355**, 222–232.
- Kery,V., Poneleit,L., Meyer,J.D., Manning,M.C. and Kraus,J.P. (1999) Binding of pyridoxal 5'-phosphate to the heme protein human cystathionine β -synthase. *Biochemistry*, **38**, 2716–2724.
- Kraus,J.P. (1994) Komrower Lecture. Molecular basis of phenotype expression in homocystinuria. *J. Inher. Metab. Dis.*, **17**, 383–390.
- Kraus,J.P. *et al.* (1999) Cystathionine β -synthase mutations in homocystinuria. *Hum. Mutat.*, **13**, 362–375.
- Maclean,K.N., Janosik,M., Oliveriusova,J., Kery,V. and Kraus,J.P. (2000) Transsulfuration in *Saccharomyces cerevisiae* is not dependent on heme: purification and characterization of recombinant yeast cystathionine β -synthase. *J. Inorg. Biochem.*, **81**, 161–171.
- Mudd,S.H., Levy,H.L. and Kraus,J.P. (2001) Disorders of transsulfuration. In Scriver,C.R., Beaudet,A.L., Sly,W.S., Valle,D., Childs,B., Kinzler,K.W. and Vogelstein,B. (eds), *The Metabolic and Molecular Bases of Inherited Disease*. Vol. 1. McGraw-Hill, New York, NY, pp. 2007–2056.
- Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.
- Ojha,S., Hwang,J., Kabil,O., Penner-Hahn,J.E. and Banerjee,R. (2000) Characterization of the heme in human cystathionine β -synthase by X-ray absorption and electron paramagnetic resonance spectroscopies. *Biochemistry*, **39**, 10542–10547.
- Omura,T., Sadano,H., Hasegawa,T., Yoshida,Y. and Kominami,S. (1984) Hemoprotein H-450 identified as a form of cytochrome P-450 having an endogenous ligand at the 6th coordination position of the heme. *J. Biochem. (Tokyo)*, **96**, 1491–1500.
- Poulos,T.L. (1987) *Heme Enzyme Crystal Structures*. Elsevier Science Publishing Co., Inc., New York, NY.
- Schneider,T.R., Gerhardt,E., Lee,M., Liang,P.H., Anderson,K.S. and Schlichting,I. (1998) Loop closure and intersubunit communication in tryptophan synthase. *Biochemistry*, **37**, 5394–5406.
- Shan,X., Dunbrack,R.L., Jr, Christopher,S.A. and Kruger,W.D. (2001) Mutations in the regulatory domain of cystathionine β -synthase can functionally suppress patient-derived mutations in *cis*. *Hum. Mol. Genet.*, **10**, 635–643.
- Skovby,F., Kraus,J.P. and Rosenberg,L.E. (1984) Biosynthesis of human cystathionine β -synthase in cultured fibroblasts. *J. Biol. Chem.*, **259**, 583–587.
- Swaroop,M., Bradley,K., Ohura,T., Tahara,T., Roper,M.D., Rosenberg,L.E. and Kraus,J.P. (1992) Rat cystathionine β -synthase. Gene organization and alternative splicing. *J. Biol. Chem.*, **267**, 11455–11461.
- Taoka,S., Ohja,S., Shan,X., Kruger,W.D. and Banerjee,R. (1998) Evidence for heme-mediated redox regulation of human cystathionine β -synthase activity. *J. Biol. Chem.*, **273**, 25179–25184.
- Taoka,S., Widjaja,L. and Banerjee,R. (1999) Assignment of enzymatic functions to specific regions of the PLP-dependent heme protein cystathionine β -synthase. *Biochemistry*, **38**, 13155–13161.
- Yao,M. *et al.* (2000) Crystal structure of 1-aminocyclopropane-1-carboxylate deaminase from *Hansenula saturnus*. *J. Biol. Chem.*, **275**, 34557–34565.
- Yap,S., Naughten,E.R., Wilcken,B., Wilcken,D.E. and Boers,G.H. (2000) Vascular complications of severe hyperhomocysteinemia in patients with homocystinuria due to cystathionine β -synthase deficiency: effects of homocysteine-lowering therapy. *Semin. Thromb. Hemost.*, **26**, 335–340.
- Zhang,R., Evans,G., Rotella,F.J., Westbrook,E.M., Beno,D., Huberman,E., Joachimiak,A. and Collart,F.R. (1999) Characteristics and crystal structure of bacterial inosine-5'-monophosphate dehydrogenase. *Biochemistry*, **38**, 4691–4700.

Received April 20, 2001; revised and accepted June 8, 2001

Corrigendum

In Figure 3B, Asp266 was wrongly labelled as Asp226 in the original publication. This error is corrected in the present version.

Chapter 3

Structural insights into mutations of cystathionine β -synthase

Meier, M., Oliveriusova, J., Kraus, J.P. and Burkhard, P. (2003) Structural insights into mutations of cystathionine beta-synthase. *Biochim Biophys Acta*, **1647**, 206-213.

The structural analysis presented in this chapter is my responsibility, with few additions of Jan P. Kraus. The CBS mutation database at www.uchsc.edu/sm/cbs/cbsdata/cbsmain.html is an effort of the Kraus lab.

Structural insights into mutations of cystathionine β -synthase

Markus Meier^a, Jana Oliveriusova^b, Jan P. Kraus^b, Peter Burkhard^{a,*}

^aM.E. Müller Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056, Basel, Switzerland

^bDepartments of Pediatrics and Cellular and Structural Biology, University of Colorado School of Medicine, Denver, CO 80262, USA

Received 12 August 2002; accepted 10 September 2002

Abstract

Cystathionine β -synthase (CBS) is a unique heme-containing enzyme that catalyses a pyridoxal 5'-phosphate (PLP)-dependent condensation of serine and homocysteine to give cystathionine. Deficiency of CBS leads to homocystinuria, an inherited disease of sulfur amino acid metabolism characterised by increased levels of homocysteine and methionine and decreased levels of cysteine. Presently, more than 100 CBS mutations have been described which lead to homocystinuria with different degrees of severity in the patients. We have recently solved the crystal structure of a truncated form of this enzyme, which enables us to correlate some of these mutations with the structure. © 2003 Elsevier Science B.V. All rights reserved.

Keywords: Cystathionine β -synthase; Mutation; X-ray crystal structure; Pyridoxal 5'-phosphate; Heme protein; Homocysteine

1. Introduction

Cystathionine β -synthase (CBS, L-serine hydrolyase, EC 4.2.1.22) is the first enzyme of the transsulfuration pathway in which the potentially toxic homocysteine is converted to cysteine. Deficiency of CBS activity is the most common cause of homocystinuria, an inherited metabolic disease characterised by dislocated eye lenses, skeletal problems, vascular disease, and mental retardation [1]. There have now been over 100 mutations described in this gene [2]; the continuously updated CBS website at <http://www.uchsc.edu/sm/cbs/cbsdata/cbsmain.htm> lists 127 mutations in 503 patient alleles. Elevated levels of homocysteine are correlated with cardiovascular diseases, neural tube defects, and Alzheimer's disease [3–5].

The human CBS is a homotetramer consisting of 63-kDa subunits, which binds two cofactors, pyridoxal 5'-phosphate (PLP) and heme [6,7]. Each CBS monomer of 551 amino acid residues binds two substrates (homocysteine and serine) and is further regulated by S-adenosyl-L-methionine (AdoMet) [8]. While the role of heme in CBS is unknown, catalysis by CBS can be explained solely by participation of PLP in the reaction mechanism [9]. We have recently

solved the X-ray crystal structure of recombinant human CBS comprising the amino acid residues 1–413 [10]. This truncated form still binds heme and PLP but its activity is no longer modulated by AdoMet. The oligomerisation state has

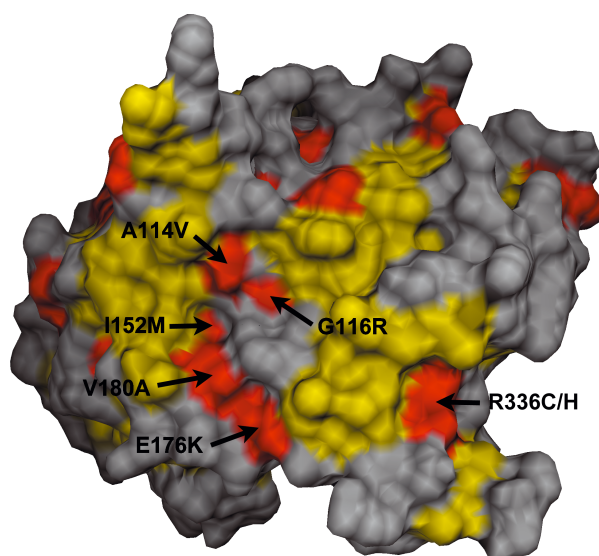


Fig. 1. Solvent accessible surface of the dimer interface between monomers A and B. Hydrophobic and aromatic residues are shown in yellow. Mutated residues are coloured in red and are labelled. The figure was prepared using the program DINO (<http://www.dino3d.org>).

* Corresponding author. Tel.: +41-61-267-20-91; fax: +41-61-267-21-09.

E-mail address: Peter.Burkhard@unibas.ch (P. Burkhard).

Table 1
Listing of CBS-mutations

Mutation	Location	Interactions	Adjacent mutations	References
P49L R58W	N-terminal extension, solvent accessible N-terminal extension, solvent exposed	– H-bond between N _ε of R58 and O _{ε1} of E62, H-bond between R58 N (mc) and N 55 O (mc), hydrophobic contacts between R58 sc and heme methyl and ethyl groups of pyrrol B	R224H –	Gaustadnes, M. ^a [20]
H65R	Heme ligand, solvent exposed	Heme Fe coordination, H-bond between H65 N (mc) and S63 O _γ	A226T	[14]
P78R	Periphery of dimer interface, loop between β-strand 1 and β-strand 2, solvent exposed	H-bond between P78 O (mc) (monomer A) and N93 N _{δ2} (monomer B)	–	[21]
G85R	Near dimer interface, start of β-strand 2, partially exposed to solvent	H-bond between G85 N (mc) and N113 O _{δ1}	R121C/H/L and E239K	[22]
P88S	Dimer interface, β-strand 2, partially exposed to solvent	H-bond between P88 O (mc) and C109 N (mc), hydrophobic contacts to K83 (monomer B) and L77 (B)	E239K	[23]
L101P	Loop between α-helix 1 and β-strand 3, sc buried, mc solvent accessible	Sc surrounded by hydrophobic side chains (C103, L105, V358, A361, Q362), H-bond between L101 N (mc) and G96 O (mc)	K102N/Q, A361T	[24]
K102N/Q	Loop between α-helix 1 and β-strand 3, solvent accessible	–	L101P, A361T	[21,25]
A114V	Heart of dimer interface, β-turn, connection between C-terminal and N-terminal domains (β-strand 3 and α-helix 2)	H-bond between A114 N (mc) and K83 O (mc), hydrophobic contact to F112 of monomer B	G116R	[20,25,26]
G116R	Dimer interface, β-turn between β-strand 3 and α-helix 2	H-bond between G116 N (mc) and N113 O (mc), H-bond between G116 O (mc) and S377 (mc)	A114V, I152M	[27]
G139R	Periphery of hypothetic interface to regulatory domain, loop between α-helix 2 and β-strand 4, solvent accessible	–	–	[19]
G148R	Residue before start of α-helix 3, active site cleft	Distorted H-bond between G148 O (mc) and I152 N (mc)	G151R, I152M, G305R	Ohura, T. ^a
I152M	Near dimer interface, α-helix 3	H-bond between I152 O (mc) and L156 N (mc), hydrophobic contact to L386 (monomer B)	G116R, G148R, G151R, A155T, V180A G151R,	[28,29]
A155T	Hypothetic interface to regulatory domain, α-helix 3, partly solvent accessible	H-bond between A155 N (mc) and G151 O (mc), H-bond between A155 O (mc) and A159N (mc)	G151R, I152M, C165Y	[14]
C165Y	Hypothetic interface to regulatory domain, β-strand 5, partly solvent accessible	H-bond between C165 O (mc) and E187 N (mc)	A155T	[14,30,31]
E176K	Periphery of dimer interface, α-helix 4, solvent accessible	H-bond between E176 O _{ε2} and T383 O _{γ1} , H-bond between E176 O _{ε2} and T383 N (mc), hydrophobic contact to M382 of monomer B	V180A, K384E	[25]
V180A	Dimer interface, α-helix 4, buried	H-bond between V180 O (mc) and L184 N (mc), hydrophobic contacts to M382 (monomer B) and L386 (B)	E176K, I152M	[31]
R224H	Heme binding pocket, connection between N- and C-terminal domain between β-strand 7 and α-helix 6, interface to active site, solvent accessible	SB between guanidium group of R224 and propionyl group of pyrrol D of heme	P49L, A226T, N228K/S	[32]
A226T	Heme binding site, connection between N- and C-terminal domain between β-strand 7 and α-helix 6, interface to active site, solvent accessible	H-bond between A226 O (mc) and L230 N (mc), hydrophobic contact to methyl group of pyrrol C of heme	R224H, H65R, N228K/S	[18]
N228K/S	α-helix 6, interface between heme binding pocket and active site	H-bond between N228 N _{δ2} and Q222 O (mc), H-bond between N228 N _{δ2} and T257 O _{γ1} , H-bond between N228 N (mc) and N225 O (mc)	R125P/Q/W, M126V, R224H, A226T, T257M, G259S	[24]

(continued on next page)

Table 1 (continued)

Mutation	Location	Interactions	Adjacent mutations	References
T257M	PLP phosphate binding loop between β -strand 8 and α -helix 8	H-bond between T257 O _{γ1} and O ₃ of PLP phosphate, H-bond between T257 O _{γ1} and G259 N (mc), H-bond between T257 N (mc) and O ₃ of PLP phosphate	N228K/S, G259S	[23]
G259S	PLP phosphate binding loop, start of α -helix 8, interface between heme binding pocket and active site	H-bond between G259 N (mc) and T257 O _{γ1} , H-bond between G259 O (mc) and G263 N (mc)	N228K/S, T257M, T262M/R	Chen, L.S. ^a
T262M/R	Interface between heme binding pocket and active site, α -helix 8	H-bond between T262 O _{γ1} and D316 N (mc), H-bond between T262 N (mc) and G258 O (mc), H-bond between T262 O (mc) and R266 N (mc)	G259S, R266G/K	[24,33,34]
R266G/K	Heme binding pocket, α -helix 8, partially solvent exposed	Van der Waals contacts between R266 sc and heme pyrrol ring C and between R266 sc and W54 sc, H-bond between R266 N _{ϵ} and D316 O _{δ1} , H-bond between R266 N (mc) and T262 O (mc)	T262M/R	[33]
G305R	Active site, next to the <i>re</i> face of the pyridine ring of the PLP, located in a long loop between β -strand 9 and β -strand 10; $\varphi = 89.5^\circ$, $\psi = 9.9^\circ$	Van der Waals contact to <i>re</i> face of the pyridine ring	G307S	[27]
G307S	Active site cleft, located in a long loop between β -strand 9 and β -strand 10, partly solvent accessible; $\varphi = 83.4$, $\psi = 168.2$	H-bond between G307 N (mc) and V255 O (mc)	G305R	[20,24,33,35–37]
R336C/H	Periphery of dimer interface, hypothetic interface to regulatory domain, α -helix 9, solvent accessible	SB between guanidium group of R336 and carboxyl group of D388; H-bond between R336 N _{ϵ} and F385 O (mc); H-bond between R336 N (mc) and F332 O (mc); H-bond between R336 O (mc) and A340 N (mc), Van der Waals contact between the guanidium group of R336 and C _{ϵ} , S _{δ} , C _{γ} of M391; not in direct contact with monomer B	M391I	[20,38]
A331E/V	α -helix 9, buried	H-bond between A331 N (mc) and N327 O (mc), H-bond between A331 O (mc) and A335 N (mc)	S352N, A355P	[30,32]
S352N	α -helix 10, buried	H-bond between S352 N (mc) and G348 O (mc), H-bond between S352 O (mc) and V356 N (mc)	A331E/V, T353M, V354M, A355P	Shih, V. ^a
T353M	α -helix 10, buried	H-bond between T353 O _{γ} and S349 O (mc), H-bond between T353 N (mc) and S349 O (mc), H-bond between T535 O (mc) and A357 N (mc)	S352N, V354M, A355P	[18,30]
V354M	α -helix 10, buried	Sc surrounded by hydrophobic sc of I92, I95, F334, L338, V358, L374; H-bond between V354 N (mc) and A350 O (mc); H-bond between V354 O (mc) and V358 (mc)	S352N, T353M, A355P	[38]
A355P	Hypothetic interface to regulatory domain, α -helix 10, solvent accessible	Sc stacked against ring plane of F334, H-bond between A355 N (mc) and G351 O (mc)	A331E/V, S352N, T353M, V354M	[24]
M391I	Connection between the catalytic domains and the regulatory domain, α -helix 11, partly solvent accessible	Packed against a hydrophobic cluster formed by F385, F332, W390 and F396; Van der Waals contact between C _{ϵ} , S _{δ} , C _{γ} of M391 and the guanidium group of R336	R336C/H, K384E/N	Koch, H.G. ^a

Description of CBS mutations. sc: side chain, mc: main chain.

^a Unpublished observation.

changed from tetrameric to dimeric and the enzyme activity is about two to three times higher than that of wt CBS.

The catalytic part of CBS has a high sequence homology to *O*-acetylserine sulfhydrylase (OASS) [11], especially in the active centre, and it belongs to the same β -family of vitamin B₆-dependent enzymes [12]. This fold consists of two domains, commonly referred to as the N-terminal and C-terminal domain. The missing residues 414–551 will be referred to as regulatory domain in this paper.

The availability of the structure of the truncated enzyme enables us now to correlate some of the point mutations described in the literature with the structure. We will mainly focus on locations of special interest such as the dimer interface, the active site, the heme-binding site, and the putative location of the regulatory domain.

2. Results and discussion

2.1. Mutations in the dimer interface

A number of CBS point mutations are located at the interface between the two monomers (Fig. 1). Mutations in the dimer interface probably destabilise the monomer–monomer interaction or would affect a potential communication between the monomers. The interface is characterised by a large hydrophobic area with Phe111 and Phe112 in the centre and few polar interactions at the periphery. All contacts between the monomers are two-fold because the two-fold dimer axis lies in the interface plane. Two known point mutations, namely, A114V and G116R, occur in the centre of the interface. Both mutations are located in the same β -turn between β -strand 3 and α -helix 2 (for the classifications of secondary structure elements, compare Ref. [10]), which is part of the connection between the N- and C-terminal domains. Both mutations are on the surface of the monomer, but only A114 is in hydrophobic contact

with F112 of the other monomer. Shan et al. [13] have shown that the enzymatic activity can be rescued in vivo by the deletion of the regulatory domain of CBS. Expression of the A114V mutant in *Escherichia coli* shows variable amounts of residual activity, which is greatly stimulated (to near wt levels) by inclusion of PLP in the assay and pyridoxal in the growth medium (our unpublished results). G116 is located at the border of a small cavity between the monomers and therefore has no direct contact to the other monomer. Both mutations are relatively close to the molecular two-fold axis. The point mutations P78R, P88S, E176K, V180A, R336C/H are located at the periphery of the interface. All these residues except for V180 are solvent accessible. The residue E176 is located very close to the two-fold axis (2.7 Å). Patients with the E176K mutation do not respond to vitamin B₆ treatment in contrast to patients having one of the other mutations of the dimer interface. This mutant has been expressed in *E. coli* and forms high molecular weight aggregates devoid of heme [14]. A list including references and short description of all mutations discussed in this paper can be found in Table 1.

2.2. Mutations in the active site

The coenzyme PLP is deeply buried in a cleft between the N-terminal and C-terminal domains and the active site is accessible only via a narrow channel. The PLP cofactor is linked to the ϵ -amino group of K119 via a Schiff base linkage forming the so-called “internal aldimine”. Four of the six known point mutations in the active site involve glycine residues: G148R, G305R, G307S, and G259S. The other two mutations are T257M and N228K/S (Fig. 2). G259 separates the active site region from the heme-binding pocket. Also the residue N228 is located in this interface. The mutation G307S occurs frequently and has been found in 87 patient alleles. It is not responsive to vitamin B₆ treatment. The residue G307 lines the entry to the active site

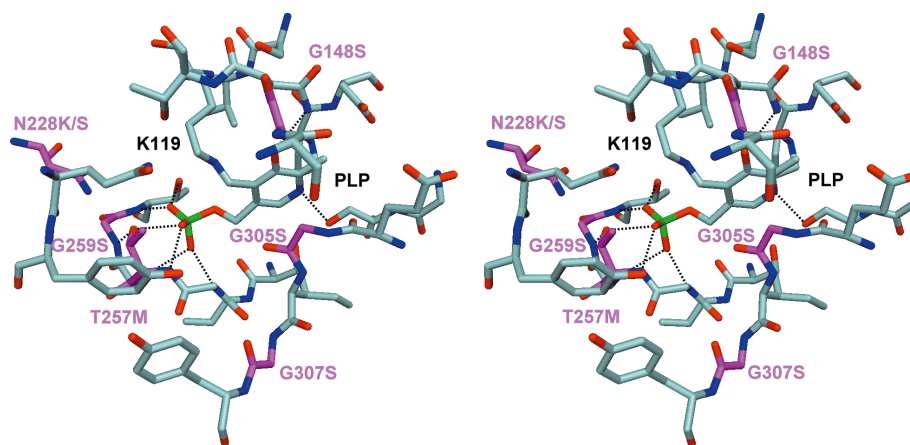


Fig. 2. Stereo picture of the active site. Atoms are colour-coded according to atom type (grey: carbon, red: oxygen, blue: nitrogen, green: phosphorus); mutated residues are labelled and their carbon atoms are coloured in magenta. Hydrogen bonds between the PLP cofactor and the surrounding residues are shown as black dotted lines. The figure was prepared using the program DINO (<http://www.dino3d.org>).

cleft and is 6.5 Å apart from the PLP cofactor. The orientation of G307 does not allow accommodating the side chain of a serine residue, because P282, S285, and Y301 are tightly packed against this side of the residue. Incorporation of such a side chain can be expected to cause a substantial conformational change in the loop containing the residues or in the adjacent chains. Probably the second substrate homocysteine binds in this region. A rearrangement of the residues in this region due to the G307S mutation would affect and probably inhibit the binding of homocysteine. The fact that this mutation is not responsive to vitamin B₆ might be surprising, because a conformational change of the residues adjacent to the cofactor should also affect its binding. However, if the mutation completely abolishes binding of homocysteine, the activity of the enzyme cannot be rescued by a supplement of PLP.

The residue G305 is in hydrophobic contact to the *re*-face of the pyridine ring of the PLP cofactor (Fig. 2). A mutation of glycine 305 to arginine will most probably affect cofactor binding and potentially also the binding of the substrate serine, assuming that serine binds to CBS in a similar fashion as the substrate analogue methionine in the complex structure of OASS [15]. The mutation is responsive to vitamin B₆ treatment. The mutation G148R is located at the opposite side of the active site cleft. This is a residue located in the “asparagine loop” which undergoes a large conformational change in OASS upon ligand binding [15]. The loop forms an extended hydrogen-bonding network to the carboxylate moiety of the ligand in OASS and is likely to do the same in CBS. Glycine 148 is a conserved residue between the two enzymes and has φ and ψ main chain angles that are only allowed for glycine residues in both, the open and the closed conformations of OASS. This indicates that the residue is important for the flexibility of the loop.

Two mutations affect the loop, which anchors the phosphate group of PLP by an extensive hydrogen-bonding

network, namely, G259S and T257M (Fig. 2). The side chain hydroxyl group of T257 forms a hydrogen bond to the phosphate. This interaction is abolished if the residue is mutated to methionine. In addition, the bulkier side chain of methionine probably occludes part of the substrate-binding site, leading to a reduced affinity of the enzyme for the substrate and hence to a reduced activity. This mutation is not B₆ responsive.

2.3. Mutations in the heme-binding site

In the structure of CBS [10], all three dimers of the asymmetric unit contain two heme molecules, which are located at distal ends of the dimers. The orientation of their ring planes is normal to the protein surface. The heme is bound in a hydrophobic pocket formed by residues 50–67, α -helices 6 and 8, and a loop preceding β -strand 10. The sulfhydryl group of C52 and the N_{ε2} atom of H65 axially coordinate the iron in the heme (Fig. 3). In two patients, histidine 65 is mutated to an arginine. This mutation has been spectroscopically analysed *in vitro* by Ojha et al. [16]. It was found that another ligand (probably H67 or P64) seems to be able to substitute for H65; however, the heme saturation of the purified enzyme and its activity are low. The mutation is not responsive to vitamin B₆ treatment.

The long side chain of R58 makes hydrophobic contacts to the methyl and ethyl groups of the pyrrol ring B of the heme (Fig. 3). In addition, it forms a hydrogen bond between the N_ε atom and the carboxyl group of E62, which is probably important for the stability of the N-terminal heme-binding loop. A patient with the mutation R58W has yet another mutation, A114V, on the same allele. While the patients carrying only the A114V mutation are vitamin B₆ responsive, the patient with both mutations is not. The mutation R58W most probably reduces the ability of the enzyme to bind heme. This might affect proper folding of the protein and would therefore

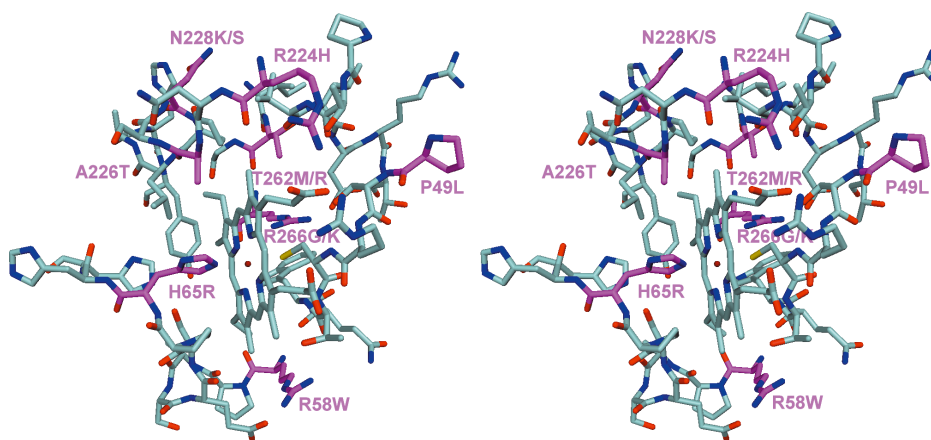


Fig. 3. Stereo picture of the heme binding pocket. Atoms are colour-coded according to atom type (grey: carbon, red: oxygen, blue: nitrogen, green: phosphorus); mutated residues are labelled and their carbon atoms are coloured in magenta. The figure was prepared using the program DINO (<http://www.dino3d.org>).

lead to reduced levels of catalytically active enzyme in the patient.

In several patients, the residue R266 is mutated to lysine. The arginine residue is in contact with the pyrrol ring C of the heme and the side chain is stacked against the plane of the indole ring of W54 (Fig. 3). In addition, there is a hydrogen bond between N_ε of R266 and O_{δ1} of D316. Patients with this mutation can be treated by pharmacological doses of vitamin B₆, indicating that the heme binding is not affected by this mutation. This outcome can be expected, because arginine and lysine are amino acids with similar properties. In contrast, the mutation R266G, which is present in only one patient, is not vitamin B₆ responsive. The loss of the interaction between the side chains of R266 and W54 is expected to lead to a collapse of the heme-binding pocket and the heme can no longer bind to the enzyme.

The mutations R224H and A226T are located in the connecting loop between the N- and C-terminal domain between β-strand 7 and α-helix 6 and separate the heme-binding pocket from the active site. The guanidium group of R224 forms a salt bridge to the propionate group of the pyrrol ring D of the heme. The side chain of A226 is in hydrophobic contact to the methyl group of the pyrrol ring C of the heme. Both mutations respond to a vitamin B₆ treatment. They can be functionally suppressed in a yeast assay by the deletion of the regulatory domain of CBS [13], indicating that these residues might play a role in the communication between the N-terminal catalytic domains and the regulatory domain of the enzyme.

2.4. Other mutations

It can be expected that quite a few CBS point mutations affect the regulatory properties of the enzyme. In the structure of CBS [10], the regulatory domain is missing (residues 414–551). Apart from modulating the enzyme activity by binding the allosteric activator *S*-adenosyl methionine, these residues are responsible for the tetramerisation. The location of the interface between the catalytic domains and the missing regulatory domain can be predicted by the presence of hydrophobic areas on the surface of the structure. There are two such patches on the surface of the dimer, which are related by the two-fold dimer axis (Fig. 4). These patches are located on the back side of the catalytic cleft formed by the N-terminal β-sheet of monomer A and the adjacent α-helices 1, 9, and 11 in the C-terminal domain of monomer B (and vice versa). A superposition of CBS and threonine deaminase [17] shows that the regulatory domain of threonine deaminase is located roughly in the same region.

One mutation of a residue of the surface of this region is R336C/H. The hydrophobic portion of the side chain of this arginine is packed against the protein surface while the guanidium group forms a salt bridge to the carboxyl group of D388. Even though the residue is close to the dimer interface, it does not contribute to any interactions between the two monomers. The mutation causes a mild disease type that is responsive to vitamin B₆ treatment. The mutation M391I that is adjacent to R336 is not responsive to B₆ supplemented diet. Two other mutations occurring in this

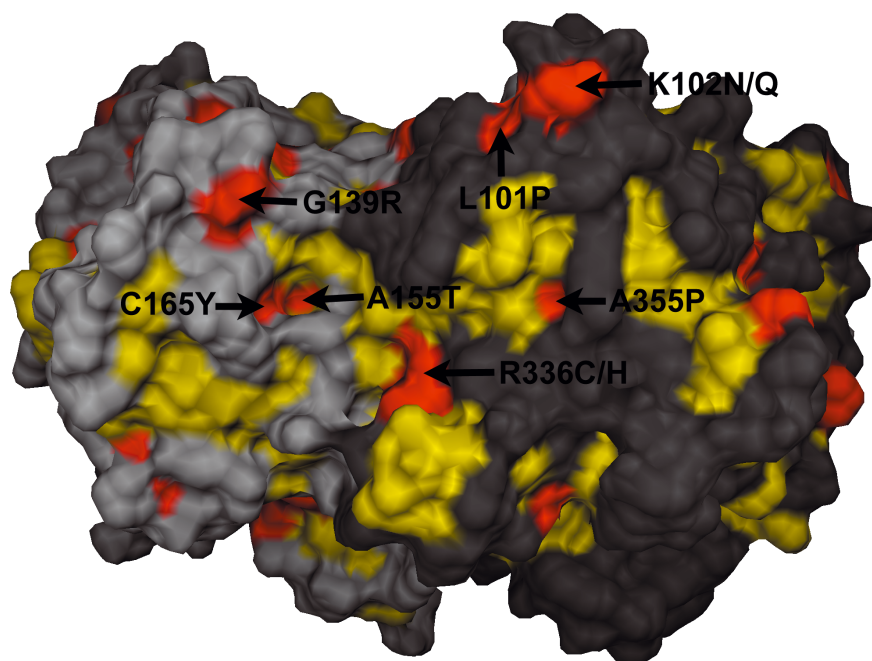


Fig. 4. Solvent accessible surface of the interface region between the catalytic domains and the missing regulatory domain. Hydrophobic and aromatic residues are shown in yellow. Mutated residues are coloured in red and are labelled. One monomer is drawn in light grey, the other monomer in dark grey. The figure was prepared using the program DINO (<http://www.dino3d.org>).

hypothetic interface are A155T and C165Y. They are located in the α -helix 3 and β -strand 5, respectively, and are close to each other. The A155T mutant has been expressed in *E. coli* and forms hemeless high molecular weight aggregates [14].

A cluster of mutations in the hypothetic interface to the regulatory domain is located in or around α -helix 10 in the C-terminal domain: A331E/V, S352N, T353M, V354M, A355P. The residue A355 is still partly solvent accessible while the others are already in the interior of the protein. The mutations A331V and T353M can be suppressed in a yeast assay by the deletion of the regulatory domain of the protein [13]. The S352N and V354M mutations are not B₆ responsive, and the same is true for most of the patients carrying the T353M mutation.

The most prevalent mutation in CBS is I278T. It is found in about 25% of all homocystinuric alleles. It is located in the middle of β -strand 9 in the β -sheet of the C-terminal domain and is partly solvent accessible. The effects of this mutation can be suppressed, when expressed in yeast, by certain point mutations in the regulatory domain of the enzyme or by complete deletion of the C-terminal region [13,18]. The I278T protein has been expressed in *E. coli*. It is only expressed in low amounts compared to wt CBS and forms high molecular weight aggregates, which do not contain heme [14]. Patients heterozygous or homozygous for this mutation respond fully to treatment with pharmacological doses of pyridoxine [19]. Surprisingly, the CBS subunits carrying this mutation are not detectable in fibroblast extracts from these patients even when cultured in pyridoxal-supplemented medium (unpublished results). The fact, that this mutation can be suppressed by mutations in the regulatory domain, suggests, that this residue is involved in the communication between regulatory domain and the catalytic domains of CBS.

Finally, it should be noted that in different patients, the same genotype can cause different phenotypes even among siblings [2]. Therefore, care must be taken when drawing conclusions about the phenotypes of mutations that have only been found in one patient.

Acknowledgements

This work was supported by grants of the Swiss National Science Foundation, by the M.E. Müller Foundation, and the Canton Basel-Stadt and by NIH grants PO1HD0805 and HL65217 to J.P.K.

References

- [1] S.H. Mudd, H.L. Levy, J.P. Kraus, in: C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Childs, K.W. Kinzler, B. Vogelstein (Eds.), *The Metabolic and Molecular Bases of Inherited Disease*, McGraw Hill, New York, 2001, pp. 2007–2056.
- [2] J.P. Kraus, M. Janosik, V. Kozich, R. Mandell, V. Shih, M.P. Sperandeo, G. Sebastio, R. de Franchis, G. Andria, L.A. Kluijtmans, H. Blom, G.H. Boers, R.B. Gordon, P. Kamoun, M.Y. Tsai, W.D. Kruger, H.G. Koch, T. Ohura, M. Gaustadnes, *Hum. Mutat.* 13 (1999) 362–375.
- [3] H. Refsum, P.M. Ueland, O. Nygard, S.E. Vollset, *Annu. Rev. Med.* 49 (1998) 31–62.
- [4] J.L. Mills, J.M. McPartlin, P.N. Kirke, Y.J. Lee, M.R. Conley, D.G. Weir, J.M. Scott, *Lancet* 345 (1995) 149–151.
- [5] R. Clarke, A.D. Smith, K.A. Jobst, H. Refsum, L. Sutton, P.M. Ueland, *Arch. Neurol.* 55 (1998) 1449–1455.
- [6] V. Kery, G. Bukovska, J.P. Kraus, *J. Biol. Chem.* 269 (1994) 25283–25288.
- [7] F. Skovby, J.P. Kraus, L.E. Rosenberg, *J. Biol. Chem.* 259 (1984) 583–587.
- [8] M. Janosik, V. Kery, M. Gaustadnes, K.N. Maclean, J.P. Kraus, *Biochemistry* 40 (2001) 10625–10633.
- [9] V. Kery, L. Poneleit, J.D. Meyer, M.C. Manning, J.P. Kraus, *Biochemistry* 38 (1999) 2716–2724.
- [10] M. Meier, M. Janosik, V. Kery, J.P. Kraus, P. Burkhard, *EMBO J.* 20 (2001) 3910–3916.
- [11] P. Burkhard, G.S. Rao, E. Hohenester, K.D. Schnackerz, P.F. Cook, J.N. Jansonius, *J. Mol. Biol.* 283 (1998) 121–133.
- [12] F.W. Alexander, E. Sandmeier, P.K. Mehta, P. Christen, *Eur. J. Biochem.* 219 (1994) 953–960.
- [13] X. Shan, R.L. Dunbrack Jr., S.A. Christopher, W.D. Kruger, *Hum. Mol. Genet.* 10 (2001) 635–643.
- [14] M. Janosik, J. Oliveriusova, B. Janosikova, J. Sokolova, E. Kraus, J.P. Kraus, V. Kozich, *Am. J. Hum. Genet.* 68 (2001) 1506–1513.
- [15] P. Burkhard, C.H. Tai, C.M. Ristroph, P.F. Cook, J.N. Jansonius, *J. Mol. Biol.* 291 (1999) 941–953.
- [16] S. Ojha, J. Wu, R. LoBrutto, R. Banerjee, *Biochemistry* 41 (2002) 4649–4654.
- [17] D.T. Gallagher, G.L. Gilliland, G. Xiao, J. Zondlo, K.E. Fisher, D. Chinchilla, E. Eisenstein, *Structure* 6 (1998) 465–475.
- [18] X. Shan, W.D. Kruger, *Nat. Genet.* 19 (1998) 91–93.
- [19] V.E. Shih, J.M. Fringer, R. Mandell, J.P. Kraus, G.T. Berry, R.A. Heidenreich, M.S. Korson, H.L. Levy, V. Ramesh, *Am. J. Hum. Genet.* 57 (1995) 34–39.
- [20] R. de Franchis, E. Kraus, V. Kozich, G. Sebastio, J.P. Kraus, *Hum. Mutat.* 13 (1999) 453–457.
- [21] R. de Franchis, V. Kozich, R.R. McInnes, J.P. Kraus, *Hum. Mol. Genet.* 3 (1994) 1103–1108.
- [22] K.N. Maclean, M. Gaustadnes, J. Oliveriusova, M. Janosik, E. Kraus, V. Kozich, V. Kery, F. Skovby, N. Rudiger, J. Ingerslev, S.P. Stabler, R.H. Allen, J.P. Kraus, *Hum. Mutat.* 19 (2002) 641–655.
- [23] G. Sebastio, M.P. Sperandeo, M. Panico, R. de Franchis, J.P. Kraus, G. Andria, *Am. J. Hum. Genet.* 56 (1995) 1324–1333.
- [24] P.M. Gallagher, E. Naughten, N.Q. Hanson, K. Schwichtenberg, M. Bignell, M. Yuan, P. Ward, S. Yap, A.S. Whitehead, M.Y. Tsai, *Mol. Genet. Metab.* 65 (1998) 298–302.
- [25] V. Kozich, M. Janosik, J. Sokolova, J. Oliveriusova, M. Orendac, J.P. Kraus, D. Elleder, *J. Inherit. Metab. Dis.* 20 (1997) 363–366.
- [26] V. Kozich, R. de Franchis, J.P. Kraus, *Hum. Mol. Genet.* 2 (1993) 815–816.
- [27] M.P. Sperandeo, M. Candito, G. Sebastio, M.O. Rolland, C. Turc-Carel, H. Giudicelli, P. Dellamonica, G. Andria, *J. Inherit. Metab. Dis.* 19 (1996) 351–356.
- [28] L.A. Kluijtmans, G.H. Boers, F.J. Trijbels, H.M. van Lith-Zanders, L.P. van den Heuvel, H.J. Blom, *Biochem. Mol. Med.* 62 (1997) 23–25.
- [29] L.A.J. Kluijtmans, G.H.J. Boers, J.P. Kraus, L.P.W.J. van den Heuvel, J.R.M. Cruysberg, F.J.M. Trijbels, H.J. Blom, University of Nijmegen, The Netherlands, 1998.
- [30] P.A. Dawson, A.J. Cox, B.T. Emmerson, N.P. Dudman, J.P. Kraus, R.B. Gordon, *Eur. J. Hum. Genet.* 5 (1997) 15–21.
- [31] L.A. Kluijtmans, G.H. Boers, J.P. Kraus, L.P. van den Heuvel, J.R. Cruysberg, F.J. Trijbels, H.J. Blom, *Am. J. Hum. Genet.* 65 (1999) 59–67.
- [32] W.D. Kruger, D.R. Cox, *Hum. Mol. Genet.* 4 (1995) 1155–1161.

- [33] C.E. Kim, P.M. Gallagher, A.B. Guttormsen, H. Refsum, P.M. Ueland, L. Ose, I. Folling, A.S. Whitehead, M.Y. Tsai, W.D. Kruger, *Hum. Mol. Genet.* 6 (1997) 2213–2221.
- [34] G. Gat-Yablonski, H. Mandel, B. Fowler, O. Taleb, B.A. Sela, *Hum. Mutat.* 16 (2000) 372.
- [35] F.L. Hu, Z. Gu, V. Kozich, J.P. Kraus, V. Ramesh, V.E. Shih, *Hum. Mol. Genet.* 2 (1993) 1857–1860.
- [36] M.Y. Tsai, N.Q. Hanson, M.K. Bignell, K.A. Schwichtenberg, *Clin. Biochem.* 29 (1996) 473–477.
- [37] M.Y. Tsai, P.W. Wong, U. Garg, N.Q. Hanson, K. Schwichtenberg, *Biochem. Mol. Med.* 61 (1997) 9–15.
- [38] M. Coude, J. Aupetit, M.T. Zobot, P. Kamoun, B. Chadeaux-Vecemans, *J. Inherit. Metab. Dis.* 21 (1998) 823–828.

Chapter 4

Introduction to coiled coils

Introduction to coiled coils

α -helical coiled coils are a versatile protein oligomerisation motif. They consist of α -helices which wind around each other to form a left-handed *superhelix*. Right-handed coiled coils can also be found (Stetefeld *et al.*, 2000), but are not discussed here. Coiled coils can form dimers, trimers, tetramers and pentamers depending on their amino acid sequence and the environment. The α -helices can be arranged in a parallel or antiparallel manner.

1. Geometry of coiled coils

Left-handed coiled coils are characterised by a 7 amino acid repeat $(abcdefg)_n$ in their primary sequence, known as heptad repeat. Positions *a* and *d* are called *core positions*. They are usually occupied by hydrophobic residues like leucine, isoleucine and valine. This forms a hydrophobic *seam* against which the α -helices are packed. The side chains of the residues at position *a* and *d* form *knobs* and *holes*, respectively. The *knobs* of one α -helix fit into the *holes* of its neighbouring α -helix and vice versa and is therefore called *knobs-into-holes* packing.

Ideal coiled coils can be described by 13 independent parameters (Table 1, Figure 2 and Figure 3). In experimental structures of coiled coils, the parameters may vary along the structure and are best analysed in a per-residue fashion (Strelkov and Burkhard, 2002). The heptad repeat can have discontinuities, known as *stutter*, *stammer* and *shift*. A *stutter* corresponds to a four-residue insert into a sequence with a continuous heptad, e.g. **abcdefg↓defgabcdefg**. The arrow in the sequence marks the discontinuity. A *stammer* corresponds to a three-residue insert. A *shift* is a one-residue insert and is equivalent to two consecutive stutters (Brown *et al.*, 1996).

2. Assignment of the heptad core positions

Two algorithms have been described for the assignment of the heptad core positions, known as *SOCKET* (Walshaw and Woolfson, 2001) and *TWISTER* (Strelkov and Burkhard, 2002). Both are required to obtain an accurate assignment. The first algorithm identifies knobs-into-holes interactions along two or more α -helices and thus the alignment of the helices. It also assigns the correct core position *a* or *d* at each *complementary* knobs-into-holes interaction (see below). However, not every core position forms a complementary knob-into-holes interaction in natural coiled coils. The *SOCKET* algorithm cannot assign such core positions. In the *SOCKET* program, the missing positions are completed such that there results a continuous heptad position assignment. But in the presence of

stutters, this procedure leads to a discontinuity at the (coincidental) location of the next complementary knobs-into-holes interaction rather than in the region of the actual geometrical distortion.

The *TWISTER* algorithm, in contrast, considers the geometry of the α -helices and assigns each core position correctly. Stutters are marked by discontinuities in the assignment that appear in the region of the actual distortion. However, the alignment of the helices must be known beforehand.

In both algorithms, the assignment of the peripheral positions is done after the core positions have been assigned. They are always assigned *in phase* to the last previous core position. Positions which cannot be assigned (e.g. in a stutter region) are denoted z. The algorithms never disagree about the assignment of core positions at the location of a complementary knobs-into-holes interaction.

2.1. SOCKET algorithm

The side chains of the residues in α -helices are represented as points *sc* in space. A point

sc is the geometric centre of the side chain atoms starting from the C_{β} atom:
$$sc = \frac{1}{n} \sum_i^n a_i,$$

where n is the number of side chain atoms and a_i the coordinate of the side chain atom i . Hydrogen atoms are ignored. For glycine, which has no side chain, the coordinate of the C_{α} atom is used as *sc*. A knobs-into-holes interaction is present if a side chain k_X (knob) of α -helix X resides within the packing cut-off distance d of four side chains h_{Y1} , h_{Y2} , h_{Y3} and h_{Y4} (hole) of the neighbouring α -helix Y . If there are more than four side chains of helix Y within the distance d , the four nearest side chains form the hole. The distance d was empirically determined to 7.0 Å. A knobs-into holes interaction is known as *complementary* when a side chain h_Y which is part of a hole in helix Y is itself a knob k_Y that fits into a hole formed by h_{X1} , h_{X2} , h_{X3} and h_{X4} on helix X , one of which acts as the previously mentioned knob k_X . A pair of helices is defined as *coiled coil* if there are at least *two* complementary knobs-into holes interactions present between them. Multi-stranded coiled coils are defined by the presence of at least *one cyclically complementary* knobs-into-holes interaction. In a cyclically complementary knobs-into-holes interaction a side chain k_X on helix X fits into a hole formed by four residues h_{Y1} , h_{Y2} , h_{Y3} and h_{Y4} of helix Y . However, one of these residues h_Y serves as a knob that fits into the third helix Z instead of helix X . If the coiled coil is three stranded, one of the hole-forming residues h_Z in helix Z fits as a knob k_Z into a hole on helix X which comprises the side chain k_X . The scheme is analogously applied to four and five-stranded coiled coils.

The hole-forming side chains h_1 , h_2 , h_3 and h_4 are consecutively numbered from N- to C-terminus of the sequence. h_1 and h_4 form the top and bottom of the hole and h_2 and h_3 the side. In the case of parallel helices, if h_3 is a knob, it resides at core position a . If h_2 is a knob, it is located at core position d . In the case of antiparallel strands the assignment is the opposite: if h_2 is a knob, it resides at core position a and if h_3 is a knob, it is located at a core position d .

2.2. TWISTER algorithm

First, the Crick angle α_n is calculated for each residue n (Table 1). A residue n is assigned to be at core position a if $\alpha_{n-1} < 0$, $\alpha_n > 0$ and $|\alpha_{n-1}| > |\alpha_n|$, or at core position d if $\alpha_n < 0$, $\alpha_{n+1} > 0$ and $|\alpha_n| < |\alpha_{n+1}|$. After the core positions have been defined, the two residues following the position a are assigned to be b and c , and the three residues following the position d are assigned to be e , f and g . The remaining residues are assigned to be z (*i.e.* position undefined).

3. Factors determining the stability of coiled coils

The factors which contribute to the stability of coiled coils can be divided into two classes of interactions. The first class stabilises the individual α -helix and acts therefore intrahelically. The second class of interactions contributes to the oligomerisation of the coiled coils and acts therefore interhelically.

3.3. Intrahelical interactions

Advantageous for the formation of a coiled coil is a stable monomeric α -helix. The coiled coil must therefore consist of residues with high α -helical propensities. The α -helical propensities of each individual amino acid are listed in Figure 3.

The monomeric α -helix is further stabilised by intrahelical ionic interactions (Burkhard *et al.*, 2000a; Spek *et al.*, 1998). These interactions can be of type i to $i+3$ or i to $i+4$ where i denotes the position of the interacting amino acids in the primary sequence. It has been proposed that an i to $i+3$ Glu-Arg salt bridge is energetically more favourable than an i to $i+3$ Arg-Glu salt bridge, and that an i to $i+4$ Arg-Glu arrangement would be preferred to an i to $i+4$ Glu-Arg ionic interaction, due to the geometry of the α -helix and the side chains (Burkhard *et al.*, 2000b).

In Chapter 7 the intrahelical ionic interactions in coiled coils are discussed in detail. However, the results discussed there were not yet available at the time when the peptides described in Chapter 5 and Chapter 6 were designed.

3.4. Interhelical interactions

The most important factor contributing to coiled-coil stability is the hydrophobic interactions at positions *a* and *d* along the helical interface. Figure 4 and Figure 5 show the stability difference values resulting from the amino acid substitution of alanine in a dimeric or trimeric parallel coiled coil. The values were obtained from two different peptide systems (Acharya *et al.*, 2002; Tripet *et al.*, 2000). The difference between the values obtained from the two systems can be very large, especially for isoleucine at an *a* position (2.65 kcal/mol) and leucine at a *d* position (2.70 kcal/mol). The reasons for these deviations remain unclear.

Interhelical ionic interactions also contribute to coiled-coil stability. The rod domain of cortaxillin I from *Dictyostelium discoideum* which forms a parallel two-stranded coiled coil, contains a network of distinct ionic interactions (Burkhard *et al.*, 2000a). These salt bridges can be classified as *g to a'*, *g to e'*, *d to a'* and *d to e'* interactions. For the parallel 2-stranded coiled coils in the protein database, the *g to e'* salt bridge is by far the most common (Figure 6).

4. Conclusion

Ever since they have been predicted in 1953 by (Crick, 1953), coiled coils have been extensively studied due to their simplicity and regularity compared to other structural motifs. The factors contributing to the stability and formation of coiled coils involve intrahelical interactions which stabilise the monomeric α -helix, and interhelical interactions which determine and stabilise their oligomerisation. In recent years, algorithms have been developed which allow automated identification of coiled-coil motifs, heptad position assignment and per-residue calculation of coiled-coil parameters from structural data. Thus, it has become possible to apply statistical methods to the coiled coils in the protein database to identify their common features and differences. The wealth of structural information in the database together with kinetic data can now be exploited to understand the principles of coiled-coil formation and to improve the design of artificial coiled coils, opening a wide range of applications.

5. Acknowledgements

I would like to thank Carmen Chan for proof-reading this chapter.

6. References

- Acharya, A., Ruvinov, S.B., Gal, J., Moll, J.R. and Vinson, C. (2002) A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. *Biochemistry*, 41, 14122-14131.
- Brown, J.H., Cohen, C. and Parry, D.A. (1996) Heptad breaks in alpha-helical coiled coils: stutters and stammers. *Proteins*, 26, 134-145.
- Burkhard, P., Kammerer, R.A., Steinmetz, M.O., Bourenkov, G.P. and Aebi, U. (2000a) The coiled-coil trigger site of the rod domain of cortexillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Structure Fold Des*, 8, 223-230.
- Burkhard, P., Meier, M. and Lustig, A. (2000b) Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci*, 9, 2294-2301.
- Chakrabarty, A. and Baldwin, R.L. (1995) Stability of alpha-helices. *Adv Protein Chem*, 46, 141-176.
- Crick, F.H.C. (1953) The packing of α -helices: Simple coiled coils. *Acta crystallographica*, 6, 689-697.
- Spek, E.J., Bui, A.H., Lu, M. and Kallenbach, N.R. (1998) Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci*, 7, 2431-2437.
- Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J. and Kammerer, R.A. (2000) Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. *Nat Struct Biol*, 7, 772-776.
- Strelkov, S.V. and Burkhard, P. (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J Struct Biol*, 137, 54-64.
- Tripet, B., Wagschal, K., Lavigne, P., Mant, C.T. and Hodges, R.S. (2000) Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position "d". *J Mol Biol*, 300, 377-402.
- Walshaw, J. and Woolfson, D.N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol*, 307, 1427-1450.
- Zimm, B.H. and Bragg, J.K. (1959) Theory of Phase Transition between Helix and Random Coil in Polypeptide Chains. *Journal of Chemical Physics*, 31, 526-535.

Figures and Tables

α-helical radius	Radius of the α -helix	$r_n = A_n O_n $
α-helical rise	Distance between adjacent residues along the α -helix axis	$h_n = \frac{ O_{n-1} O_n + O_n O_{n+1} }{2}$
initial phase of α-helix	Starting angle of the α -helix	j_0
α -helical phase yield	Angle resulting from the turn of the α -helix from one residue to the next	$\Delta\varphi_n = \frac{\omega_{A_{n-1}O_{n-1}O_nA_n} + \omega_{A_nO_nO_{n+1}A_{n+1}}}{2}$
α-helical pitch	Distance along the axis of the α -helix which corresponds to an exact 360° turn of the α -helix	$p_n = \frac{2\pi \times h_n}{\Delta\varphi_n}$
Number of residues per α -helical turn		$\frac{2\pi}{\Delta\varphi_n}$
Superhelical multiplicity	Number of α -helices forming the coiled coil	m
Coiled-coil radius	Radius of the superhelix	$R_n = \frac{\sum O_n^m C_n }{m}$
Coiled-coil rise	Distance between adjacent residues along the superhelix axis	$H_n = \frac{ C_{n-1} C_n + C_n C_{n+1} }{2}$
Initial phase of superhelix	Starting angle of the superhelix	Φ_0
Superhelical phase yield	Angle resulting from the turn of the superhelix from one residue to the next	$\Delta\Phi_n = \frac{\sum \frac{\omega_{O_{n-1}^m C_{n-1} C_n O_n^m} + \omega_{O_n^m C_n C_{n+1} O_{n+1}^m}}{2}}{m}$
Coiled-coil pitch	Distance along the axis of the superhelix which corresponds to an exact 360° turn of the superhelix	$P_n = \frac{2\pi \times H_n}{\Delta\Phi_n}$
Crick phase	Residual α -helical phase defined relative to a coordinate frame rotating with the superhelix. The sign of the angle is given by the direction of the helix axis which is positive going from N- to C-terminus.	$\alpha_n = \nu_{C_n O_n A_n}$
Crossing angle	Crossing angle of the helices m and m' at residue n	$\frac{\omega_{O_n^m O_{n-1}^m O_{n-1}^{m'} O_n^{m'}} + \omega_{O_{n+1}^{m'} O_n^{m'} O_n^m O_{n+1}^m}}{2}$

Table 1: Coiled-coil parameters. Coiled coils can be described solely by 13 independent parameters. Seven of them are in the table and typed in bold. The other six independent parameters position and orient the coiled-coil axis in space and are not in the table. The remaining parameters in the table are derived. The parameters vary along the coiled coil and are calculated per residue n . A , O and C signify points in space (see Figure 1 and Figure 2 for their definition). ν_{ABC} represents the angle formed by the points A, B, C . ω_{ABCD} represents the dihedral angle formed by the points A, B, C and D . $|AB|$ denotes the distance between the points A and B .

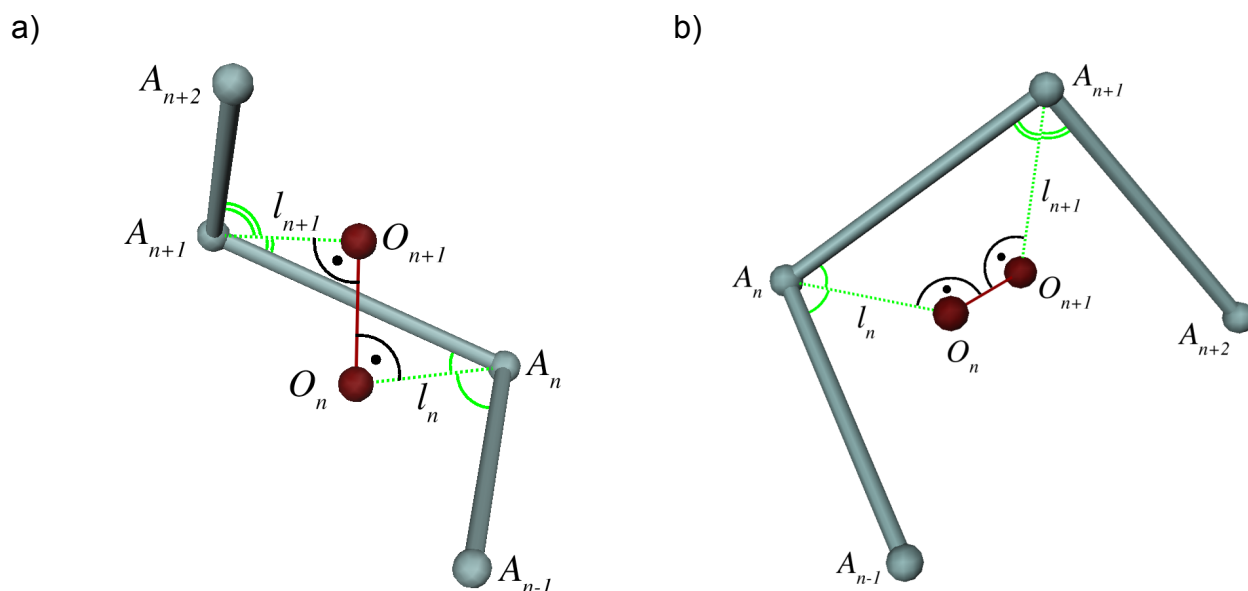


Figure 1: Construction of the local α -helix axis. The helix axis is the spline line formed by points O_n such that an individual point O_n corresponds to residue n . The point A_n is the C_α position of the residue n . The points O_n and O_{n+1} are constructed on the bisection l_n and l_{n+1} of the angles $A_{n-1}A_nA_{n+1}$ and $A_nA_{n+1}A_{n+2}$, respectively, such that the line O_nO_{n+1} is perpendicular to both l_n and l_{n+1} . a) side view, b) top view.

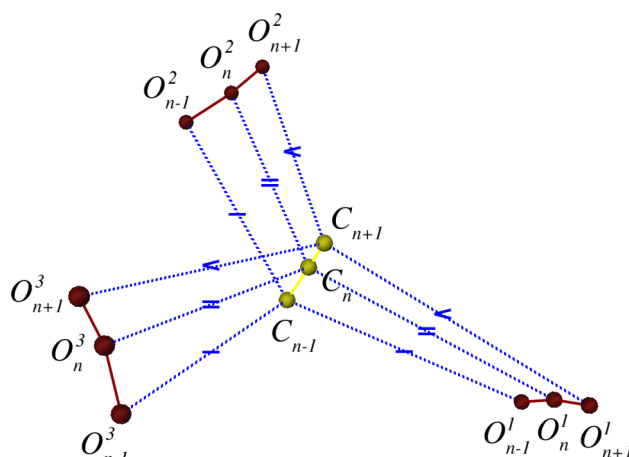


Figure 2: Construction of the local coiled-coil axis. The axis is formed by points C_n which correspond to residue n . The axis points C_n are defined as the geometric mean of the α -helix axis points O_n^m where m denotes the m^{th} α -helix of the coiled coil.

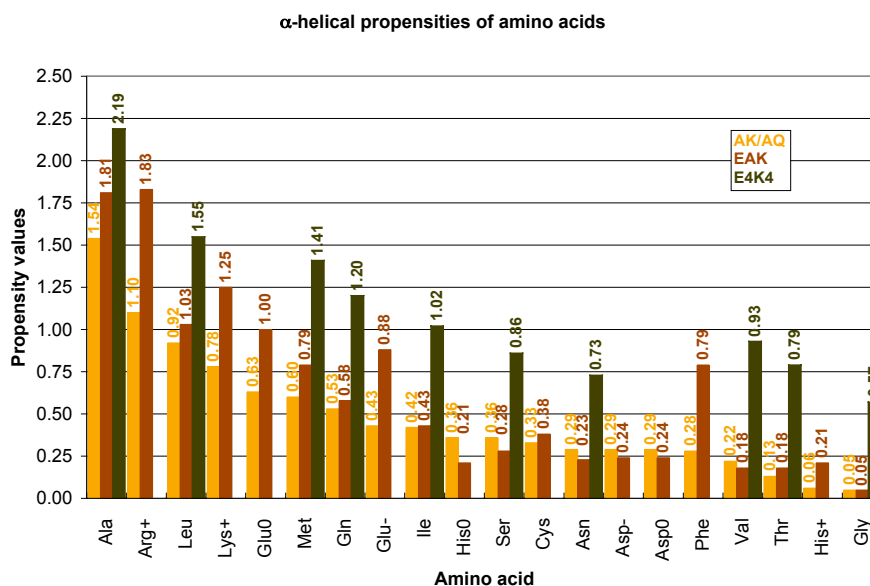


Figure 3: α -helical propensity of each biogenic amino acid. Values are obtained from (Chakrabarty and Baldwin, 1995). The values represent the helix propagation parameter s from the Zimm-Bragg theory (Zimm and Bragg, 1959) and were obtained from different peptide systems. The system of $(AAKAA)_n$ and $(AAQAA)_n$ is denoted AK/AQ, the system Ac-YEAAAKEAXAKEAAKA-NH₂ is called EAK, and the system Ac-YSEEEEEKKKXXXEEEEKKK-NH₂ is denominated E4K4. The values are highly correlated between the different peptide systems, but they deviate in the absolute scale. Amino acids with $s > 1$ are classified as helix forming, $s \approx 1$ as helix indifferent and $s < 1$ as helix breaking. For a detailed discussion see (Chakrabarty and Baldwin, 1995).

Coiled coil stability differences of amino acid substitutions at an α position

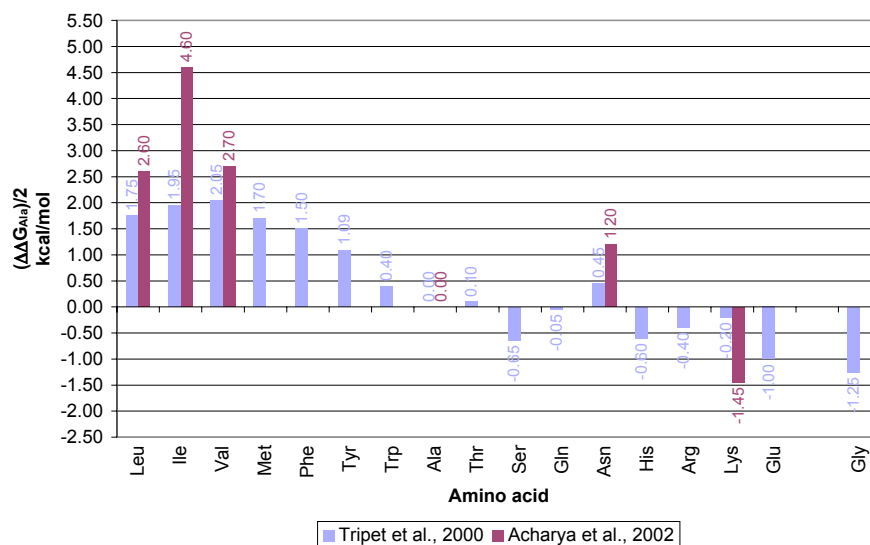


Figure 4: Differences of coiled-coil stability resulting from the substitution of alanine at core position α . The data were obtained from two different peptide systems (Acharya *et al.*, 2002; Tripet *et al.*, 2000) and represent hydrophobicity and packing interactions of the substituted residues within a two-stranded or three-stranded parallel coiled coil. The peptide systems have been designed such that ionic interactions between the substituted residue and its neighbouring residues are avoided. Positive values indicate a gain and negative numbers a loss of stability.

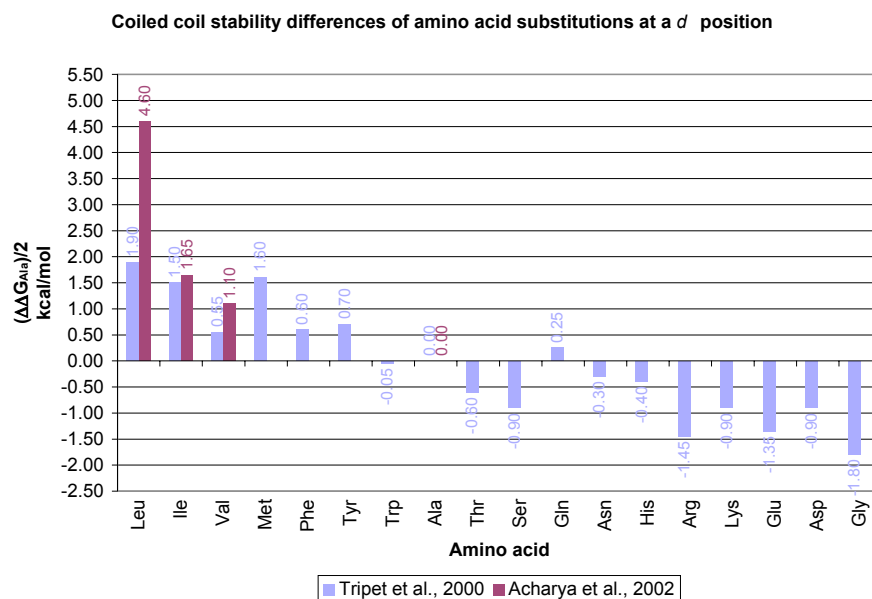


Figure 5: Differences of coiled-coil stability resulting from the substitution of alanine at core position *d*. The data were obtained from two different peptide systems (Acharya *et al.*, 2002; Tripet *et al.*, 2000) and represent hydrophobicity and packing interactions of the substituted residues within a two-stranded or three-stranded parallel coiled coil. The peptide systems have been designed such that ionic interactions between the substituted residue and its neighbouring residues are avoided. However in the case of aspartic acid, a hydrogen bond between its side chain and a neighbouring asparagine side chain might have formed, falsifying the value. Positive values indicate a gain and negative numbers a loss of stability.

Interhelical salt bridges in parallel 2-stranded coiled coils

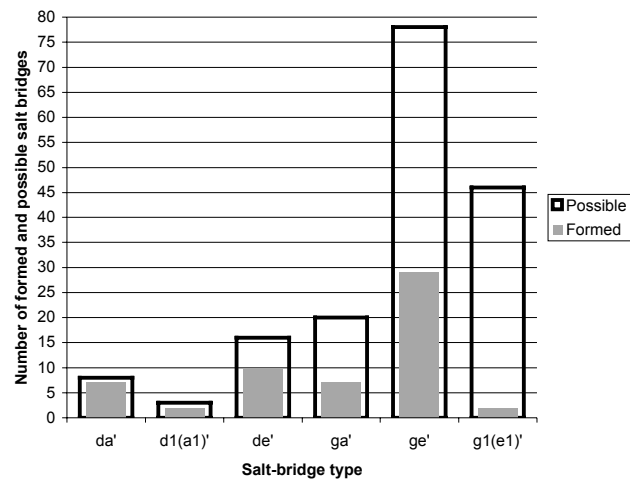


Figure 6: Interhelical salt bridges in parallel 2-stranded coiled coils. The salt bridges were identified by the program *SBSCC* described in Chapter 7. A distance cut-off of 3.8 Å and a side-chain *B*-factor cut-off of 80 Å² was used to define a formed salt bridge. The salt-bridge types are termed by their heptad positions and relative numbers, *i.e.* in the classical *ge'* ionic interaction the first residue resides at a *g* position in the first strand and the second residue at an *e* position shifted by 1 heptad towards the *C*-terminus in the second strand (synonym *g1(g2)'*). In the non-classical salt-bridge type *g1(e1)'* the residues reside in adjacent heptads.

Chapter 5

Design of a minimal protein oligomerization domain by a structural approach

Burkhard, P., Meier, M. and Lustig, A. (2000) Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci*, **9**, 2294-2301.

I contributed with CD-spectroscopy to this chapter. All crystallographic work was done by P. Burkhard and ultracentrifugation by Ariel Lustig.

ACCELERATED COMMUNICATION

Design of a minimal protein oligomerization domain by a structural approach

PETER BURKHARD,¹ M. MEIER,¹ and ARIEL LUSTIG²

¹M.E. Muller Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

²Department of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

(RECEIVED August 15, 2000; FINAL REVISION October 6, 2000; ACCEPTED October 9, 2000)

Abstract

Because of the simplicity and regularity of the α -helical coiled coil relative to other structural motifs, it can be conveniently used to clarify the molecular interactions responsible for protein folding and stability. Here we describe the de novo design and characterization of a two heptad-repeat peptide stabilized by a complex network of inter- and intrahelical salt bridges. Circular dichroism spectroscopy and analytical ultracentrifugation show that this peptide is highly α -helical and 100% dimeric under physiological buffer conditions. Interestingly, the peptide was shown to switch its oligomerization state from a dimer to a trimer upon increasing ionic strength. The correctness of the rational design principles used here is supported by details of the atomic structure of the peptide deduced from X-ray crystallography. The structure of the peptide shows that it is not a molten globule but assumes a unique, native-like conformation. This de novo peptide thus represents an attractive model system for the design of a molecular recognition system.

Keywords: coiled coil; crystal twinning; ionic interactions; protein de novo design; protein folding; protein oligomerization; salt bridge

Gaining an understanding of the relationship between the amino acid sequence of a protein and its structure poses a challenging problem, primarily because proteins have such highly complex structures. However, recent progress in the de novo design of peptides has achieved significant progress in the construction of protein structures that assume a predefined fold. Harbury et al. (1998) have successfully designed a model peptide that assumes a right-handed coiled-coil conformation. Such a fold was predicted for the extremely thermostable protein tetrabrachion (Peters et al., 1996) but has only recently been proven to be a right-handed coiled coil (Stetefeld et al., 2000). Not only α -helical proteins but also β -sheet containing peptides (Kortemme et al., 1998), diiron proteins (Lombardi et al., 2000), or proteins with a specific function (Baltzer et al., 1999) have recently been designed.

Despite this considerable progress in protein design, it is still difficult to design proteins that fold into predetermined three-dimensional structures. The solution to this problem was attempted via quantifying the relative energetic contributions of short- and long-range interactions in simplified model systems such as small polypeptides (Struthers et al., 1996). These independently folded polypeptide motifs, in turn, have been successfully used as tem-

plates for construction of mini-proteins that were shown to exhibit distinct structural and functional properties (Severin et al., 1997; Baltzer et al., 1999).

The principle of short- and long-range interactions can also be found in the coiled-coil structural motif, which has been investigated in more detail, mainly by modifying the sequence of bZIP proteins (Harbury et al., 1993; Gonzalez et al., 1996; Moitra et al., 1997). Coiled coils consist of two to five amphipathic α -helices that twist around one another to form a multistranded supercoil (Lupas, 1996). Sequences of parallel left-handed coiled-coil proteins are characterized by a heptad repeat pattern of seven amino acids denoted **a** to **g** harboring mostly apolar residues in the **a** and **d** positions. The stability of the coiled coil is achieved by the systematic packing of the side chains of the amino acids in the **a** and **d** positions along a hydrophobic seam. This is called “knobs-into-holes” packing and was first postulated by Crick (1953). It has also been shown that distinct coiled-coil trigger sites within heptad-repeat-containing amino acid sequences may be necessary to mediate coiled-coil formation (Kammerer et al., 1998; Steinmetz et al., 1998). The coiled-coil trigger site of cortexillin contains a distinct pattern of inter- and intrahelical salt bridges in addition to the hydrophobic interactions occurring along the dimer interface (Burkhard et al., 2000).

The design principles for producing the shortest possible coiled-coil peptide—ideally only two heptad-repeats long—can be di-

Reprint requests to: Peter Burkhard, M.E. Müller Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

vided into factors that contribute to either its monomeric α -helical stability (short-range interactions) or to its oligomeric packing (long-range interactions). The success of such a design depends on how correctly the underlying principles of coiled-coil formation and protein de novo design have been understood (see also Cohen & Parry, 1990; Munoz & Serrano, 1995). Hence, structural information is of great importance both to verify the correctness of the design and to gain insight into the atomic details of current models to correlate kinetic data with structural information. These insights can then guide the future design of de novo designed peptides with distinct structural and functional properties.

The small size of such coiled-coil peptides offers practical advantages: the newly engineered molecules can be produced efficiently and rapidly by chemical syntheses and can easily be manipulated chemically. Therefore, it is an attractive model system to study the factors responsible for protein folding and stability. Also it may be efficiently used to design a simple molecular recognition motif for applications, such as a drug targeting system (Trail & Bianchi, 1999), a protein purification and detection system (Tripet et al., 1996), and it can also be employed for the design of hybrid hydrogels (Wang et al., 1999) or biosensors (Chao et al., 1998).

Results and discussion

Factors contributing to monomeric α -helix stability

Obviously, a highly α -helical peptide would contain α -helix favoring amino acid residues, i.e., residues with a high α -helical propensity (for a review see Chakrabarty & Baldwin, 1995). In our work we have mainly focused on incorporating intrahelical salt bridges as the key stabilizing factor for monomeric α -helices (Spek et al., 1998; Burkhard et al., 2000). So far, these have mostly been classified into only two types, namely, i to $i + 3$ and i to $i + 4$ intrahelical salt bridges. Considering the geometry of an α -helix and the different lengths of the side chains of different residues, it becomes clear that an i to $i + 3$ Glu-Arg salt bridge should be energetically more favorable than an i to $i + 3$ Arg-Glu arrangement, and that an i to $i + 4$ Arg-Glu arrangement would be preferred over the i to $i + 4$ Glu-Arg salt bridge. In addition, the helix dipole would favor negatively charged side chains oriented toward the N-terminus and positively charged side chains toward the C-terminus, thus favoring positively charged residues at positions preceding the negatively charged side chains. In our design, we have preferred Arg residues over Lys residues as positively charged residues because Arg may form two charged hydrogen bonds to the carboxyl group of a Glu residue and because the guanidinium group of Arg has a decreased flexibility compared to the side chain of Lys and hence is entropically more favorable than Lys (see also Merutka & Stellwagen, 1991; Huyghues-Despointes et al., 1993). Indeed, Lys to Arg substitutions can be observed in thermophilic enzymes (Fagain, 1995) and Arg seems to be more often involved in salt bridges in coiled coils than Lys (Musafia et al., 1995; Burkhard et al., 2000).

An optimal helix capping motif was designed to compensate for the helix dipole and to generate an optimal terminal hydrogen bonding network (Doig & Baldwin, 1995; Gong et al., 1995; Lu et al., 1999). Amidation at the C-terminus and succinylation at the N-terminus of the peptides removes or even inverts the unfavorable terminal charges with regard to the helix dipole and can establish a favorable hydrogen bonding network at the N-terminus.

Finally, negatively charged residues were positioned near the N-terminus while positively charged residues were placed near the C-terminus to take account of the helix dipole (Morgan & Mayo, 1998).

Factors contributing to oligomeric α -helix stability

Hydrophobic interactions occurring along the coiled-coil interface are believed to be the key driving force for the stability of the coiled coil. Stabilization of a parallel two-stranded coiled coil can be achieved, for example, with Ile in **a** positions and Leu in **d** positions (Harbury et al., 1993). While results from experimental measurements on the contribution of interhelical ionic interactions to coiled-coil stability have not been conclusive (Lavigne et al., 1996; Lumb & Kim, 1996), the recent discovery that interhelical salt bridges including charged residues even at **a** positions are a key feature of the coiled-coil trigger site of corticillin I (Burkhard et al., 2000) point to the importance of interhelical salt bridges for coiled-coil formation and stability (Krylov et al., 1994). The helix dipole will favor negatively charged side chains oriented toward the N-terminus and positively charged side chains toward the C-terminus. Thus, the arrangement with Arg at **g** and Glu at **e'** should be energetically more favorable than the opposite arrangement. For similar reasons as discussed for the intrahelical salt bridges, Arg residues stabilize the coiled-coil structure to higher extent than Lys residues (Krylov et al., 1998).

Based on these design principles, we have de novo designed the peptide Succ-DELERRIRELEARIK-NH₂ aimed at forming a stable parallel coiled coil with a length of only two-heptad repeats. The peptide is, in principle, able to form many intrahelical salt bridges in addition to an optimal **g-e'** Arg-Glu interhelical salt bridge (Fig. 1E).

Biophysical properties of the peptide

The oligomerization state of the peptide was determined by analytical ultracentrifugation (Table 1). In sedimentation equilibrium measurements, the peptide was found to be 93% dimeric under physiological conditions (150 mM NaCl, pH 7.0) at room temperature. Even though these results have an average error of measurement of about 20%, one can say that with higher temperatures and/or lower peptide concentration the dimer is slightly less stable. This is in agreement with the circular dichroism (CD) measurements (see below). Interestingly, when increasing the ionic strength to 2 M NaCl, the peptide switches its oligomerization state from the dimeric to the trimeric state despite the fact that its core residues were expected to favor the dimeric state (Harbury et al., 1993; Tripet et al., 2000). Such a transition between oligomerization states has not been reported before and is similar to the stabilization of a trimeric coiled coil in response to binding of a hydrophobic ligand (Gonzalez et al., 1996). Most likely this transition is due to a better shielding of the hydrophobic side chains from the very polar environment in a trimeric coiled coil compared to the situation in a dimer even though the packing of the side chains would better accommodate a dimeric structure (Harbury et al., 1993). In contrast, when replacing the Leu residues in the **d** positions of the peptide by Ile, the peptide stays dimeric under low and high ionic strength conditions (P. Burkhard, in prep.). This is again unexpected as previous experiments would predict a trimeric state (Harbury et al., 1994; Tripet et al., 2000). This clearly indicates that specific residues at specific heptad repeat positions

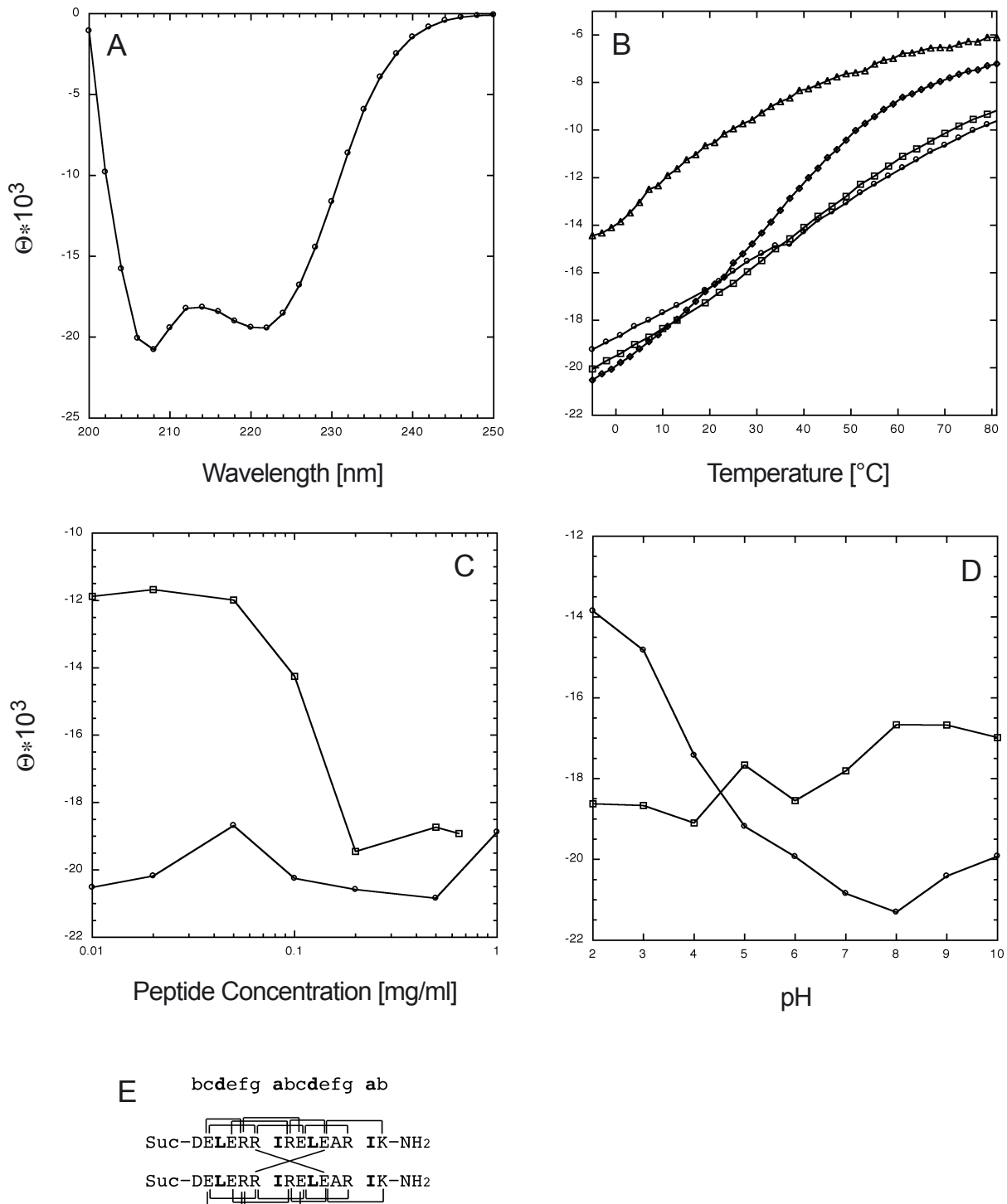


Fig. 1. CD analysis of the peptide. **A:** CD spectrum of the peptide. The peptide was analyzed in 5 mM sodium phosphate buffer (pH 7.0) and 150 mM NaCl at 0 °C. **B:** Melting curve of the peptide under different ionic strength conditions and at different pH monitored by the CD signal change at 222 nm, $[\Theta]_{222}$. The peptide concentration was always 1 mg/mL. (○) pH 7.0, 0 M NaCl; (□) pH 7.0, 0.15 M NaCl; (◆) pH 7.0, 2 M NaCl; (△) pH 2.0, 0 M NaCl. **C:** Concentration dependence of the peptide at 0 M NaCl (○) and 2 M NaCl (□) ionic strength as monitored by the CD signal change at 222 nm, $[\Theta]_{222}$. The peptide was analyzed in 10 mM Tris buffer (pH 7.0) at 0 °C. **D:** pH dependence of the peptide at 0 M NaCl (○) and 2 M NaCl (□) ionic strength as monitored by the CD signal change at 222 nm, $[\Theta]_{222}$. **E:** Sequence of the peptide with the possible intra- and interhelical saltbridges depicted as brackets and the hydrophobic core residues in bold.

Table 1. Oligomerization state of the peptide as observed by analytical ultracentrifugation^a

NaCl (M)	Peptide concentration (mg/mL)	Temperature (°C)	Mw-AUC ^b (kDa)	Oligomerization state ^c
0.15	1.0	5	3.5	73% dimer ^d
0.15	1.0	20	3.9	93% dimer ^d
0.15	1.0	30	3.3	63% dimer ^d
0.15	0.3	5	3.7	83% dimer ^d
0.15	0.3	20	3.5	73% dimer ^d
0.15	0.1	20	3.3	63% dimer ^d
2.0	0.1	20	5.6	88% trimer ^e

^aThe peptide was analyzed in 10 mM Tris buffer at pH 7.0 in sodium chloride at different concentrations and temperatures.

^bObserved molecular mass.

^cThe oligomerization state is given in percentages of dimers or trimers as calculated from the “observed” molecular mass relative to the “real” molecular mass.

^dAssuming a monomer/dimer equilibrium.

^eAssuming a monomer/trimer equilibrium.

cannot determine the oligomerization state per se, but that the oligomerization state is also dependent on the remainder of the sequence as well as on the buffer conditions.

According to the CD spectrum, the peptide is highly α -helical (Fig. 1A). Thermal melts of the peptide under different ionic strength conditions show a destabilizing effect of high ionic strength at higher temperature (Fig. 1B). The unfolding profile of the peptide at high ionic strength shows a higher degree of cooperativity, indicating that the unfolding of the peptide is a two-state process in which the α -helices are much more stable when they are in the coiled-coil conformation (Thompson et al., 1993). At low ionic strength, the thermal melting curve is much less cooperative and shows also at higher temperature considerable α -helicity, indicating that the monomeric α -helices themselves are very stable due to the many stabilizing intrahelical salt bridges. Also, the α -helicity of the peptide is concentration dependent at high ionic strength conditions while at low ionic strength it is not (Fig. 1C). This does not necessarily mean that the coiled coil is more stable at low ionic strength but rather that the CD signal cannot discriminate between dimers and monomers since the monomer helices are very stable at low ionic strength. This is different at high ionic strength conditions where a transition between two different states at about 0.1 mg/mL can be observed, presumably corresponding to the transition from the trimeric to the monomeric state. This corresponds to a K_d of $\sim 50 \mu\text{M}$ for coiled-coil formation.

Most remarkably, the α -helicity of the peptide is highest around neutral pH (Fig. 1B,D). This is in sharp contrast to previous observations that coiled coils containing Glu residues are more stable at low pH than at neutral pH, despite the loss of ion pairs by protonation of acidic residues at low pH (Lumb & Kim, 1995). The intrinsically higher stabilization of the coiled coil by protonated Glu has been attributed to both its higher α -helical propensity (Chakrabarty & Baldwin, 1995) and more hydrophobic character, which allows it to better pack at the dimer interface (Kohn et al., 1995). The designed peptide, however, is most stable (at low ionic strength) when its side chains are charged, indicating that the ionic interactions stabilize the α -helical conformation substantially (Fig. 1D). Furthermore, at high ionic strength and neutral pH, the

coiled coil is less stable, because the contribution of the ionic interactions to coiled-coil stability is reduced due to the increased shielding of the charged side chains. This implies that the loss of stability due to shielding of the charged side chains dominates the gain of stability due to the strengthened hydrophobic interaction. This is further supported by the fact that at low pH the melting curve is nearly concentration independent (data not shown), indicating a monomolecular reaction, i.e., at low pH the peptide is mostly in its monomeric state. Hence, the ionic interactions do in fact contribute considerably to coiled-coil stability (Krylov et al., 1994; Spek et al., 1998; Burkhard et al., 2000; compare also Lavigne et al., 1996; Lumb & Kim, 1996).

Structure of the peptide

The crystal structure of the peptide at 1.2 Å resolution contains four three-stranded coiled coils in the asymmetric unit. According to the design principles employed, the peptide was found to form a parallel coiled coil (Fig. 2; Table 2). Apart from the hydrophobic packing of the core residues, the designed interhelical salt bridge of the **g-e'** type between residues Glu6 and Arg11' appears to be the prominent and conserved feature of the structure. In all 12 possible positions within the asymmetric unit, this particular interhelical salt bridge is formed with optimal interatomic distances (Figs. 2, 3A). Moreover, the side chains of the residues involved have very low temperature factors that are comparable to those of the side chains of the core residues (Fig. 2B). This indicates that the side chains of this interhelical salt bridge are in a unique conformation. In all four trimers within the asymmetric unit, one additional interhelical salt bridge per trimer between Glu9 and Lys15' of the type **c-b'** is formed (Fig. 3B).

Several intrahelical salt bridges according to the design principles have been found including those between Glu2 and Arg6 (Fig. 3A). None of them, however, is found in all 12 places within the asymmetric unit. The peptide was crystallized with a surprisingly low solvent content of only 25%, implying a very dense packing of the four trimers in the asymmetric unit (Fig. 4). There are many salt bridges formed between neighboring molecules within the asymmetric unit as well as to symmetry related molecules and also to bound sulfate ions. These “crystal-packing” salt bridges are in competition with the salt bridges within one trimer. The fact that none of the 12 interhelical **g-e'** salt bridges is involved in such “crystal-packing” salt bridges points again to the importance of this salt bridge for proper trimer formation. The intrahelical salt bridge arrangement, in contrast, exhibits considerable variability. As mentioned in the design principles, most of the intrahelical salt bridges are of the less favorable types for monomeric helix stabilization and may therefore more easily be involved in “crystal-packing” salt bridges.

The crystal packing as shown in Figure 4 exhibits nearly exact fourfold NCS symmetry parallel to the crystallographic *c*-axis through half of the unit cell. After half of the unit cell, the fourfold NCS axis switches its position to the center of the *C*-plane. This explains why this crystal form can be twinned with the twin operator (0 1 0, 1 0 0, 0 0 1). The twin fraction for this crystal as refined by the program SHELXL (Sheldrick & Schneider, 1997) was 6%. Despite the twinning of the crystal, the structure was solved by single anomalous dispersion phasing from a heavy atom derivative.

The X-ray structure also explains the high α -helix inducing potential of the N-terminal succinylation motif (Munoz & Serrano, 1995): the succinyl moiety forms as many as four hydrogen bonds

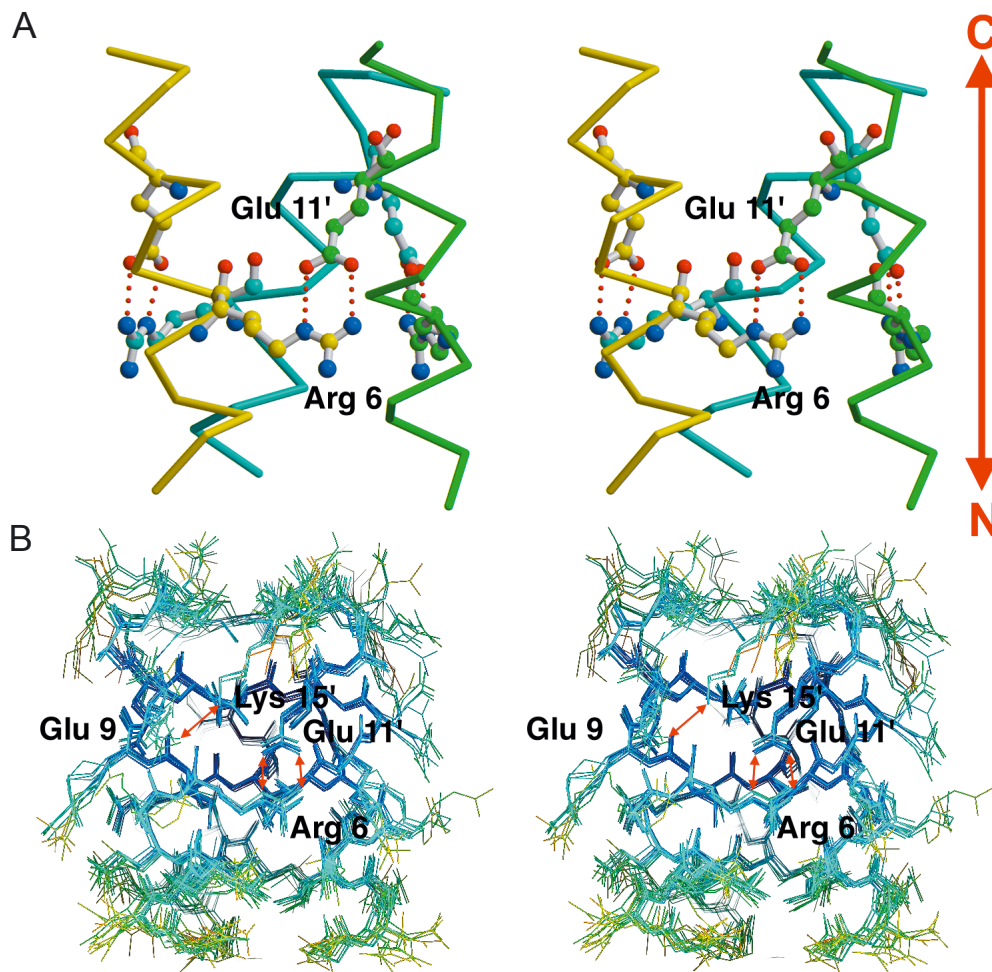


Fig. 2. Stereo view of the peptide trimer. **A:** One trimer is shown as a C_{α} trace with the interhelical salt bridges as red dotted lines and the residues Arg6 and Glu11 depicted as ball-and-stick model. **B:** In the same orientation, all 12 possible superpositions of the four trimers in the asymmetric unit are colored according to their atomic B -factors ranging from dark blue (low B -factors) to red (high B -factors). The side chains and the termini have the higher conformational variability than the peptide backbone with the exception of the side chains of Arg6 and Glu11, which form an interhelical salt bridge. Remarkably, in all 12 monomers these side chains have the same conformation and also their B -factors are in the same range as those of the side chains of the hydrophobic core residues. In four places, there exists also a c - b' interhelical salt bridge between Glu9 and Lys15'. The structures were drawn with the program O (Jones et al., 1991), MOLSCRIPT (Kraulis, 1991), and Raster3D (Merritt & Bacon, 1997).

with the N-terminal peptide nitrogens thus acting as an effective N-terminal capping box (Fig. 3C). The design of the N-terminal capping box including three negatively charged residues to compensate the positive pole of the helix dipole and a terminal succinylation that can make hydrogen bonds to the peptide nitrogen of residues 2, 3, and 4 is nicely confirmed by the X-ray structure (Fig. 3C). Even though there is considerable flexibility of the succinyl moiety, the hydrogen bonding pattern is conserved in most of 12 helices in the asymmetric unit. The C-terminus deviates from exact threefold symmetry (Fig. 3B). The three Ile14 residues all have different but well-defined side-chain conformations, and the distance between the C_{α} positions of the three Lys15 residues are all different. This asymmetry can be observed in all four trimers of the asymmetric unit and in all four trimers one additional interhelical salt bridge between Glu9 and Lys15' of the type c - b' can be found. This salt bridge, however, is only formed between

the two helices that are closest to each other. The different side-chain conformations of Ile14 allows two helices to pack closer to each other and enables the charged residues Glu9 and Lys15' to form a salt bridge (Fig. 1B). Similarly, one could also argue that the formation of a salt bridge between Glu9 and Lys15' forces the side chains of Ile14 to adopt a different conformation, which is only possible at the termini of the coiled coil. This would then also explain why this peptide is able to switch its oligomerization state to a trimer at high salt concentration despite the fact that the hydrophobic core residues (Ile at **a** and Leu at **d**) are thought to induce dimer formation (Harbury et al., 1993; Triplet et al., 2000).

Conclusions

In this work we have demonstrated that ionic interactions can improve coiled-coil stability considerably. In particular, the inter-

Table 2. Data collection, phasing, and refinement statistics

Data collection statistics	
Resolution (Å)	1.2
Observed reflections	244,675
Unique reflections	78,313
Completeness (%)	
20–2.0 Å	99.5
Overall	81.5
Phasing statistics (30.0–1.4 Å)	
R_{Cullis}^c	0.629
Phasing power ^b	1.81
Refinement statistics (20.0–1.2 Å)	
R -factor (%) ^c	15.9
R_{free} (%) ^d	22.0
Mean B -factor protein atoms (Å ²)	21.51
Mean B -factor water atoms (Å ²)	35.99
RMSD bond distances (Å)	0.010
RMSD bond angles (°)	2.5

$$^a R_{\text{Cullis}} = \frac{\sum ||F_{PH}| - |F_P + F_H||}{\sum ||F_{PH}| - |F_P||}$$

$$^b \text{Phasing power} = \left[\frac{\sum |F_H|^2}{\sum (|F_{PH}| - |F_P|)^2} \right]^{1/2}$$

$$^c R\text{-factor} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{calc}}|}$$

$$^d R_{\text{free}} = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{calc}}|}$$

helical salt bridge of the **g-e'** type with an Arg residue in **g** position and Glu in **e'** position strongly stabilizes the oligomeric structure. Together with an optimal hydrophobic seam (Leu in **d** position and Ile in **a** position) and ideal helix capping boxes, we were able to design a two-heptad repeat long peptide that forms a stable parallel coiled-coil dimer under physiological conditions. Based on these findings, we are now designing short peptides able to induce spe-

cific oligomerization states as well as a hetero-dimeric coiled coil as a molecular recognition system (Chao et al., 1998). The small size of such coiled-coil peptides offers practical advantages: the newly engineered molecules can be produced efficiently and rapidly by chemical syntheses and easily manipulated chemically. Most importantly, this simple molecular recognition motif can be used eventually to design an efficient protein purification and detection system (Tripet et al., 1996) or a drug targeting system (Trail & Bianchi, 1999). Furthermore, it can also be employed in applications such as biosensors (Chao et al., 1998) or hybrid hydrogels (Wang et al., 1999).

Materials and methods

Synthetic peptides

The peptide was purchased from Neosystem (Strasbourg, France). The purity of the peptide (>95%) had been verified by qualitative amino acid and mass spectral analyses.

Analytical ultracentrifugation

AUC was performed on an Optima XL-A analytical ultracentrifuge (Beckman Instruments, Palo Alto, California) equipped with a 12 mm Epon double-sector cell in an An-60 Ti rotor. Sedimentation equilibrium runs were performed at 5 °C at rotor speeds of 48,000 and 50,000 rpm and a peptide concentration 1.0 mg/mL. Average molecular masses were evaluated by using a floating baseline computer program that adjusts the baseline absorbance to obtain the best linear fit of $\ln(\text{absorbance})$ vs. the square of the radial distance. A partial specific volume of 0.73 mL/g was used for low ionic strength measurement and of 0.78 mL/g for the high ionic strength measurements, respectively.

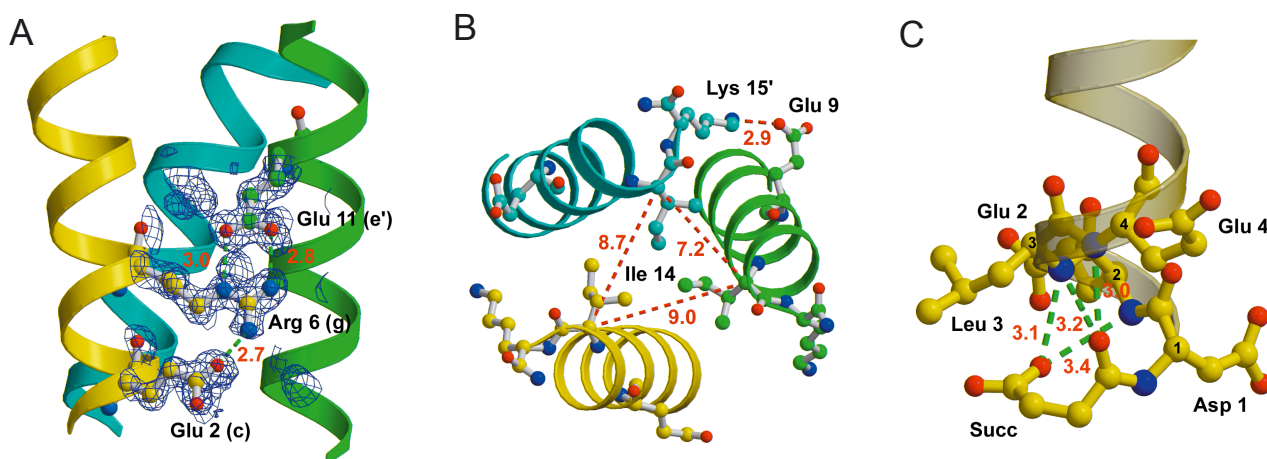


Fig. 3. Structural details of the designed peptide. **A:** A $2F_o - mF_c$ electron density of the residues involved in the highly conserved interhelical salt bridge between Arg6 **g** and Glu11 **e'** together with an i to $i + 4$ intrahelical salt bridge between Glu2 and Arg6. The interatomic distances (displayed as green dotted lines) are optimal to form hydrogen-bonded salt bridges. **B:** The C-terminus of the peptide is asymmetric. Between the two helices that are closest to each other (green and cyan), an interhelical salt bridge from Glu9 **c** to Lys15 **b'** is formed while between the other helices this salt bridge does not exist. All three Ile14 core residues have a different side-chain conformation. **C:** The N-terminal succinylation motif exhibits an extended hydrogen bonding network marked by green dotted lines. Despite the conformational variability found at the N-terminus, this hydrogen bonding pattern is found in all 12 N-termini of the structure thus supporting the high helix forming propensity of this N-terminal modification (Munoz & Serrano, 1995). The picture was drawn with the programs MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt & Bacon, 1997).

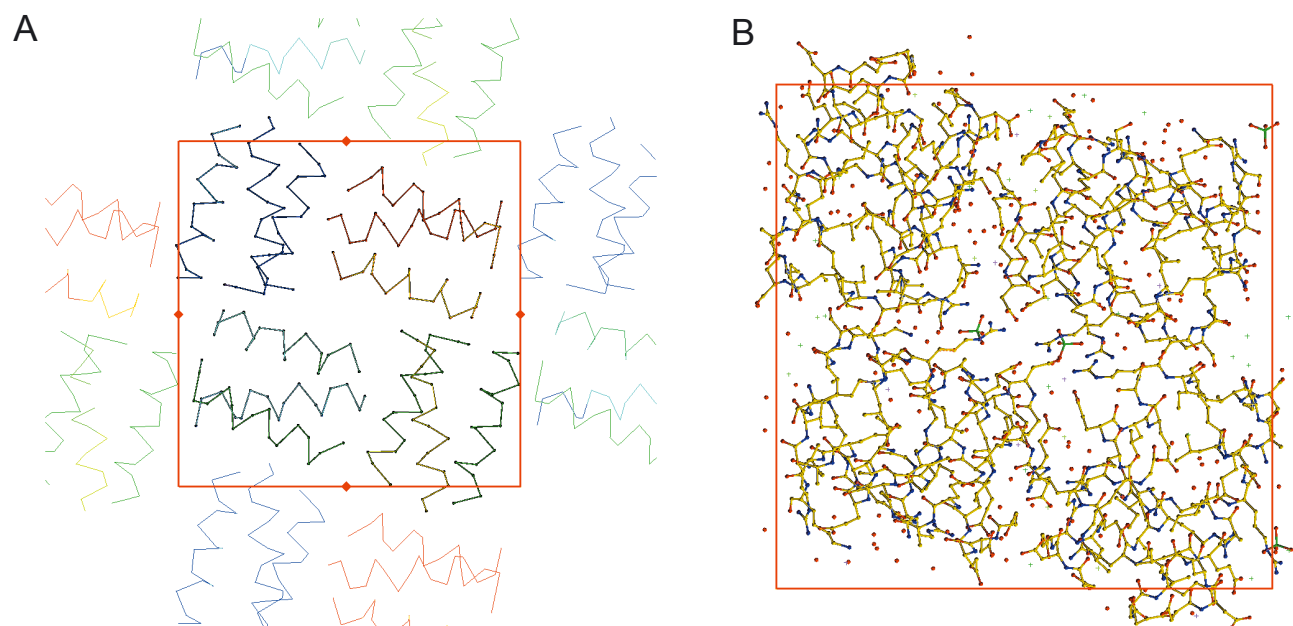


Fig. 4. Crystal packing of the peptide. **A:** The four trimers are shown as a C_{α} trace colored according to the residue number from red (low) to blue (high). The symmetry related molecules are drawn around the unit cell (thinner lines). The fourfold NCS axes are indicated by red squares. **B:** Only the four trimers of the asymmetric unit are shown with all atoms displayed colored according to their atom types. The picture was drawn with the program O (Jones et al., 1991).

CD spectroscopy

CD analysis of the synthetic peptide was performed on an Aviv 62 DS spectropolarimeter. Melting curves and ionic strength dependence were measured at 222 nm.

Crystallization

Crystals were grown in 24-well Limbro plates by vapor diffusion using the hanging-drop method. The 1 mL well solution contained 900 μL saturated ammonium sulfate, and 100 μL of 1 M Na-acetate buffer at pH 5.6. The 4 μL drop contained 2 μL peptide at a concentration of 60 mg/mL and 2 μL of well solution. Crystals grew within 14 days to cubes with dimensions up to $0.3 \times 0.3 \times 0.3 \text{ mm}^3$. The crystals belong to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions $a = 44.3 \text{ \AA}$, $b = 44.9 \text{ \AA}$, and $c = 81.0 \text{ \AA}$. There are four trimers in the asymmetric unit that gives a V_M of $1.65 \text{ \AA}^3/\text{Da}$ corresponding to a solvent content of only 25%.

Data collection, processing, and phasing

The X-ray diffraction data set from the trimethyl lead acetate (TLA) derivative was collected at 100 K on the X11 beamline (EMBL, DESY, Hamburg, Germany) at $\lambda = 0.91 \text{ \AA}$ close to the absorption edge of lead ($\lambda = 0.951 \text{ \AA}$). This single data set was used for initial phasing based on the anomalous scattering signal as well as for subsequent refinement of the structure. Data were processed, integrated, and scaled using DENZO and SCALEPACK (Otwinowski & Minor, 1997), respectively. Initial heavy atom positions were obtained from anomalous difference Patterson maps. Heavy atom parameters were refined using the program CNS (Brünger

et al., 1998). Phasing was further improved by solvent-flattening and histogram matching.

Model building and refinement

The initial atomic model was “automatically” built with the program wARP (Perrakis et al., 1997), improved using the graphics program O (Jones et al., 1991), and refined using the program SHELXL (Sheldrick & Schneider, 1997) including anisotropic B -factor and twinning refinement. The refined twin fraction with the twin operator (0 1 0, 1 0 0, 0 0 1) was 6%. No noncrystallographic symmetry restraints have been used during refinement. The final model comprises four trimers of the peptide, nine lead ions, four sulfate ions, and 161 water molecules. The final model has an R -factor of 15.9% (20.0–1.2 \AA), and the free R -factor was calculated with 5% of the native data set aside prior to refinement and is 22.0%. Root-mean-square deviations (RMSDs) from ideality in bond lengths and angles are 0.010 \AA and 2.5° , respectively (Table 2).

Accession number

Coordinates have been deposited with the Research Collaboratory for Structural Bioinformatics under the accession code 1HQJ.

Acknowledgments

We are indebted to Dr. U. Aepli for helpful discussions and all support. This work was supported by the Canton Basel-Stadt, by grants from the Swiss National Science Foundation and the Maurice E. Muller Foundation of Switzerland. SDG.

References

- Baltzer L, Broo KS, Nilsson H, Nilsson J. 1999. Designed four-helix bundle catalysts—The engineering of reactive sites for hydrolysis and transesterification reactions of p-nitrophenyl esters. *Bioorg Med Chem* 7:83–91.
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
- Burkhard P, Kammerer RA, Steinmetz MO, Bourenkov GP, Aebi U. 2000. The coiled-coil trigger site of the rod domain of cortexillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Struct Fold Des* 8:223–230.
- Chakrabartty A, Baldwin RL. 1995. Stability of alpha-helices. *Adv Protein Chem* 46:141–176.
- Chao H, Bautista DL, Litowski J, Irvin RT, Hodges RS. 1998. Use of a heterodimeric coiled-coil system for biosensor application and affinity purification. *J Chromatogr B Biomed Sci Appl* 715:307–329.
- Cohen C, Parry DA. 1990. Alpha-helical coiled coils and bundles: How to design an alpha-helical protein. *Proteins* 7:1–15.
- Crick FCH. 1953. The packing of α -helices: Simple coiled coils. *Acta Crystallographica* 6:689–697.
- Doig AJ, Baldwin RL. 1995. N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci* 4:1325–1336.
- Fagain CO. 1995. Understanding and increasing protein stability. *Biochim Biophys Acta* 1252:1–14.
- Gong Y, Zhou HX, Guo M, Kallenbach NR. 1995. Structural analysis of the N- and C-termini in a peptide with consensus sequence. *Protein Sci* 4:1446–1456.
- Gonzalez L Jr, Plescs JJ, Alber T. 1996. An engineered allosteric switch in leucine-zipper oligomerization. *Nat Struct Biol* 3:510–515.
- Harbury PB, Kim PS, Alber T. 1994. Crystal structure of an isoleucine-zipper trimer. *Nature* 371:80–83.
- Harbury PB, Plescs JJ, Tidor B, Alber T, Kim PS. 1998. High-resolution protein design with backbone freedom. *Science* 282:1462–1467.
- Harbury PB, Zhang T, Kim PS, Alber T. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262:1401–1407.
- Huyghues-Despointes BM, Scholtz JM, Baldwin RL. 1993. Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci* 2:80–85.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard. 1991. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119.
- Kammerer RA, Schulthess T, Landwehr R, Lustig A, Engel J, Aebi U, Steinmetz MO. 1998. An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc Natl Acad Sci USA* 95:13419–13424.
- Kohn WD, Kay CM, Hodges RS. 1995. Protein destabilization by electrostatic repulsions in the twostranded alpha-helical coiled-coil/leucine zipper. *Protein Sci* 4:237–250.
- Kortemme T, Ramirez-Alvarado M, Serrano L. 1998. Design of a 20-amino acid, three-stranded betasheet protein. *Science* 281:253–256.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:946–950.
- Krylov D, Barchi J, Vinson C. 1998. Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J Mol Biol* 279:959–972.
- Krylov D, Mikhailenko I, Vinson C. 1994. A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J* 13:2849–2861.
- Lavigne P, Soennichsen FD, Kay CM, Hodges RS. 1996. Interhelical salt bridges, coiled-coil stability, and specificity of dimerization. *Science* 271:1136–1137.
- Lombardi A, Summa CM, Geremia S, Randaccio L, Pavone V, DeGrado WF. 2000. Inaugural article: Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proc Natl Acad Sci USA* 97:6298–6305.
- Lu M, Shu W, Ji H, Spek E, Wang L, Kallenbach NR. 1999. Helix capping in the GCN4 leucine zipper. *J Mol Biol* 288:743–752.
- Lumb KJ, Kim PS. 1995. Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper [see Comments]. *Science* 268:436–439.
- Lumb KJ, Kim PS. 1996. Interhelical salt bridges, coiled-coil stability, and specificity of dimerization. *Science* 271:1137–1138.
- Lupas A. 1996. Coiled coils: New structures and new functions. *Trends Biochem Sci* 21:375–382.
- Merritt EA, Bacon DJ. 1997. *Raster3D: Photorealistic molecular graphics*. San Diego, California: Academic Press.
- Merutka G, Stellwagen E. 1991. Effect of amino acid ion pairs on peptide helicity. *Biochemistry* 30:1591–1594.
- Moitra J, Szilak L, Krylov D, Vinson C. 1997. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry* 36:12567–12573.
- Morgan CS, Mayo SL. 1998. Surface design of an alpha helical protein. *Protein Sci* 7:85.
- Munoz V, Serrano L. 1995. Helix design, prediction and stability. *Curr Opin Biotechnol* 6:382–386.
- Musafia B, Buchner V, Arad D. 1995. Complex salt bridges in proteins: Statistical analysis of structure and function. *J Mol Biol* 254:761–770.
- Otwinowski Z, Minor W. 1997. *Processing of X-ray diffraction data collected in oscillation mode*. San Diego, California: Academic Press.
- Perrakis A, Sixma TK, Wilson KS, Lamzin VS. 1997. wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Crystallogr D Biol Crystallogr* 55:448–455.
- Peters J, Baumeister W, Lupas A. 1996. Hyperthermostable surface layer protein tetrabrachion from the archaeobacterium *Staphylothermus marinus*: Evidence for the presence of a right-handed coiled coil derived from the primary structure. *J Mol Biol* 257:1031–1041.
- Severin K, Lee DH, Kennan AJ, Ghadiri MR. 1997. A synthetic peptide ligase. *Nature* 389:706–709.
- Sheldrick GM, Schneider TR. 1997. *SHELXL: High resolution refinement*. San Diego, California: Academic Press.
- Spek EJ, Bui AH, Lu M, Kallenbach NR. 1998. Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci* 7:2431–2437.
- Steinmetz MO, Stock A, Schulthess T, Landwehr R, Lustig A, Faix J, Gerisch G, Aebi U, Kammerer RA. 1998. A distinct 14 residue site triggers coiled-coil formation in cortexillin I. *EMBO J* 17:1883–1891.
- Stetefeld J, Jenny M, Schulthess T, Landwehr R, Engel J, Kammerer RA. 2000. Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer [in Process Citation]. *Nat Struct Biol* 7:772–776.
- Struthers MD, Cheng RP, Imperiali B. 1996. Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science* 271:342–345.
- Thompson KS, Vinson CR, Freire E. 1993. Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. *Biochemistry* 32:5491–5496.
- Trail PA, Bianchi AB. 1999. Monoclonal antibody drug conjugates in the treatment of cancer. *Curr Opin Immunol* 11:584–588.
- Tripet B, Wagschal K, Lavigne P, Mant CT, Hodges RS. 2000. Effects of side-chain characteristics on stability and oligomerization state of a de novo-designed model coiled-coil: 20 amino acid substitutions in position “d”. *J Mol Biol* 300:377–402.
- Tripet B, Yu L, Bautista DL, Wong WY, Irvin RT, Hodges RS. 1996. Engineering a de novo-designed coiled-coil heterodimerization domain off the rapid detection, purification and characterization of recombinantly expressed peptides and proteins [published Erratum appears in *Protein Eng*, 1997, 10(3):299]. *Protein Eng* 9:1029–1042.
- Wang C, Stewart RJ, Kopecek J. 1999. Hybrid hydrogels assembled from synthetic polymers and coiled-coil protein domains. *Nature* 397:417–420.

Chapter 6

Removing an Interhelical Salt Bridge Abolishes Coiled-Coil Formation in a *de Novo* Designed Peptide

Meier, M., Lustig, A., Aebi, U. and Burkhard, P. (2002) Removing an interhelical salt bridge abolishes coiled-coil formation in a *de novo* designed peptide. *J Struct Biol*, **137**, 65-72.

The design of the presented peptide was invented by P. Burkhard. Ultracentrifugation was performed by Ariel Lustig. CD spectroscopy, crystallography and the analysis of the structure is my responsibility.

Removing an Interhelical Salt Bridge Abolishes Coiled-Coil Formation in a *de Novo* Designed Peptide

Markus Meier,* Ariel Lustig,† Ueli Aebi,* and Peter Burkhard*¹

*M.E. Müller Institute for Structural Biology, and †Department of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Received November 15, 2001, and in revised form March 13, 2002

α -Helical coiled coils represent a common protein oligomerization motif that are mainly stabilized by hydrophobic interactions occurring along their coiled-coil interface, the so-called hydrophobic seam. We have recently *de novo* designed and optimized a series of two-heptad repeat long coiled-coil peptides which are further stabilized by a complex network of inter- and intrahelical salt bridges. Here we have extended the *de novo* design of such two heptad-repeat long peptides by removing the central and most important *g-e'* Arg to Glu (*g-e'*RE) ionic interhelical interaction and replacing these residues by alanine residues. The effect of the missing interhelical ionic interaction on coiled-coil formation and stability has been analyzed by CD spectroscopy, analytical ultracentrifugation, and X-ray crystallography. We show that the peptide, while being highly α -helical, is no longer able to form a parallel coiled-coil structure but rather assumes an octameric globular helical assembly devoid of any coiled-coil interactions. © 2002 Elsevier Science (USA)

Key Words: coiled coil; protein *de novo* design; ionic interactions; protein folding; protein oligomerization.

INTRODUCTION

The α -helical coiled-coil structural motif mediates subunit oligomerization of a large number of proteins (Burkhard *et al.*, 2001). Typically, coiled coils consist of two to five right-handed amphipathic α -helices that “coil” around one another to form a left-handed supercoil (Lupas, 1996). Sequences of parallel left-handed coiled-coil proteins are characterized by a heptad repeat pattern of seven amino acids denoted **a** to **g** harboring mostly apolar residues in the **a** and **d** positions. The stability of the coiled coil is achieved by the systematic packing of the side chains of the amino acids in

the **a** and **d** positions along a hydrophobic seam. This is called “knobs-into-holes” packing, and was first postulated by Crick (1953). Distinct sites within heptad-repeat-containing amino acid sequences, so-called coiled-coil trigger sites, may be necessary to induce coiled-coil formation (Kammerer *et al.*, 1998; Steinmetz *et al.*, 1998). We have shown that the coiled-coil trigger site of cortexillin contains a distinct pattern of inter- and intrahelical salt bridges in addition to the hydrophobic interactions occurring along the dimer interface (Burkhard *et al.*, 2000a) with interhelical ionic interactions occurring even at **a** and **d** positions of the coiled coil (Fig. 1a).

Based on these findings we have recently *de novo* designed a stable two-heptad-repeat long coiled-coil peptide that is stabilized by an extended network of inter- and intrahelical salt bridges (Burkhard *et al.*, 2000b; Fig. 1b). The design of this parent peptide was further refined by testing different other possibilities of intra- and interhelical salt bridge arrangements ultimately leading to an improved thermodynamic stability of an optimized peptide which has a melting temperature above 50°C (Burkhard *et al.*, 2002). In this latter study we have also demonstrated the importance of an optimal arrangement of salt bridges and especially the importance of the **e-g'**RE (naming convention according to Burkhard *et al.*, 2002) interhelical salt bridge.

As yet, results from experimental measurements on the contribution of interhelical ionic interactions to coiled-coil stability have not been conclusive (cf. Lavigne *et al.*, 1996; Lumb and Kim, 1996). However, as noted before, interhelical salt bridges including charged residues even at **a** positions in the coiled-coil trigger site of cortexillin I (Burkhard *et al.*, 2000a) point to their importance for coiled-coil formation and stability (Kohn *et al.*, 1998; Krylov *et al.*, 1998). Starting from the same parent peptide (cf. Burkhard *et al.*, 2000b) we have now further extended our study to verify the importance of the interhelical salt bridge for coiled-coil formation and stability in the context of

¹ To whom correspondence should be addressed. Fax: +41 61 267 21 09. E-mail: peter.burkhard@unibas.ch.



TABLE I
Data Collection, Phasing, and Refinement Statistics

Dataset	Native	Derivative (trimethyl lead acetate)	Native
X-ray source	Elliott GX-20 rotating anode generator	Elliott GX-20 rotating anode generator	Beam line ID 29 ESRF Grenoble
Crystallization conditions	2 M MgSO ₄ , 100 mM sodium acetate, pH 4.6	2 M MgSO ₄ , 100 mM sodium acetate, pH 4.6	100 mM bicine pH 9.0
Space group	P3 ₁ 21	P3 ₁ 21	P3 ₁ 21
Unit cell <i>a</i> , <i>b</i> , and <i>c</i> (Å)	<i>a</i> = <i>b</i> = 56.51 <i>c</i> = 113.61	<i>a</i> = <i>b</i> = 56.49 <i>c</i> = 113.30	<i>a</i> = <i>b</i> = 56.63 <i>c</i> = 113.87
α, β , and γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Wavelength (Å)	1.542	1.542	0.9763
Resolution (Å)	2.6	2.7	2.0
Completeness (%)			
Overall	99.0	99.7	96.4
Outer shell	90.0 (2.69–2.60 Å)	99.8(2.80–2.69 Å)	83.3 (2.06–2.00 Å)
<i>R</i> _{sym} (%) ^a			
overall	7.8	8.9	7.4
outer shell	48.6 (2.69–2.60 Å)	57.5 (2.80–2.69 Å)	46.3 (2.06–2.00 Å)
Mean redundancy	>9	4–5	6–7
Phasing power ^b		Acentric iso.: 1.330 Acentric ano.: 0.984 Centric: 1.094	
FOM ^c		Acentric: 0.32953 Centric: 0.30714	
<i>R</i> factor (%) ^d			24.2
free <i>R</i> factor (%)			27.8
Mean <i>B</i> -factor protein (Å ²)			53.53
Mean <i>B</i> -factor water (Å ²)			59.63
RMS deviation from ideal bond length (Å)			0.010
RMS deviation from ideal bond angle (°)			1.20
RMS deviation from ideal dihedral angle (°)			16.10
RMS deviation from ideal improper angle (°)			0.80
Number of independent molecules in the asymmetric unit			8
Number of atoms in asymmetric unit			1163

$$^a R_{\text{sym}} = \frac{\sum_{hkl} \sum_i I_i - \langle I \rangle / \sum_{hkl} \sum_i I_i}{\sum_{hkl} \sum_i I_i}$$

^b Phasing power (FH/lack of closure).

^c FOM, figure of merit: $((\cos \phi)^2 + (\sin \phi)^2)^{1/2}$.

$$^d R \text{ factor} = \frac{\sum_{hkl} |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_{hkl} |F_{\text{obs}}|}{\sum_{hkl} |F_{\text{obs}}|}$$

these newly designed coiled-coil peptides. In the present design we have removed the interhelical **g-e'** salt bridge by replacing the residues in the **g** and **e'** position (Arg and Glu) of the parent peptide by Ala residues that exhibit a high propensity for α -helix formation (Fig. 1c). We show that this *de novo* designed peptide, while still being highly α -helical, is no longer able to form a parallel coiled-coil but rather assumes a globular octameric structure devoid of any coiled-coil interactions.

MATERIALS AND METHODS

Synthetic peptides. The peptide Suc-DELAIRELAARIK-NH₂ was purchased from Neosystem (Strasbourg, France) and its purity (>95%) was verified by qualitative amino acid and mass spectral analyses.

Analytical ultracentrifugation. AUC was performed on an Optima XL-A analytical ultracentrifuge (Beckman Instruments, Palo Alto, CA) equipped with a 12-mm Epon double-sector cell in an An-60 Ti rotor. Sedimentation equilibrium runs were performed at 20°C at rotor speeds of 32,000 rpm and 40,000 rpm and a peptide concentration of 1.0 mg/ml. Average molecular masses were evaluated by using a floating baseline computer program that adjusts the baseline absorbance to obtain the best linear fit of $\ln(\text{absorbance})$ versus the square of the radial distance. A partial specific volume of 0.73 ml/g was used for low ionic strength measurement and of 0.78 ml/g for the high ionic strength measurements, respectively.

CD spectroscopy. CD analysis of the synthetic peptide was performed on a Cary 61 spectropolarimeter. Melting curves and ionic strength dependence were measured at 222 nm.

Crystallization. Crystals were grown in 24-well Limbro plates by vapor diffusion using the hanging-drop method (McPherson,

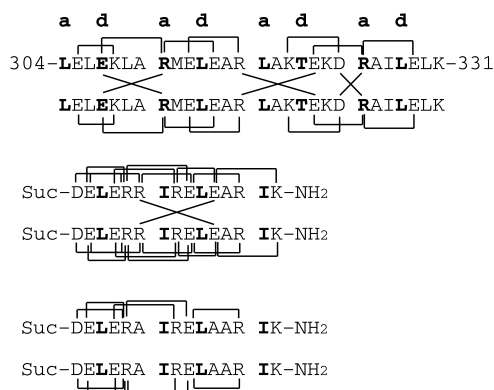


FIG. 1. Possible inter- and intrahelical salt bridges in the sequences of (a) cortexillin (Burkhard *et al.*, 2000a), (b) the parent peptide (Burkhard *et al.*, 2000b), and (c) the novel peptide lacking the interhelical salt bridge. Interhelical salt bridges are depicted as lines between the sequences; intrahelical salt bridges are depicted as brackets above and below the sequences.

1982). The 1-ml well solution contained 2 M MgSO₄ and one of the following buffers: 100 mM sodium acetate of pH 4.6, 100 mM sodium cacodylate of pH 6.5, 100 mM Hepes of pH 7.5, or 100 mM bicine of pH 9.0. The 4- μ l drop contained 2 μ l peptide at a concentration of 60 mg/ml and 2 μ l of well solution. Crystals grew within 14 days to a size of 0.3 \times 0.3 \times 0.3 mm³. The crystals belong to the trigonal space group P3₁21 exhibiting unit cell dimensions of $a = b = 56.5$ \AA , and $c = 113.6$ \AA (Table D). There are 8 independent molecules in the asymmetric unit. A twofold symmetry axis relates the chains A, B, C, D to the chains E, F, G, H. Thus the basic unit is a tetramer. The molecules occupy a volume of about 40% of the asymmetric unit.

Data collection and processing. The first data set of the native peptide and the data set from the trimethyl lead acetate (TLA) derivative were collected from crystals grown at pH 4.6 on a modified Elliott GX-20 rotating anode generator running at 40 kV and 50 mA and a temperature of 100 K using a MAR research image plate detector (Hamburg, Germany). A second native data

set from a crystal grown at pH 9.0 was collected at the beam line ID 29 of the ESRF in Grenoble at a wavelength of $\lambda = 0.976$ \AA . All diffraction data were processed, integrated, and scaled using Denzo and Scalepack (Otwinowski and Minor, 1997), respectively.

Structure determination and refinement. The structure was solved by single isomorphous replacement methods (SIR). Initial heavy atom positions were obtained from isomorphous and anomalous difference Patterson maps. Heavy atom parameters were refined using the program Sharp (De La Fortelle and Bricogne, 1997). Phasing was further improved by solvent-flattening, histogram matching, and twofold noncrystallographic symmetry (NCS) averaging using the program DM (CCP4, 1994). The position and orientation of each helix of the asymmetric unit could be determined using the experimental phases. Once the dimensions of the octamer and the exact solvent content were known, density modification including twofold NCS averaging over the NCS-related helices yielded a high-quality map using only the in-house data of up to 2.6 \AA into which the entire model could be built. At later stages of the refinement the resolution was extended to 2.0 \AA , using the native data set collected at the synchrotron. An atomic model was built using the graphics program O (Jones *et al.*, 1991) and refined using the CNS program package (Brunger *et al.*, 1998). No NCS constraints were used in the refinement. This model comprises eight monomers of the peptide, one magnesium ion, one chloride ion, and 105 water molecules. The final model has an R factor of 24.2% (30.2–2.0 \AA) and a free R factor of 27.8% which was calculated with 5.1% of the native data set aside prior to refinement. Rms. deviations from ideality in bond lengths and angles are 0.010 \AA and 1.2 $^\circ$, respectively (Table I).

Accession number. Coordinates have been deposited with the Research Collaboratory for Structural Bioinformatics under Accession Code 1L4X.

RESULTS AND DISCUSSION

Biophysical Characterization of the Peptide

To address the importance of the interhelical **g-e'RE** salt bridge of the parent peptide (Burkhard *et al.*, 2000b), a peptide was designed in which the residues involved in this ionic interaction (Arg 6 and

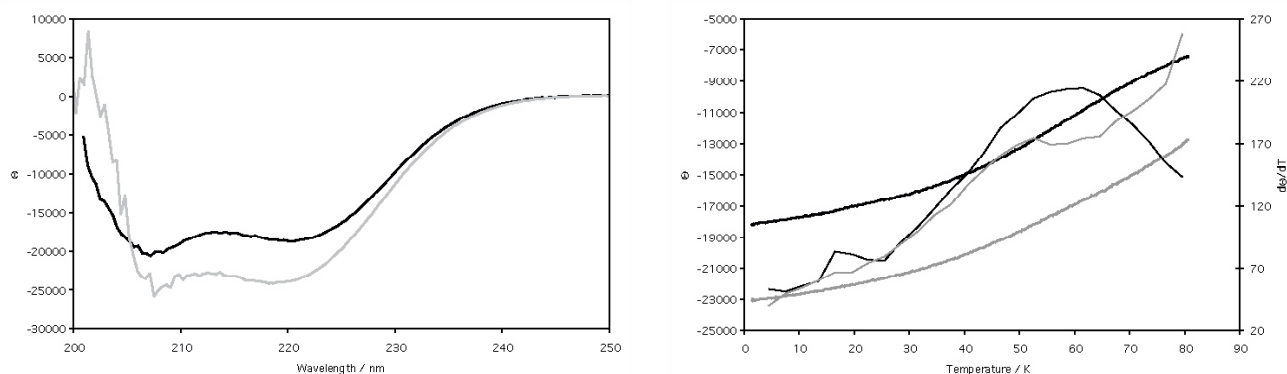


FIG. 2. CD analysis of the peptide. (Left) CD spectrum of the peptide. The peptide at a concentration of 1 mg/ml was analyzed in 10 mM Tris/HCl buffer (pH 7.5) with no salt (black) and in 2 M magnesium sulfate (gray) with 100 mM Tris/HCl. (Right) Melting curves (thick lines) and the corresponding first-order derivatives (thin lines) of the peptide under different ionic strength conditions monitored by the CD signal change at 222 nm, $[\Theta]_{222}$. The peptide at a concentration of 1 mg/ml was analyzed in 10 mM Tris/HCl buffer (pH 7.5) with no salt (black) and in 2 M magnesium sulfate (gray) with 100 mM Tris/HCl.

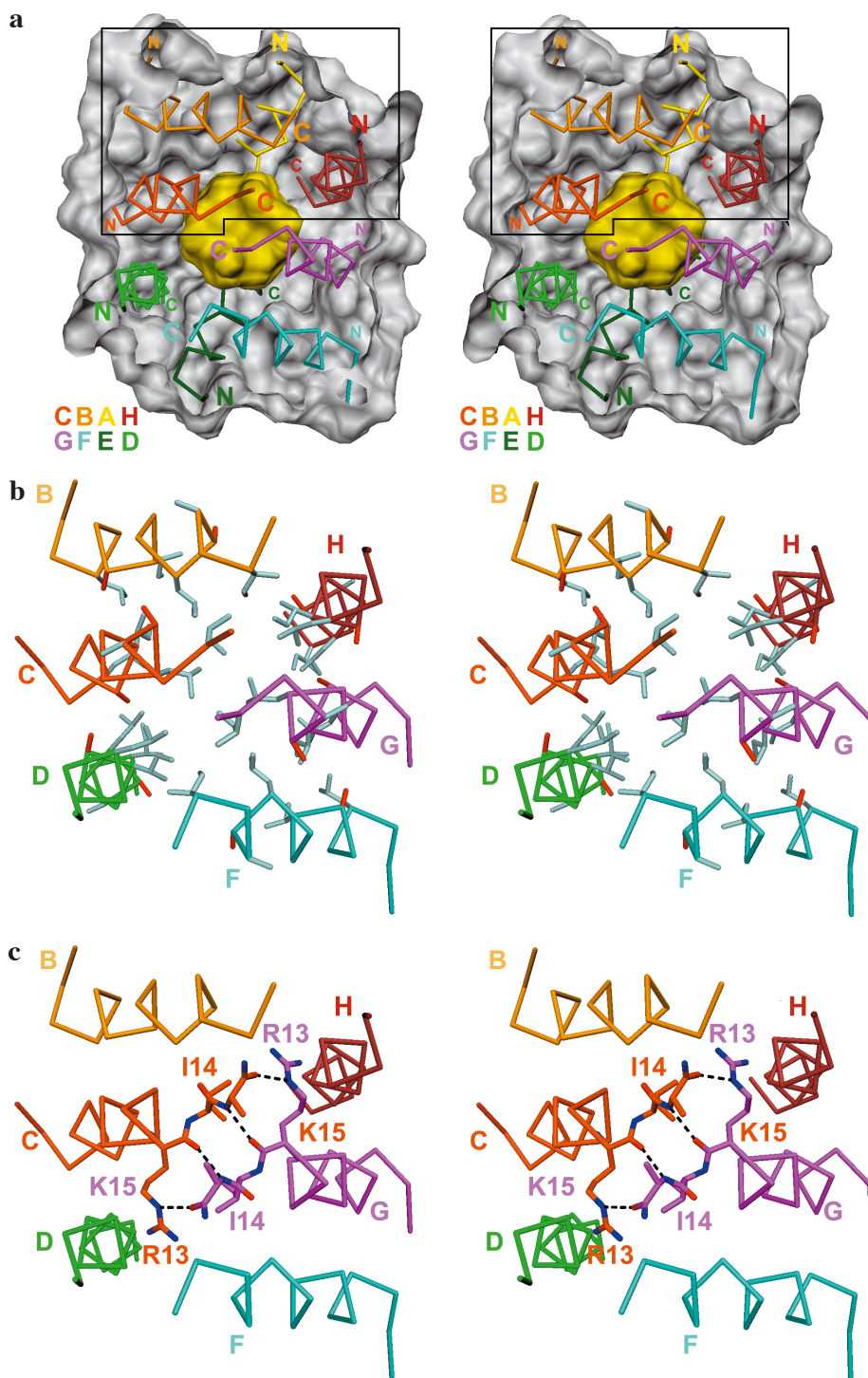


FIG. 3. Helical arrangement in the globular octameric structure. (a) The helices are shown in different colors and their N- and C-termini are labeled. The (outer) solvent-accessible surface is shown in gray while the surface of the central hydrophobic cavity is colored in yellow. (b) All hydrophobic residues of the helices B, C, D, F, G, and H are depicted as ball-and-stick models with their carbon atoms in gray. They are oriented toward the center of the globular structure, thus forming a hydrophobic cavity. The mutated residues Ala 6 and Ala 11 at the positions of the former interhelical salt bridge (Arg 6 and Glu 11 in the parent peptide) are depicted in red. Larger side chains at these positions would sterically interfere with the packing of the helices as observed in the globular structure. (c) A short β -sheet is formed by the residues Arg 13, Ile 14, and Lys 15 of the helices C and G, respectively. Furthermore, a hydrogen bond between the N_{ϵ} atom of Arg 13 and the C-terminal amide is formed at both ends of the β -sheet. The side chains of Lys 15 of helices C and G are disordered and therefore the C_{β} atom is shown only.

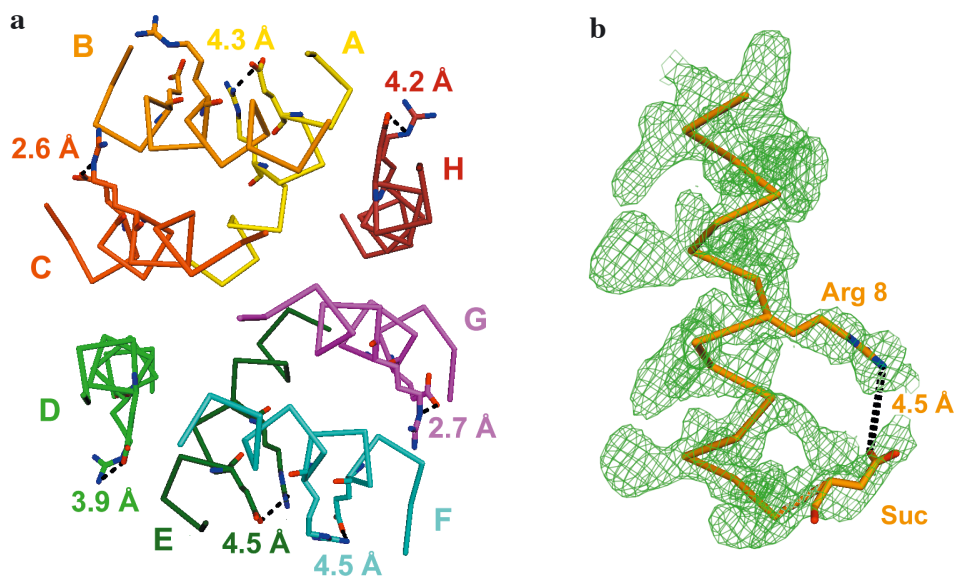


FIG. 4. Intrahelical salt bridges in the octameric structure. (a) The intrahelical salt bridge of the type i to $i + 4$ between Glu 4 and Arg 8 is conserved in all helices except for helix B. In helix C it is the only intrahelical salt bridge of all five possible salt bridges (Fig. 1c) which can be observed in the structure. (b) A 2Fo-mFc electron density map of the helix B. In this helix and in helix F (not shown) an intrahelical salt bridge of the type i to $i + 8$ is formed between the N-terminal succinyl moiety and Arg 8. This type of salt bridge cannot be formed in coiled coils.

Glu 11) were replaced by Ala. This modification also decreases the number of possible intrahelical salt bridges to five (Fig. 1). The oligomerization state of the *de novo* designed peptide Suc-DELELAIRE-LAARIK-NH₂ lacking the interhelical salt bridge was determined by analytical ultracentrifugation. By sedimentation equilibrium measurements at room temperature and at pH 7.5 the extrapolated average molecular weight in 150 mM NaCl is about 7.0 kDa while in 2 M NaCl it is about 10.5 kDa. This roughly corresponds to a trimeric or tetrameric state under low ionic strength conditions and to a pentameric or hexameric state under high ionic strength conditions. Circular dichroism (CD) at pH 7.5 and a peptide concentration of 1.0 mg/ml shows that the peptide is highly α -helical and that the degree of α -helicity is higher under high ionic strength conditions than under low ionic strength (Fig. 2). Based on the first-order derivative of the melting curve, the melting point of the α -helices in low ionic strength conditions is $62^\circ \pm 5^\circ\text{C}$. In 2 M MgSO₄ the melting point appears to be higher than the measured temperature range (0–81°C), since the maximum of the first-order derivative is higher than 81°C.

Structure of the Peptide

The peptide was crystallized in a solution with a high salt content (Table I). In contrast to the analytical ultracentrifugation experiments which gave

oligomerization states between trimeric and hexameric assemblies, structural analysis of the peptide by X-ray crystallography revealed that the peptide forms an octameric globular α -helical assembly (Fig. 3). In the crystal structure eight monomeric α -helices are arranged to form a globular structure which exhibits a twofold symmetry relationship between two tetramers. Hence, the basic unit of this globular structure is a tetramer (Fig. 3a). At this point however, it remains unclear whether this tetramer represents an independent folding unit. Helices B, C, and D and their NCS related helices F, G, and H are each arranged in a sheet-like manner (Fig. 3b). Helices B and C run approximately parallel to one another, whereas helix C is oriented antiparallel to helix D. These two α -helical sheets are inclined toward each other such that they form a V-shaped structure. The cleft of this V-shaped structure is covered by helix A and its symmetry-related helix E. The hydrophobic seam of all eight helices is oriented toward the center of the structure, thereby forming a hydrophobic cavity with a volume of 614 Å³ (Figs. 3a and 3b). This hydrophobic cavity contains a network of water molecules.

Since all hydrophobic residues are oriented toward the center of the globular structure (Fig. 3b), there do not exist any coiled-coil interactions within the structure. Helices A and E are neither parallel nor antiparallel to any other helix nor do the two

TABLE II
Interhelical Electrostatic Interactions in the Octamer^a

Tetramer 1 ^b			Tetramer 2 ^b		
From	To	Distance (Å)	From	To	Distance (Å)
Suc 0 (A) ^c	Lys 15 (B)	3.1	Suc 0 (E) ^c	Lys 15 (F)	4.1
Arg 8 (A)	Glu 4 (B)	3.7	Glu 4 (E)	Arg 8 (F)	4.0
Glu 9 (A)	Lys 15 (H)	3.9	Arg 8 (E)	Glu 4 (F)	3.7
Asp 1 (C)	Arg 13 (D)	3.9	Glu 9 (E)	Lys 15 (D)	3.3
Glu 2 (C)	Arg 13 (D)	2.8	Glu 2 (F)	Arg 8 (G)	4.3
Arg 5 (C)	Glu 9 (D)	4.3	Glu 2 (G)	Arg 13 (H)	3.6
Arg 13 (C)	Suc 0 (D) ^c	4.3	Arg 5 (G)	Glu 9 (H)	4.3
Glu 4 (D)	Lys 15 (F)	4.1	Arg 13 (G)	Suc 0 (H)	4.3

^a The cutoff for ionic interactions is 4.5 Å.

^b The interactions on the same row in the left (Tetramer 1) and right (Tetramer 2) columns are related by the twofold symmetry relationship within the octamer.

^c The name of the helix is given in parentheses; Suc, succinyl moiety.

α -helical sheets formed by helices B, C, D and F, G, H, respectively, coil around each other. Hence, there exist no knobs-into-hole interactions, since the hydrophobic seams of the helices do not form the interface between the helices. As predicted, the absence of the interhelical **g-e'RE** salt bridge does indeed abolish any coiled-coil formation.

The higher stability of the octamer in high ionic strength buffer (Fig. 2) suggests that it is largely stabilized by hydrophobic interactions. However, even at high ionic strength as found under the crystallization conditions, interhelical electrostatic interactions can be observed in the crystal structure. All charged residues (including the N-terminal succinyl cap) are at least once involved in one of these interhelical electrostatic interactions. A list which contains all 16 interhelical ionic interactions is provided in Table II. Most interhelical salt bridges occur twice; thus the symmetry-related tetramers dis-

play the same interhelical ionic interactions. There are four exceptions, where the salt bridge occurs only in one of the two tetramers (see Table II). All of the five possible intrahelical salt bridges (cf. Fig. 1c) are formed at least twice in the octamer (Table III). The *i* to *i* + 4 salt bridge between Glu 4 and Arg 8 is most strongly conserved. It is present in all helices except B (Fig. 4a). In helices B and F the residue Arg 8 is also involved in an *i* to *i*+8 salt bridge to the N-terminal succinyl moiety (Fig. 4b).

A short distorted β -sheet is formed by the residues 13 to 15 of the chains C and G and two hydrogen bonds between main chain atoms of these residues are formed (Fig. 3c). Two additional hydrogen bonds are formed between the N_ε atom of Arg 13 in helix C and the C-terminal amide group of helix G (and vice versa). The electron-density map calculated from the final model is poorly defined at the ends of the NCS-related helices C and G. The succinyl cap and

TABLE III
Intrahelical Electrostatic Interactions in the Octamer^a

From	To	Type	Chain ID (distance Å)
"Classical" intrahelical salt bridges			
Asp 1	Arg 5	<i>i</i> to <i>i</i> + 4	D (2.5), E (2.8), G (3.1), H (2.9)
Glu 2	Arg 5	<i>i</i> to <i>i</i> + 3	A (3.6), B (3.4), E (2.8), F (2.4)
Glu 4	Arg 8	<i>i</i> to <i>i</i> + 4	A (4.3), C (2.6), D (3.9), E (4.5), F (4.5), G (2.7), H (4.2)
Arg 5	Glu 9	<i>i</i> to <i>i</i> + 4	A (4.0), E (4.5)
Glu 9	Arg 13	<i>i</i> to <i>i</i> + 4	B (3.0), F (3.1)
Other salt bridges			
Suc ^b 0	Arg 5	<i>i</i> to <i>i</i> + 5	B (2.7), C (3.4)
Suc ^b 0	Arg 8	<i>i</i> to <i>i</i> + 8	B (4.5), F (3.3)

^a The cutoff for ionic interactions is 4.5 Å. The distance of the interaction is given in parentheses.

^b Suc, succinyl moiety.

TABLE IV
B Factors and Root Mean Square Differences
for Each Helix

Helix	Mean <i>B</i> factor (Å ²)	Main chain <i>B</i> factor (Å ²)	Side chain <i>B</i> factor (Å ²)	rmsd ^a (Å)	rmsd ^b (Å)
A	57.94	50.67	63.02		0.202 (E)
B	58.14	52.71	61.17	0.336	0.136 (F)
C	50.72	47.82	52.36	1.173	0.153 (G)
D	50.24	46.20	51.99	0.463	0.556 (H)
E	54.75	46.23	60.61	0.202	
F	56.86	50.57	60.25	0.330	
G	50.47	47.03	53.05	1.164	
H	49.14	46.61	50.63	0.719	

^a Root mean square difference to helix A as reference.

^b Root mean square difference to the NCS-related helix. The related helix is given in parentheses.

the side chains of Asp 1 have different conformations in both helices. These side chains might assume alternative conformations in addition to those shown in the model, but these conformations cannot reliably be refined at this moderate resolution. The side chains of Lys 15 in helices C and G are disordered and are not visible in the electron density. A list of *B*-factors and root mean square differences of the *C*_α positions for each helix can be found in Table IV. The model has been refined to an *R* factor of 24.2% and a free *R* factor of 27.8% using the program *CNS* (Brunger *et al.*, 1998; Table I).

Diffraction quality crystals have been grown in 2 M MgSO₄ at room temperature (20°C) in a wide range of different pHs from 4.6 up to 9.0, always yielding the same crystal form. From a comparison of the data sets collected at pH 4.6 and 9.0 (Table I) it can be seen that the pH of the crystallization conditions did not influence the arrangement of the helices in the octamer. A second crystal form was obtained in (NH₄)₂SO₄; however, these crystals formed spheres consisting of fine needles and were not suitable for X-ray diffraction experiments. The crystals grown in MgSO₄ melt immediately when the temperature is increased by only a few degrees. This high temperature sensitivity can be explained by the relatively loose crystal packing. There are only three crystal contacts formed at very localized areas on the surface, covering only about 22% of the surface of the octameric structure. One crystal contact is generated by applying the crystallographic twofold axis. The other two contacts are not related by crystallographic symmetry but are formed between NCS-related residues. The NCS-related contacts are mediated by a magnesium and a chloride

ion. Thus a central octamer and three symmetry-related octamers are all lying in one plane. This results in a loose packing with large solvent-filled cavities leading to a solvent content of about 60%.

We thank the staff of the beam line ID 29 at the ESRF in Grenoble for their kind support, especially Dr. Germaine Sainz. This work was supported by the Canton Basel-Stadt, by grants from the Swiss National Science Foundation and by the Maurice E. Müller Foundation of Switzerland.

REFERENCES

- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921.
- Burkhard, P., Ivaninskii, S., and Lustig, A. (2002) Improving coiled-coil stability by optimizing ionic interactions. *J. Mol. Biol.* **318**, 901–910.
- Burkhard, P., Kammerer, R. A., Steinmetz, M. O., Bourenkov, G. P., and Aebi, U. (2000a) The coiled-coil trigger site of the rod domain of corticillin I unveils a distinct network of interhelical and intrahelical salt bridges. *Structure Fold Des.* **8**, 223–230.
- Burkhard, P., Meier, M., and Lustig, A. (2000b) Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci.* **9**, 2294–2301.
- Burkhard, P., Strelkov, S. V., and Stetefeld, J. (2001) Coiled coils: A highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82–88.
- CCP4 (Collaborative Computing Project No. 4) (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr-D* **50**, 760–763.
- Crick, F. C. H. (1953) The packing of α -helices: Simple coiled coils. *Acta Crystallogr.* **6**, 689–697.
- De La Fortelle, E., and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods, in Carter, C. W. J. (Ed.), *Methods in Enzymology Part A*, p. 277, Academic Press, San Diego.
- Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
- Kammerer, R. A., Schulthess, T., Landwehr, R., Lustig, A., Engel, J., Aebi, U., and Steinmetz, M. O. (1998) An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc. Natl. Acad. Sci. USA* **95**, 13419–13424.
- Kohn, W. D., Kay, C. M., and Hodges, R. S. (1998) Orientation, positional, additivity, and oligomerization-state effects of interhelical ion pairs in alpha-helical coiled-coils. *J. Mol. Biol.* **283**, 993–1012.
- Krylov, D., Barchi, J., and Vinson, C. (1998) Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* **279**, 959–972.
- Lavigne, P., Soennichsen, F. D., Kay, C. M., and Hodges, R. S. (1996) Interhelical salt bridges, coiled-coil stability, and specificity of dimerization. *Science* **271**, 1136–1137.
- Lumb, K. J., and Kim, P. S. (1996) Interhelical salt bridges,

- coiled-coil stability, and specificity of dimerization. *Science* **271**, 1137–1138.
- Lupas, A. (1996) Coiled coils: New structures and new functions. *Trends Biochem. Sci.* **21**, 375–382.
- McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, J. Wiley, New York.
- Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *in* Carter, C. W. J. (Ed.), *Methods in Enzymology* p. 277, Part A, Academic Press, San Diego.
- Steinmetz, M. O., Stock, A., Schulthess, T., Landwehr, R., Lustig, A., Faix, J., Gerisch, G., Aebi, U., and Kammerer, R. A. (1998) A distinct 14 residue site triggers coiled-coil formation in cortexillin I. *EMBO J.* **17**, 1883–1891.

Chapter 7

Statistics of intrahelical salt bridges in coiled coils

All work presented in this chapter is my responsibility.

Statistics of intrahelical salt bridges in coiled coils

1. Introduction

A successful *de novo* design of protein structures which assume a predefined fold is crucially dependent how well the underlying principles and factors which lead to the specific fold have been understood. In Chapter 4 we explained why coiled coils are especially well suited for a rational design of protein structure and discussed the factors which have to be considered for a successful design of a stable coiled coil. In Chapter 5 and Chapter 6 we experimentally showed that not only the hydrophobic interactions between the helices forming the coiled coil are important, but also electrostatic interactions. Salt bridges in protein structures have been analysed in detail by previous studies and the importance of *i* to *i*+3 and *i* to *i*+4 ionic interactions for the stability of α -helices has been shown before (Finkelstein *et al.*, 1991; Klingler and Brutlag, 1994; Musafia *et al.*, 1995; Walther and Argos, 1996). However, to our knowledge, no statistical study to this extent as presented in this chapter, focussing mainly on coiled coils, has been performed previously.

2. Materials and methods

The program *SBSCC* was developed to produce a statistics of the salt bridges in all coiled-coil structures currently present in the *RCSB* database. It is written in C++ programming language and runs on several platforms. Currently, the *UNIX* platforms *Mac OS X (FreeBSD)*, *Tru64 UNIX (DEC alpha OSF1)* and *Microsoft Windows* operating systems are supported. *SBSCC* can be operated in two modes: Scanning and analysis. The scanning mode needs a local copy of the *RCSB* database (Berman *et al.*, 2000) which can be downloaded from the internet (<ftp://ftp.rcsb.org>) or ordered on CD-ROM (<http://www.rcsb.org/pdb/cdrom.html>). In the current state, the *RCSB* database consists of a collection of compressed files in the *PDB* format containing structural information. In the scanning mode, each file is searched for possible coiled coils. For this task, *SBSCC* relies on the program *SOCKET* (Walshaw and Woolfson, 2001) which analyses structural data for knob-into-holes interactions. As result, *SBSCC* produces separate lists with the *PDB ID* codes and content information of each database entry which contains parallel two-stranded, antiparallel two-stranded, parallel three-stranded, antiparallel three-stranded, parallel four-stranded, antiparallel four-stranded, parallel five-stranded and antiparallel five-stranded coiled coils.

In the analysis mode, *SBSCC* identifies all salt bridges in a collection of *PDB* input files and classifies them according to their location in the protein structure. A salt bridge is

considered to be formed if the distance between the two nearest nitrogen or oxygen atoms in the oppositely charged groups of two possible interaction partners is smaller than a certain cut-off specified on the command line. Hydrogen-bonded salt bridges are not separately treated. It is possible to specify a side chain temperature factor limit for these residues if they reside in an X-ray crystal structure. The program also counts the number of different amino acids occurring at each coiled-coil heptad position. For the heptad position assignment the *TWISTER* algorithm (Strelkov and Burkhard, 2002) is used which improves the algorithm used by the *SOCKET* program (see Chapter 4). In the analysis mode, a number of coiled-coil filters is available which allow to eliminate undesired coiled-coil geometries. Structures can contain parallel or antiparallel coiled coils which may be two- or more-stranded. Using these filters, one specific coiled-coil type can be selected, and the others are ignored. Filter options include coiled-coil length, orientation and oligomeric state. For X-ray crystal structures which form the majority of data in the protein database, it is possible to exclude low quality models. Limits can be set for the resolution, reliability factor (R) and free reliability factor (R_{free}) of the deposited data. The R and R_{free} limits can only be applied to structures which have been deposited in *PDB* format 2.x and for which these numbers have been supplied.

2.1. Limitations of the *SOCKET* program

In 2001, a database of coiled-coil structures was published by *The Protein Design Group* at the University of Sussex, UK (Walshaw and Woolfson, 2001). The database was based on the *PDB* release #89 which represented a state of the protein database prior to 1999. Since then, the protein database has grown considerably, reaching 22'448 entries in September 2003.

The *SOCKET* program which was developed to compile the Sussex coiled-coil database had severe limitations:

No crystallographic (and non-crystallographic if only a part of the asymmetric unit was deposited) symmetry operations were applied to crystal structures to build the complete biological assembly. Therefore, all coiled coils sitting on crystallographic symmetry axes were missed. In the case of parallel two-stranded coiled coils the percentage is 11%, after all.

SOCKET cannot handle *PDB* files which use insertion codes in the residue numbering scheme instead of plain numbers, e.g. 201A, 201B, etc., or entries containing residues with alternate conformations. The percentage of such files in the *PDB* is large. This rules

out a manual correction of the residue numbers, because it would consume too much time.

SOCKET depends on the program *DSSP* (Kabsch and Sander, 1983) for the secondary structure assignment. *DSSP* assigns α -helices according to IUPAC rule 6.2, *i.e.* a residue is included in a secondary structure only if its NH **and** CO groups form the appropriate hydrogen bonds. A newer algorithm (Hutchinson and Thornton, 1996) conforms to IUPAC convention rule 6.3 according to which a residue is considered to be part of an α -helix if either its NH **or** CO groups are involved in the appropriate hydrogen bonds. As a consequence, one extra residue is added to both ends of each helix. *SOCKET* has the option to extend the helices to mimic the Hutchinson & Thornton algorithm. However, adjacent helices which share a common residue are joined, even if there is a kink between them. This often yields false positives of knobs-into-holes interactions.

SOCKET yields wrong positives if the *PDB* entry contains overlapping helices. This can happen if two crystallographic alternate conformations are present in the asymmetric unit, or if similar models were deliberately superimposed for visualisation. Not infrequently, such overlaps were also due to errors in the *PDB* entry. Two superimposed helices are wrongly identified as coiled coil, because they satisfy the distance criteria used in the algorithm for the recognition of the knobs-into-hole interactions.

SOCKET cannot be run in batch mode, *i.e.* it must be run separately on each individual file.

The program does not use dynamic memory allocation which raises a big problem, because of the growing size and complexity of structures in the *PDB*.

SBSCC has been designed to avoid these problems. Protein database entries complying to format 2.x contain a *BIOMT* card which supply the symmetry operations needed to generate a complete biological entity from the deposited coordinates. *SBSCC* considers these entries and generates a symmetry expanded input file for *SOCKET*. It also removes alternate conformations (only the conformation with the highest occupancy is kept) and replaces insertion codes, thus converting the original *PDB* file into a *SOCKET* compatible format. The *PDB* format cannot accommodate more than 63 chains per model because the chain identifier is limited to a single alphanumeric letter (0-9, A-Z and a-z and space). In viral capsids this number can easily be exceeded if the symmetry mates of the chains in the asymmetric unit are generated. *SBSCC* limits the number of chains and deletes all chains further than 10 Å away from the coordinate centre of any

residue in the asymmetric unit. Such chains are superfluous, because the interactions between the symmetry mates are redundant.

The *PDB* format cannot accommodate more than 99999 atomic coordinates in a single model, because the atomic serial number is limited to 5 digits. Since *SOCKET* does not depend on atomic serial numbers, *SBSCC* simply restarts the numbering from 1 if this limit is exceeded.

SBSCC implements the modified *DSSP* algorithm by (Hutchinson and Thornton, 1996) and writes the assigned secondary structure information (α -helices only) into a *DSSP*-like file format which can subsequently be read by *SOCKET*. Kinks between adjacent helices are taken into consideration and such helices remain separate.

SBSCC works in batch mode and extracts *PDB* entries directly from the database. The entries which should be extracted can be supplied through a file containing a list of *PDB IDs*. If required, additional coordinate files in the *PDB* format (compressed or uncompressed) can be used as input. It is also possible to specify a list of undesired *PDB IDs* which should be excluded, even if they are present in the input list. This is useful if the input list was obtained from an automated procedure.

2.2. Limitations of *SBSCC*

The usage of the *BIOMT* card in the *PDB* file has some disadvantages. *PDB* structures entered before the introduction of the format 2.x (1996) do not contain those matrices. Hence, the biological unit cannot be generated. Coiled coils sitting on crystallographic symmetry axes will therefore still be missed in these entries.

The *BIOMT* matrices may not contain crystallographic symmetry information, but may contain information derived from other sources (e.g. electron microscopy) to position the atomic coordinates in a electron density map of the biological unit. Such cases are usually detectable, because they frequently generate overlapping chains.

In quite a few cases the matrices were incorrect, duplicating already existing chains or causing distorted and wrongly placed symmetry mates. If incorrect matrices cause very close contacts (C_{α} - C_{α} distance between non-adjacent residues $< 1.85 \text{ \AA}$) *SBSCC* ignores the structure and asks the user for a manual inspection. Otherwise, the error is not noticed. Only half the van der Waals is used for the check, because too many structures would be rejected with a larger radius.

Such problems could be avoided if the symmetry mates were generated directly from the crystallographic symmetry card. However, this would consume much more computational

time and the definition of a biological assembly requires human interpretation in many cases.

The *RCSB* database is currently undertaking efforts to unify their data in a new *XML*-format and correct inconsistencies in the deposited data (Westbrook *et al.*, 2002). This will hopefully help to avoid such problems in future.

If the size of the model exceeds the atomic serial number limit of 99999 in official *PDB* entries, the model is split into several models separated by *MODEL/ENDMDL* cards. However, those cards are also used for multiple *NMR* models or special demands by the authors. *SBSCC* (and *SOCKET*) currently ignores all models except the first one.

Currently, *SBSCC* is still dependent on the *SOCKET* program for the identification of the helices forming a coiled coil. However, the *SOCKET* algorithm for the detection of knobs-into-holes interactions is already integrated in *SBSCC* as *C++* code. It is used to find the alignment of the coiled-coil helices. Only little work is needed to implement the complete functionality of *SOCKET* into *SBSCC*. This will solve all problems with the missing dynamic memory allocation of *SOCKET*, and the dependence on the limited *PDB* format can be abandoned.

2.3. Differences between *SOCKET* and *SBSCC*

The coiled-coil length returned by *SOCKET* and *SBSCC* may be different. The *TWISTER* algorithm used for the heptad assignment in *SBSCC* requires that all strands forming the coiled coil have the same length. In contrast, *SOCKET* yields an individual length for each strand. *SBSCC* takes the furthest limits of the knobs-into-holes interactions of all coiled-coil helices as a first boundary for the coiled coil. The individual limits of each strand are then extended to match this boundary. If the extension is not possible because it would exceed the borders of the α -helix, the coiled-coil boundary is trimmed to the maximal possible size. *SBSCC* always returns a length which is larger or equal to the size given by *SOCKET*.

2.4. Statistical methods

For the analysis of data, numerous statistical methods are available. Different methods have to be used for different scales of the data, *i.e.* nominal scale (yes-no arbitration, categories), ordinal scale (data are ranked in an order) and metrical scale (interval scale and ratio scale). These scales are ranked, as given in the previous sentence. The nominal scale has the lowest rank and the ratio scale the highest. Methods developed for a

lower scale can also be used in a higher scale, but not vice versa. However, the price is a lower sensitivity of the method. Parametric statistical methods assume a mathematical model how the data are distributed, e.g. a normal distribution. The obtained results are then only valid if the underlying model is correct for these data which is often difficult to prove. In contrast, non-parametric tests do not have this limitation, but are less sensitive, because they work on a lower scale. An example is the Wilcoxon rank sum test described below which works on the ordinal scale, but can also be applied to metrical data.

Before a statistical test is applied, two hypotheses must be formulated. The null hypothesis (no effect is observed) and the alternative hypothesis (an effect is observed). The test then yields a p -value (*probability value*). This value corresponds to the probability to make an error if we discard the null hypothesis in favour of the alternative hypothesis. If the p -value is below a certain confidence level, the null hypothesis can be discarded. In the confirmatory data analysis this confidence level must be chosen before the experiment is performed and the data are ascertained. In the explorative data analysis this rule is less strict, and all reasonable confidence levels are allowed, as long as the confidence level is stated. Usually, the 95% level ($p < 0.05$) is used. This study belongs to the second category. The strength of a statistical test is to discard the null hypothesis, not to prove it. If the p -value is larger than the confidence level, this can have two meanings: Either the null hypothesis is indeed true, or the number of observations is not large enough to prove the opposite.

The statistical computations were done with the free open-source program *R* (Team, 2003). The following methods were used:

2.4.1. Fisher's exact test

The Fisher's exact test is a two-sample test to decide whether the observed contingents in two compared populations are equal (null hypothesis) or different (two-sided alternative hypothesis). It is also possible to choose the one-sided alternative hypotheses *greater* or *less*. The contingents are written in a 2 by 2 contingency table. The null hypothesis is equivalent to the hypothesis that the odds ratio in the contingency table equals one. The odds of an event is the number of events divided by the number of non-events, e.g. the number of formed salt bridges divided by the number of salt bridges which did not form. The odds ratio is obtained by dividing two odds. The Fisher's exact test offers the same functionality like Pearson's χ^2 -test (Pearson, 1966), but yields the exact p -value instead of an approximation (Fisher, 1935).

For example, consider the 3DR configuration in non-coiled coils and antiparallel 2-stranded coiled coils (Table 5). In non-coiled coils, SBSCC found in 38% of configurations that the ionic interaction formed. In antiparallel coiled coils this number was 61%. Is this number valid or simply a chance, because there were not enough such configurations observed? The p -value of Fisher's exact test is $7.814 \cdot 10^{-3}$. Thus, even with a confidence level of 99% ($p < 0.01$), the observed difference is still meaningful, *i.e.* statistically significant.

2.4.2. Wilcoxon rank sum test

The Wilcoxon rank sum test is a non-parametric two-sample test whether the medians of the distributions x and y are equal (null hypothesis) or different (two-sided alternative hypothesis). It is also possible to choose the one-sided alternative hypotheses *greater* or *less*. This test is equivalent to the Mann-Whitney test (Lorenz, 1992). Since the test is non-parametric, the data do not need to be normally distributed, but the test works on the ordinal scale and is therefore less sensitive than the Student's t -test for normal distributions (Hollander and Wolfe, 1973). In this work, a version of the test was used which computes exact conditional (on the data) p -values and quantiles using the Shift-Algorithm by Streitberg & Röhmel for both tied and untied samples (see the help text to the `wilcox.exact`-function from the `exactRankTests`-library in *R* for an explanation). Samples are tied if two (or more) values are identical and therefore get the same rank. This requires a correction, because they cannot be brought in an unambiguous order.

Example: In antiparallel coiled coils the total number of configurations containing aspartic acid (3DK, 3KD, 3DR, 3RD, 4DK, 4KD, 4DR, 4RD) was 262. The total number of configurations containing glutamic acid (3EK, 3KE, 3ER, 3RE, 4EK, 4KE, 4ER, 4RE) was 558. Is this difference meaningful? To verify this, the variance of these numbers must be included in the comparison. The numbers of the configurations containing aspartic acid were therefore conceived as distribution. The numbers of the configurations containing glutamic acid were conceived as another distribution. The Wilcoxon rank sum test was applied on the distributions, yielding a p -value of $1.554 \cdot 10^{-4}$. Thus, even assuming a confidence level of 99% ($p < 0.01$), the medians of the distributions differ and the observed difference is statistically significant.

2.4.3. Wilcoxon signed rank test

The Wilcoxon signed rank test is a non-parametric two-sample test for pair differences (Hollander and Wolfe, 1973). Whereas the Wilcoxon rank sum test compares two medians, this test compares the values in the distributions x and y pairwise, *i.e.* x_1 is com-

pared to y_1 , x_2 to y_2 *etc.* The null hypothesis states that the distribution x - y is symmetric about 0, *i.e.* there is no real difference between x and y . In this work, a version of the test was used which computes exact conditional (on the data) p -values and quantiles using the Shift-Algorithm by Streitberg & Röhmel for both tied and untied samples (see the help text to the *wilcox.exact*-function from the *exactRankTests*-library in *R* for an explanation).

Example:

We want to investigate whether salt-bridge configurations are more frequent if the first amino acid is negatively and the second amino acid positively charged than the opposite arrangement. It makes therefore sense to compare only pairing configurations, *i.e.* to compare the 3DK against the 3KD, the 3EK against the 3KE, the 3DR against the 3RD arrangement *etc.*, instead of comparing the median of two distributions. Applying the Wilcoxon signed rank test on the data, yields a p -value of $7.812 \cdot 10^{-3}$ for non-coiled coils (Table 4). Thus, even assuming a confidence level of 99% ($p < 0.01$), configurations with a negative charge on the first and a positive on the second amino acid are indeed preferred to the opposite arrangement.

2.4.4. Poisson distribution

The Poisson distribution is a model of probability for count data where the number of counted events is theoretically unlimited (Lorenz, 1992). The probability p_k that the result of a count is k is $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$, $k \in \mathbf{N}_0$, $\lambda \in \mathbf{R}^+$. An example is the number of salt-bridge configurations in the protein database, because the number of proteins in the database is theoretically unlimited.

The parameter λ is the mean *and* variance of the distribution. If the count data follow the Poisson distribution, its parameter λ is estimated by a single count n and the accuracy of the count is therefore known ($\lambda \approx n$). The lower (μ_l) and upper (μ_u) 95 % confidence limits are obtained by the following formulas: $\mu_l = \left(\frac{1.96}{2} - \sqrt{n} \right)^2$, $\mu_u = \left(\frac{1.96}{2} + \sqrt{n+1} \right)^2$

The formulas are approximations and are only valid if $n > 100$. The exact confidence limits for $n \leq 100$ are listed in (Lorenz, 1992). The 95 % confidence limits predict that if the count was repeated with other datasets of the same size the counted number n in each dataset would lie inside these limits with a probability of 95 %.

2.4.5. Poisson linear regression

Regression analysis is used to explain or model a relationship between a single variable Y and one or more variables X_1, \dots, X_p . Y is called *dependent variable* or *response*. It is a

vector containing the observations $Y = (y_1, y_2, \dots, y_n)^T$. X_1, \dots, X_p are called *independent variables, explanatory variables* or *predictors*. They can be represented as a matrix

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{pmatrix},$$

where $x_{i1}, x_{i2}, \dots, x_{in}$ denote the observed values for the variable X_i .

If $p = 1$ the method is called *simple regression* and if $p > 1$ *multiple regression*.

In linear regression, the function $Y = \beta X + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$ is fitted by finding estimates for the unknown parameters $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$. The term $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ represents random components called *errors*. Traditional least square regression is limited to a continuous response, uncorrelated errors and identically normally distributed errors.

Count data are non-negative integers and therefore not continuous. Such data often have a Poisson distribution. A generalised linear fit for Poisson distributed data fits the function $\log(Y) = \beta X + \varepsilon$ where the logarithmic link function ensures that the predicted values are always non-negative. The fit is performed using a maximum likelihood method. See (Faraway, 2002) for an introduction into linear regression.

The goodness of fit can be indicated using R^2 -values which range from 0 to 1, where 1 signifies best and 0 worst fit. We use an R^2 -value based on deviance residuals which was recommended by (Cameron and Windmeijer, 1996). It is defined as

$$R_{DEV,P}^2 = \frac{\sum_{i=1}^N \{y_i \log(\hat{\mu}_i / \bar{y}) - (\hat{\mu}_i - \bar{y})\}}{\sum_{i=1}^N y_i \log(y_i / \bar{y})}, \text{ where } \hat{\mu}_i \text{ denotes the predicted value for the observa-}$$

tion y_i and \bar{y} the predicted mean of the response.

2.5.6. Spearman's rank correlation coefficient

Spearman's rank correlation coefficient r_s is a robust, parameter-free method to measure an approximately linear correlation between two variables $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$. It ranges from -1 to +1, where -1 signifies maximum negative correlation, 0 no correlation and +1 maximum positive correlation. The observations are ranked and r_s is

$$\text{calculated on the ranks: } r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \text{ where } n \text{ is the number of observations and } d_i \text{ the}$$

difference of rank between the observations x_i and y_i . The calculation requires a correction if the data are tied (Lorenz, 1992).

A significance test called *Spearman's test* is available to investigate if r_s deviates significantly from 0. The implementation in the program *R* does not work in case of tied data.

2.6. Scaling of amino acid frequencies

The frequencies of the different amino acids in the proteome of a biological organism are different (Figure 1). The observed frequencies of salt-bridge configurations therefore also contain the frequency of the amino acids in proteins. If any effect is to be elucidated by comparing amino acid frequencies, they must be scaled by their average frequency of occurrence in proteins. The following procedure was used: The average amino acid frequencies from (Gerstein, 1998) were conceived as probabilities ($p_D = 0.051$ for aspartic acid, $p_E = 0.065$ for glutamic acid, $p_K = 0.075$ for lysine and $p_R = 0.042$ for arginine). The probability to find a combination of two amino acids in a protein sequence is then the product of their respective probabilities ($p_{DE} = p_{ED} = p_D \cdot p_E$ etc.). The scaled numbers $n_{s,t}$ were obtained by dividing the counted number n_t of configurations of type t by their respective probabilities p_t . Thereon, the new numbers were rescaled such that the total number of configurations remained equal before and after scaling.

$$n_{s,t} = \frac{n_t \sum_t n_t}{p_t \sum_t \frac{n_t}{p_t}}$$

$n_{s,t}$	Scaled number of configurations of type t
n_t	Counted number of configurations of type t
p_t	Probability of occurrence for a configuration of type t

2.7. Preparation of the reference data set

Often, the comparison of two small, rather low quality datasets does not yield useful results, because the results are not statistically significant due to the large uncertainties in the data. It is therefore advantageous if a high quality reference dataset is available that is based on a large amount of data. The comparison of the small datasets can then be done indirectly against the reference set.

For this reason, a statistical analysis of the intrahelical salt bridges in α -helices which do not reside in coiled coils was performed. A set of representative protein chains from the protein database was obtained from the *PISCES* culling server (Wang and Dunbrack, 2003). *PISCES* allows to select chains from protein structures which have a sequence identity lower than a freely selectable limit. This limit was chosen to be 25%. Because of the large number of structures which contain α -helices, a more restrictive resolution cut-

off of 2.5 Å could be used instead of 3.0 Å as in the coiled-coil statistics. A free *R*-factor limit of 31.0 % and a side-chain *B*-factor cut-off of 80 Å² for salt bridges was applied. *SBSCC* compared the list of *PDB* IDs obtained from *PISCES* against the *PDB* IDs which were identified to contain knobs-into-holes interactions (Table 1) and removed matching entries from the list. *SBSCC* was modified to read *PISCES* output lists and to extract single chains from *PDB* entries. The final data set comprises 152340 α -helical residues from 1769 protein chains.

2.8. Preparation of the coiled-coil datasets

SBSCC was run on the complete protein database in September 2003, yielding different lists of *PDB* entries containing 2-, 3-, 4- and 5-stranded parallel and antiparallel coiled coils (Table 1).

For the further statistical analysis of the occurrence of the salt bridges only X-ray diffraction data with a resolution better than 3.0 Å and a free *R*-factor lower than 31.0 % were used. If available, X-ray diffraction data were preferred to NMR data. Coiled coils shorter than 15 residues were also discarded. The data were manually inspected for *de novo* sequences, fusion sequences of different coiled coils (mostly *GCN4*), inverted sequences, right-handed coiled coils and chimeric coiled coils. Such structures were also removed. If there were duplicate models, only the structure with the highest resolution was taken. Only one structure of the same coiled coil in the asymmetric unit was kept. A side-chain *B*-factor cut-off of 80 Å² for ionic interactions was applied, *i.e.* a salt bridge with one or both side chain *B*-factors larger than 80 Å² was classified as *not formed*. Table 2 lists which structural models were used in the statistical analysis.

De novo sequences were removed, because they are based on an assumption which factors induce a successful coiled coil. If they were included, these factors would migrate into the dataset and might be overvalued. Furthermore, *de novo* designed sequences are frequently a variation of natural sequences or previously designed sequences, introducing sequence bias into the data set. This also applies to fusion sequences with *GCN4* or the like, and the designed variants of the alanine zipper in the major outer lipoprotein of *Escherichia coli* (Liu and Lu, 2002) which were also omitted. There were some coiled coils in the protein database with inverted amino acid sequence from C- to N-terminus. Other coiled coils were chimeric structures, *i.e.* the strands in the coiled coil were derived from different species. Such inverted or chimeric structures were omitted to be as authentic to nature as possible. Nevertheless, these structures are also subject to the principles of coiled-coil formation and the *PDB* cannot be taken as a representative sam-

ple of the proteins in biological organisms anyway (Gerstein, 1998). Maybe this proceeding was too conservative, however, only about 4 inverted or chimeric structures had to be omitted. The one right handed coiled coil (Stetefeld *et al.*, 2000) was not included because it contains an undecad repeat instead of a heptad repeat pattern.

3. Definitions

The following terms will be used for all statistical analyses: A ionic interaction is called *potential* or *possible* if the corresponding residues are in the proper configuration to form a salt bridge, *e.g.* a 3DK configuration with an aspartate and a lysine residue in this sequence, separated by two residues. Configurations which are related to each other by an exchange of the amino acids in their relative positions are termed XZ and ZX configurations, respectively (*e.g.* 3DR versus 3RD). An electrostatic interaction is termed *formed* if the distance d between the nearest nitrogen or oxygen atoms in the oppositely charged groups of the two interacting side chains is smaller than 3.8 Å and the mean B -factor of each side chain is smaller than 80 Å². According to (Marqusee and Baldwin, 1987) the term *salt bridge* should only been applied to hydrogen bonded ion pairs. However, in this study we did not distinguish between hydrogen-bonded and non-bonded ionic interactions. We will therefore use the term *salt bridge* and *ionic interaction* interchangeably.

The rotamer definition and nomenclature of side chains follows the recommendation of (Lovell *et al.*, 2000). Torsion angles of $180 \pm 60^\circ$ are denoted t , $+60 \pm 60^\circ$ p and $-60 \pm 60^\circ$ m .

The term *significant* is only used if it is statistically justified. A confidence level of 95 % is applied if not otherwise stated.

In tables and figures, the following abbreviations are used: a2-cc for antiparallel 2-stranded, p2-cc for parallel 2-stranded, a3-cc for antiparallel 3-stranded, p3-cc for parallel 3-stranded, a4-cc for antiparallel 4-stranded, p4-cc for parallel 4-stranded and p5-cc for parallel 5-stranded coiled coils. The abbreviation n-cc is used for non-coiled coils, *i.e.* α -helices which are not part of a coiled coil.

4. Results

4.1. Size of the datasets

In September 2003, SBSCC identified 1366 entries in the protein database which contain knobs-into-holes interactions. Table 1 contains a summary of the screening results. Of these entries, coiled-coil datasets were compiled as described in Section 2. *Materials and methods* (Table 2).

The statistics of the antiparallel 2-stranded coiled coils was created from 241 unique coiled coils of this type, comprising 10070 residues. The group of antiparallel 2-stranded coiled coils is most abundant in the protein database, yielding the most sound coiled-coil dataset. The extent of this dataset is 6.6 % of the size of the non-coiled-coils dataset.

The statistics of the parallel 2-stranded coiled coils was derived from 66 unique such structural motifs. They comprised 3478 amino acid residues. The number of parallel 2-stranded coiled coils in the protein database is much smaller than the number of antiparallel 2-stranded coiled coils. Its size is only 35 % of the size of the antiparallel 2-stranded data set, or 2.3 % of the size of the non-coiled-coil reference data set. Thus, it is much more difficult to draw significant conclusions.

4.2. Average coiled-coil length

The average length of the investigated antiparallel 2-stranded coiled coils is 20.9 residues per helix (3.0 heptad repeats). The explored parallel 2-stranded coiled coils have an average length of 26.4 residues per helix (3.8 heptad repeats). However, many α -helices in the protein database which were found to exert knobs-into-holes interactions, have a coiled-coil length shorter than 15 residues. Since they were excluded, they did not contribute to the average length.

4.3. Residues involved in ionic interactions

451 i to $i+4$ or i to $i+3$ salt-bridge configurations were counted in parallel 2-stranded coiled coils, and 820 such arrangements were found in antiparallel 2-stranded coiled coils. Roughly every 5th residue in parallel 2-stranded coiled coils and every 8th residue in antiparallel 2-stranded coiled coils is part of *at least* one such a configuration. However, only 12.4 % of the configurations form the salt bridge in parallel 2-stranded coiled coils. In antiparallel 2-stranded coiled coils the proportion is higher, that is 22.8 %. In the non-coiled-coils reference set roughly every 8th residue takes part in *at least* one configuration capable to form an i to $i+4$ or i to $i+3$ intrahelical electrostatic interaction. 22.4 % of these arrangements actually form the interaction. The deviation in the probability of salt-bridge formation in parallel 2-stranded coiled coils compared to antiparallel 2-stranded coiled coils and non-coiled coils is highly significant¹, but there is no apparent difference between antiparallel 2-stranded coiled coils and non-coiled coils, in the average.

¹ $p = 1.142 \cdot 10^{-7}$ against non-coiled coils and $p = 4.909 \cdot 10^{-6}$ against antiparallel 2-stranded coiled coils in Fisher's exact test

4.4. Exposure of the coiled coils to solvent

Salt bridges in a coiled coil compete for interaction partner residues within the coiled coil and its surrounding. By counting the ratio of foreign to domestic (within the coiled coil) ionic interactions relative to the coiled coils, an estimate can be obtained of the solvent exposure of the coiled coils in a dataset. In antiparallel 2-stranded coiled coils 441 foreign and 333 domestic interactions were counted, giving a ratio of 1.32. In parallel 2-stranded coil coils the number of foreign interactions was 106, compared to 125 domestic ones, yielding a ratio of 0.85. Thus, the parallel 2-stranded coiled coils are clearly more solvent exposed than the antiparallel 2-stranded coiled coils¹.

4.5. Intrahelical salt bridges

4.5.1 Formed intrahelical salt bridges

Figure 2a, Figure 3a and Figure 4a show the formed intrahelical salt bridges identified by SBSCC in non-coiled coils, antiparallel and parallel 2-stranded coiled coils, respectively. The grey columns represent the unscaled counts of formed salt bridges and the magenta columns show the data scaled by the average frequencies of the amino acids in proteins. Beside the i to $i+3$ and i to $i+4$ ionic interactions, also several other types of ionic interactions were found. Of these, the i to $i+1$ salt bridges 1RD seem to be quite frequent in non-coiled coils and antiparallel 2-stranded coiled coils. While the importance of i to $i+3$ and i to $i+4$ salt bridges in α -helices was recognized before (Musafia *et al.*, 1995), i to $i+1$ ionic interactions were hitherto not believed to contribute to helix stability (Walther and Argos, 1996). The i to $i+1$ ionic interactions will be discussed elsewhere, because at present SBSCC only counts the number of configurations which can form i to $i+3$ and i to $i+4$ electrostatic interactions.

In non-coiled coils and parallel 2-stranded coiled coils the most frequent ionic interaction is the 3ER salt bridge. In antiparallel 2-stranded coiled coils it is the 4KD electrostatic interaction in the raw data, or the 3DR salt bridge if the data are scaled by the frequency of occurrence of amino acids. However, to estimate the probability of a particular configuration to form salt bridges, it is better to look at the ratio of formed salt bridges to the number of such configurations (see below).

¹ $p = 3.312 \cdot 10^{-3}$ in Fisher's exact test

4.5.2 Configurations which can form intrahelical ionic interactions

The number of configurations which can form salt bridges should correlate with the α -helical propensities of the residues involved in the ionic interactions and their frequency of occurrence in proteins, because these residues have to play two roles: Their main chain has to assume an α -helical geometry, and the charged side chains can stabilize the helix additionally by forming intrahelical salt bridges. If some residues can form stable α -helices better than others, they should occur more frequently in this type of secondary structure. Therefore, also the probability is higher that it appears in a possible salt-bridge configuration. According to (Chakrabartty and Baldwin, 1995), arginine, lysine and glutamate residues have a high helix forming propensity, whereas for aspartic acid it is low (see Chapter 4). Hence, in non-coiled coils as well as in antiparallel and parallel 2-stranded coiled coils the number of possible salt bridges containing aspartic acid (3DK, 3KD, 3DR, 3RD, 4DK, 4KD, 4DR, 4RD) is roughly half as frequent as the number of salt bridges containing glutamic acid (3EK, 3KE, 3ER, 3RE, 4EK, 4KE, 4ER, 4RE) in the respective position (Figure 2b, Figure 3b and Figure 4b).

This difference is somewhat reduced when the number of potential salt bridges is scaled by the frequency of occurrence of the amino acids, but still persists for all salt-bridge types without exception (Table 3). The difference is largest in parallel 2-stranded coiled coils and smallest in non-coiled coils, but the Wilcoxon rank sum test confirms that it is statistically significant for all types (Table 3).

If the number of salt-bridge configurations would only reflect α -helical propensities, the observed frequencies for XZ arrangements for potential electrostatic interactions should be equal to ZX arrangements. However, in non-coiled coils, nature prefers if the first amino acid is negatively and the second amino acid positively charged (Table 4). Thus, the DK configuration is more frequent than the KD arrangement, EK than KE, DR than RD and ER than RE. This is true for both, i to $i+3$ and i to $i+4$ configurations, though it is less pronounced in the i to $i+4$ arrangements. No such preference could be established in parallel coiled coils, even when the data from 3- and 4-stranded coiled coils are included, and in antiparallel 2-stranded coiled coils neither (Table 4).

4.5.3. Percentage of formed intrahelical salt bridges

For a particular salt-bridge configuration, the ratio of formed to possible electrostatic interactions yields the probability of this arrangement to form the interaction, provided the number of counted configurations is large enough to be statistically sound.

Figure 2c, Figure 3c and Figure 4c show the number of potential salt bridges versus the number of formed ionic interactions for the salt bridge types 3DK, 3KD, 3EK, 3KE, 3DR, 3RD, 3ER, 3RE, 4DK, 4KD, 4EK, 4KE, 4DR, 4RD, 4ER and 4RE in non-coiled coils, antiparallel and parallel 2-stranded coiled coils. Figure 2d, Figure 3d and Figure 4d show the number of formed salt bridges divided by number of potential ionic interactions, yielding the percentage of formed salt bridges. The figures are colour coded in seven bins according to the distance of the ionic interactions: 0.0-2.0 Å, 2.0-2.3 Å, 2.3-2.6 Å, 2.6-2.9 Å, 2.9-3.2 Å, 3.2-3.5 Å and 3.5-3.8 Å. Salt bridges with larger distances are not shown.

Figure 2e, Figure 2f, Figure 3e, Figure 3f, Figure 4e, Figure 4f show the relative frequency of ionic interactions in each bin. There is a maximum in the 2.6-2.9 Å and 2.9-3.2 Å bins, and a steep decrease of the relative frequency in 3.2-3.5 Å bin for i to $i+3$ type interactions or in the 3.5-3.8 Å bin for i to $i+4$ interactions. For larger distances the relative frequency increases again slowly and approximately linearly. The course of the curve is asymmetric between 2.0-3.8 Å for most types of salt bridges because distances smaller than the sum of the van der Waals radii of the interaction partners cannot occur, whereas the maximum distance is limited by the spacing of the amino acids on the helix.

4.5.3.1. Percentage of formed salt bridges in non-coiled coils

On the average, 22% of the possible salt bridges are formed in non-coiled coils. Three types of salt bridges emerge strongly from this average: The 4KD arrangement with 48 % and the 3DR and 3ER configurations both with 38 % of formed salt bridges. The 4ER and 4RE arrangement are also above the average with 30 % formed salt bridges (Figure 2d). These 5 types of ionic interactions also happen to be the most frequent ones (Figure 2a).

4.5.3.2. Percentage of formed salt bridges in antiparallel 2-stranded coiled coils

The comparison of the formed salt bridges between antiparallel 2-stranded coiled coils and non-coiled coils is interesting (Figure 2d, Figure 3d and Table 5). The 4KD ionic interaction which is the most dominant one in non-coiled coils forms equally well in both datasets. However, it has been outstripped by the 3DR salt bridge which forms very efficiently in 61 % of these configurations in antiparallel 2-stranded coiled coils, compared to the still remarkable 38 % in non-coiled coils. This is also notable in Figure 3e and Figure 3f. The difference is highly significant. In non-coiled coils the 4DR and 4RD configurations constitute salt bridges with approximately the same below-average probability of 22 % and 20 %. However, in antiparallel 2-stranded coiled coils the probability of formation of the 4DR electrostatic interaction has greatly increased and is formed in 43 % of these configurations. In contrast, the 4RD salt bridge has lost any relevance and is only formed

in rare 3 % of these configurations. Both deviations from the non-coiled coils are statistically significant.

4.5.3.3. Percentage of formed salt bridges in parallel 2-stranded coiled coils

Comparing the formed ionic interactions to the non-coiled coils reference data set points out that the parallel 2-stranded coiled coils deviate much more from the reference set than the antiparallel ones. Also a direct comparison between the parallel and antiparallel 2-stranded coiled coils yields interesting differences (Figure 4d, Figure 4e, Figure 4f). *p*-values of the Fisher's exact test can be found in Table 6a and Table 6b.

Whereas the 3DR salt bridge is much more efficiently formed in antiparallel 2-stranded coiled coils compared to the non-coiled coils, the opposite is true for the parallel 2-stranded coiled coils. It is only formed in 14 % of the 3DR configurations.

The 3RE salt bridge is never constituted in parallel 2-stranded coiled coils, whereas in non-coiled coils and antiparallel 2-stranded coiled coils it is formed with the same probability, *i.e.* 20 % and 19 %, respectively.

The same is also true for the 4DK configuration. It does not form salt bridges in parallel 2-stranded coiled coils, but in non-coiled coils it forms with 17 % probability. The comparison to the antiparallel 2-stranded coiled coils is not significant, because of the small number of counts.

The 4KE configuration also shows a reduced potential to form electrostatic interactions in parallel 2-stranded coiled coils (5 %) compared to non-coiled coils (21 %) and antiparallel coiled coils (18 %).

The last significant difference is the 4ER configuration which is likewise less eager to form salt bridges in parallel 2-stranded coiled coils (13 %), compared to non-coiled coils (30 %) and antiparallel 2-stranded coiled coils (35 %).

Summarizing, in parallel 2-stranded coiled coils, the probability to form ionic interactions is reduced in 5 of the 16 investigated configurations, whereas in antiparallel 2-stranded coiled coils it is increased in two configurations and only decreased in one, compared to the non-coiled coil reference set. Thus, the configurations which have the largest probability to form salt bridges in parallel 2-stranded coiled coils are the 4KD with 39 % and 3ER with 32 % probability.

4.5.3.4. Probability of salt-bridge formation in XZ versus ZX configurations

There is a difference in the probability of formation for the configurations 3DR/3RD, 3ER/3RE, 4DK/4KD and 4EK/4KE in non-coiled coils. The differences for the other pairs

are not significant (Table 7). There are contrasts in these differences between the i to $i+3$ and the i to $i+4$ types. The i to $i+3$ configurations form with different probability if they contain arginine, the i to $i+4$ configurations if they contain lysine. And of the i to $i+3$ configurations with such differences, the salt bridges form with larger probability if the first amino acid is negatively charged, whereas of the i to $i+4$ configurations those ionic interactions form more frequently when the first amino acid is positively charged. An exception is the difference in the probability of salt bridge formation between the 4DR and 4RD configurations which only occurs in antiparallel 2-stranded coiled coils (see Section 4.5.3.2.).

4.5.4 Distribution of intrahelical salt bridges along the heptad repeat

Figure 3g and Figure 3h show the distribution of the intrahelical ionic interactions along the heptad repeats of antiparallel and Figure 4g and Figure 4h of parallel 2-stranded coiled coils. Configurations where one of the partner residues sits in a core position occur distinctly less frequently than configurations which are purely peripheral¹. However, salt bridges do not form with different probability if one partner residue is at a core position compared to purely peripheral salt bridges². Ionic interactions are never formed at pure core positions, *i.e.* from a to a d in i to $i+3$ configurations or from d to an a position in i to $i+4$ arrangements, and also configurations of oppositely charged residues occur rarely at such positions. There is an exception, though: SBSCC identified one 3EK electrostatic interaction of 2.9 Å distance from an a to a d position in the parallel 2-stranded coiled coil of a heme activator protein HAP1/DNA complex of *Saccharomyces cerevisiae* (PDB ID 1HWT, (King *et al.*, 1999)). The salt bridge is formed in both complexes in the asymmetric unit (Figure 6a). However, this ionic interaction actually induces termination of the coiled coil (Figure 6b), thus confirming that salt bridges cannot occur between two core positions.

In antiparallel 2-stranded coiled coils, the i to $i+3$ salt bridges from a g to a c position form with a higher probability of 32 % than the average of 20 %. The i to $i+4$ salt bridges from an a to an e position form with a probability of 51% compared to the average of 25 % (Table 8).

In parallel 2-stranded coiled coils only the i to $i+4$ interaction from an a to an e position is significantly more frequently formed, *i.e.* 47 %, than the average of 14 % (Table 9).

¹ $p = 1.059 \cdot 10^{-12}$ for antiparallel and $p = 9.724 \cdot 10^{-10}$ for parallel 2-stranded coiled coils in the one-sided Wilcoxon rank sum test

² $p = 0.4642$ for antiparallel and $p = 0.4595$ for parallel 2-stranded coiled coils in the two-sided Wilcoxon rank sum test

When comparing the occurrence of configurations at certain heptad positions it is noteworthy that the *cf* position is much more populated in antiparallel than in parallel 2-stranded coiled coils¹.

5. Discussion

5.1. Sequence bias in the coiled-coil datasets

There are fundamental differences between the dataset of non-coiled coils, that is of pure α -helices, and the one of coiled coils. The statistics of non-coiled coils was ascertained from linear α -helical sequences with maximal 25% identity between them, derived from a large number of structures of comparable high resolution (2.5 Å). This yielded a high quality, unbiased reference data set. By contrast, the coiled-coil statistics was drawn from two- (or more-) stranded sequences which can have a different degree of identity between the strands, in the frequent case of homodimers even 100 %. Redundant coiled coils have been removed, but no further attempts were made to reduce the relatedness of the remaining coiled-coil sequences. The same sequence can form different coiled coils with diverse other sequences, all of them appear in the statistics. Therefore, the coiled-coil datasets are sequence biased and appropriate care must be taken when comparing them, because observed differences could be due to this bias. Taking entire coiled coils as a basis for the statistics is justified, however, because the same sequence can form salt bridges to a different extent, depending on the context of the coiled coil.

5.2. Effect of solvent exposure

Ionic interactions on the surface of proteins are less strong than those in the interior, because the charges are shielded on the surface by the surrounding water molecules and the side chains can adopt more conformations than in the close packed interior. The parallel 2-stranded coiled coils are more exposed to solvent than the antiparallel ones (see Section 4.4.). This might explain the lower probability of salt bridge formation observed in parallel 2-stranded coiled coils compared to their antiparallel counterparts.

5.3. Helix dipole

In non-coiled coils, salt-bridge configurations where the first amino acid is negatively and the second amino acid positively charged are more frequent than the opposite configurations (see Section 4.5.2 and Table 4). This phenomenon can be understood as a com-

¹ $p = 4.557 \cdot 10^{-4}$ in Fisher's exact test

compensation for the helix dipole. The individual dipole moments of the peptide bonds are linearly aligned along the helix axis. This causes a total dipole moment antiparallel to the helix axis, *i.e.* the helix corresponds to a dipole with a positive charge on the *N*-terminal end and a negative charge on the *C*-terminal end of approximately ± 0.5 unit charge (Hol, 1985; Wada, 1976). The dipole destabilises the helix if it is not compensated.

In antiparallel 2-stranded coiled coils the dipoles of the two helices cancel each other due to their antiparallel orientation. Consequently, no further compensation is required, and no preference for salt-bridge configurations with the negatively charged amino acid in the first and the positively charged in the second position is observed.

Curiously, no such preference is observed for parallel 2-stranded coiled coils either. The reason remains unclear.

The electric field generated by the helix dipole is strong at the ends of the helix, but weak between (Hol, 1985). Therefore, an eventual influence on the formation of the salt bridges would only manifest itself at the ends of the helix. If this is the case, remains to be shown in future studies.

5.4. Importance of side chain conformations for the formation of salt bridges

Side chain conformations are restricted in α -helices. The *p* orientation of the χ_1 angle of the side chain is sterically hindered by the main chain and occurs very rarely (Klingler and Brutlag, 1994). Also residues which form salt bridges have therefore to assume χ_1 angles of *t* or *m* orientation. Figure 7 shows stereo diagrams of the χ_1 angles for residues spaced at *i* to *i*+3 and *i* to *i*+4.

To form an optimal ionic interaction between an *i* and *i*+3 position, the χ_1 angles of the side chains should both in the *m* orientation. Combinations of *t/t*, *m/t* and *t/m* angle orientations¹ make the side chains to point apart from each other and it is therefore less likely that salt bridges are formed with such angle orientations (Klingler and Brutlag, 1994). Because the *i*+3 position is shifted to the left compared to the *i* position when viewed from the *N*- to *C*-terminus, the side chain at the *i*+3 position should be longer than the side chain at the *i* position in order to make an optimal interaction (Burkhard *et al.*, 2000). Thus, a 3ER configuration should form electrostatic interactions more frequently than a 3RE configuration, and a 3DR arrangement more often than a 3RD arrangement. In both cases such a difference is indeed observed (see Section 4.5.3.4.) and the effect is

¹ The orientation before the slash corresponds to the side chain located at position *i*, the orientation after the slash to the side chain at position *i*+3 or *i*+4, respectively.

stronger between the 3DR/3RD configurations, where the difference of length between the side chains is larger than between the 3ER/3RE arrangements.

For an optimal ionic interaction between an i and $i+4$ position, the χ_1 angle at position i should be in t orientation and the angle at position $i+4$ in m orientation. An m/t combination of χ_1 angle orientations cannot form ionic interactions, because the side chains point in opposite directions. The χ_1 angle orientations discussed in (Klingler and Brutlag, 1994) apply to van-der-Waals contacts between the side chains, mostly. The restrictions towards these angles to form salt bridges are less strict. Therefore, the m/m and t/t combinations are also suitable to form ionic interactions. In contrast to the i to $i+3$ interaction, the position $i+4$ is only slightly shifted to right compared to the i position when viewed from the N - to C -terminus. Thus the two positions are basically located above each other. Side chains which have an m/t χ_1 angle combination should have approximately the same length for an optimal interaction. For the m/m combination, the side chain at the $i+4$ position should be longer than the side chain at the i position. For the t/t combination the situation is opposite and the longer side chain should reside at the i position. I therefore doubt the reasoning of (Burkhard *et al.*, 2000) that a 4RE arrangement should be preferred to an 4ER configuration. Indeed, configurations involving arginine do not prefer a particular arrangement in non-coiled coils (see Section 4.5.3.4.). The exception of the 4DR/4RD configurations appears in antiparallel coiled coils only and must therefore have different reasons for its occurrence. The preference for lysine at the i position is probably due to the higher flexibility of this side chain compared to arginine and the availability of optimal rotamer conformations, *i.e.* conformations which are low of energy and which simultaneously provide an optimal geometry for salt-bridge formation.

An analysis of the rotamers of the side chains involved in 3DR and 3RD salt bridges is presented below. The rotamers of the other ionic interactions will be presented elsewhere.

5.5. Probability of salt-bridge formation of the 3DR configuration

Table 10 shows the side chain rotamers of 3DR salt bridges in antiparallel 2-stranded coiled coils and 3RD ionic interactions in non-coiled coils. Aspartic acid has only two rotamers in α -helices, α_m-10° and α_t60° . The α_m-10° is much more abundant and therefore energetically more favourable than the α_t60° conformation. Arginine has 34 observed rotamers. In α -helices, the most frequent ones are the $m_{tm}-85^\circ$ (13%), $m_{tt}180^\circ$ (9%), $t_{tp}85^\circ$ (5%), $t_{tp}180^\circ$ (5%), $m_{tp}180^\circ$ (4%), $m_{tt}85^\circ$ (4%), $m_{tt}-85^\circ$ (4%). All other rotamers occur less frequently (< 4%). In the 3DR ionic interaction the side chains can both

reside in the most favourable conformation. This is also the most frequently observed case (Figure 8). This explains the high probability of the 3DR configuration to form salt bridges in α -helices and coiled coils. By contrast, to form a 3RD electrostatic interaction it is not possible for both side chains to assume an energetically favoured conformation simultaneously. Indeed, most of the observed arginine conformations in the 3RD salt bridge cannot be assigned to a defined rotamer at all. Thus, the energy gained from the constitution of the salt bridge is partly lost by the adverse geometry of the side chains, and the ionic interaction is weakened.

The probability to form electrostatic interactions is reduced in the 3DR configurations of the parallel 2-stranded coiled coils dataset. Only in three of 21 observed such configurations is the interaction actually formed. The average *B*-factor of the side chains in 3DR configurations which formed salt bridges is 60.7 Å² in parallel 2-stranded coiled coils, whereas it is 26.4 Å² in the antiparallel ones, *i.e.* the side chains in the parallel 2-stranded coiled coils are more flexible. Table 11 shows the environment of the side chains of the 3DR configurations in the respective crystal structures. Almost all of the side chains of the 3DR configuration in parallel 2-stranded coiled coils are exposed to solvent. Also those side chains in 3DR configurations of antiparallel 2-stranded coiled coils which do **not** constitute the salt bridge tend to be exposed. In contrast, those side chains which **do** form the ionic interaction tend to be buried. It seems that the salt bridge is preferentially formed if the aspartate has the possibility to form additional attractive polar interactions, because 91 % of the aspartates in formed 3DR salt bridges constitute one or more additional interactions. In the 3DR configurations which do not form the salt bridges, only 64 % of the aspartates make additional polar interactions. The conclusion is that the 3DR salt bridge unfolds its strength prevailingly in a buried or partly buried environment where there is the possibility to make further interactions with other protein residues.

5.6. Comparison with earlier results

(Klingler and Brutlag, 1994) performed a study which used pairwise residue correlations to find structural constraints in α -helices. They found significant correlations for 4KD, 4KE, 4EK and 3DR pairs. In our study we can confirm that the 4KD configuration forms salt bridges with highest probability, and the 3DR arrangement in non-coiled coils and antiparallel 2-stranded coiled coils as well. The 4KE and 4EK interactions do not attract attention in our study, however (see Section 4.5.3.).

The group of R. L. Baldwin used alanine-based peptides to address the implications of ionic interaction on helix formation (Huyghues-Despointes *et al.*, 1993; Marqusee and

Baldwin, 1987). The measured ellipticities of their peptides in Table 12a directly mirror the overall contribution of the residues on the stability of the peptides. They found the following pattern of stabilisation for the configuration couples EK/KE, DR/RD and ER/RE: $4AB > 4BA \approx 3AB > 3BA$, where A denotes the acidic and B the basic residue of the corresponding pair of configurations. This pattern does not correlate with the probabilities of salt-bridge formation found for non-coiled coils, indicating that the contribution of ionic interactions to stability is more complex in real proteins than in simple peptides. However, the empirically observed pattern $3AB > 3BA$ does correspond well to the behaviour observed in our statistics for the 3DR/3RD and 3ER/3RE configurations (see Section 4.5.3.4.).

(Smith and Scholtz, 1998) were able to measure the contributions of the ionic side chain interactions of lysine with glutamic or aspartic acid to peptide stability in simple alanine-based peptides (Table 12b). The reported ΔG_{sc} values reflect pure side chain interaction energies. Contributions due to the helix propensity values of the residues, helix capping effects, interaction of the charged side chains with the helix dipole, and other factors contributing to helix stability were subtracted. Their ΔG_{sc} values for glutamate-lysine interactions are similar regarding the sequence of the amino acids, in both i to $i+3$ and i to $i+4$ spacing. If one of the partner residues is aspartic acid, however, the interaction energy more than doubles when lysine resides at the i position and aspartic acid at the $i+3$ or $i+4$ position. The strongest interaction energy is reported for the 4KD configuration. This agrees well with the dominant role of the 4KD configuration we observe in our statistics. However, we detect no preference in terms of salt-bridge formation for either the 3DK or 3KD configuration, where it is expected from the ΔG_{sc} values, but we see a preferential salt-bridge formation of the 4KE configuration compared to the 4KE arrangement which is not expected (see Section 4.5.3.4.).

(Fernandez-Recio and Sancho, 1998) calculated statistical energy values from the frequency of amino acid pairs spaced at i to $i+3$ and i to $i+4$ in α -helices and compared them to measured side chain interactions energies. They found no statistically significant correlation and concluded that α -helices are not, in general, stabilised by side-chain interactions. They found, however, that the number of each residue pair involving lysine and aspartic or glutamic acid except 3KD is different from the number that would be expected due to a random distribution of the residues in the α -helix.

5.7. Dissecting the importance of intrahelical ionic interactions for helix stabilisation

Dividing the number of observed ionic interactions by the total number of configurations which can form such electrostatic interactions yields a probability value for each configuration that it forms the ionic interaction. If nature constitutes ion pairs to stabilise α -helices and coiled coils, we would expect to observe residue pairs which have a high probability to form the ionic interactions more frequently than pairs with a low probability. The probability to find an interacting amino acid pair in a protein structure can surprisingly often be well described by a Boltzmann distribution $p(ab) \sim \exp(-E_{ab}/c)$, where E_{ab} is the free energy of interaction of the pair and c a constant (Shortle, 2003). However, this model is rarely implemented in statistical packages. We therefore use a simple Poisson model as a first approximation (see Section 2. *Materials and methods*). But the logarithmic link function used in the Poisson regression thus obtains a physical justification.

We plotted the counted number of each amino acid pair against its probability to form the ionic interaction. It appeared that the points were randomly distributed with Spearman's rank correlation coefficient $r_s = 0.16$, obviously confirming the results of (Fernandez-Recio and Sancho, 1998). However, we concluded that a possible dependence might be confounded by the different average frequencies of the amino acids. We repeated the plots against the frequencies of configurations which were scaled against the average frequencies of amino acids in several genomes (Figure 9). Now, in non-coiled coils the situation changed and the number of configurations shows a clear dependence on the probability of formation (Figure 9a). Spearman's rank correlation coefficient shows a positive correlation with $r_s = 0.58$ when all data points are included. The value is significantly different from 0 as indicated by a p -value of 0.02. To illustrate the dependency, a Poisson linear fit with a log link function was calculated which is a standard regression method for count data. The fit is poor with an $R^2_{DEV,P}$ value of 0.28 due to the obvious outlier 4KD. If the outlier is excluded, the goodness of fit improves to $R^2_{DEV,P} = 0.74$ which is remarkably good.

Antiparallel 2-stranded coiled coils show only a low significant correlation¹ between the scaled frequency of a particular configuration and its probability to form salt bridges, in general (Figure 9b). The plot is horizontally divided into two sections with the configurations containing glutamic acid in the upper part and those containing aspartic acid in the lower part. Thus, the selection is clearly governed by the helix propensity of the residues which is high for glutamic but low for aspartic acid. The relative proportions of the

¹ $r_s = 0.48$ and $p = 0.06151$

charged amino acids in the average genome, non-coiled coils, antiparallel and parallel 2-stranded coiled coils are displayed in Figure 1b. They largely determine the position of the data points on the ordinate. If the data containing glutamic acid of Figure 9b are treated separately, they show now a highly significant correlation of $r_s = 0.73$ and $p = 0.04583$. The dependency is illustrated by a Poisson log-linear fit (red curve) which has an excellent goodness of fit of $R^2_{DEV,P} = 0.93$ if the point 3RE is excluded as an outlier. In contrast, the data points comprising aspartic acid display no significant correlation¹. The data points in Figure 9b can be assigned to three clusters: The first cluster comprises the 3RE, 4RE, 3ER and 4ER configurations which have a relatively high probability to form ionic interactions and an above-average occurrence. The second cluster comprises the 4DR, 4KD and 3DR arrangements which have a very high probability to form ionic interaction but a rare occurrence. The remaining cluster comprises all other configurations. They have a low probability to form ionic interactions and low occurrence.

The analysis of parallel 2-stranded coiled coils is particularly sensitive to sequence duplicates, because of many homodimers in the dataset. The probabilities of salt-bridge formation are accurate but the frequencies of the configurations are sequence biased (see Section 5.1.). Like in antiparallel 2-stranded coiled coils the residue pairs comprising glutamic acid occur in the upper part of the plot and those comprising aspartic acid in the lower part (Figure 9c). Separating the data gives an r_s of 0.57 for those data points containing glutamic acid and an r_s of 0.65 for those containing aspartic acid. The result is not statistically significant, however². 3ER is the only configuration which has both, a high probability to form the ionic interaction and a frequent occurrence in parallel 2-stranded coiled coils.

5.8. Consequences for coiled-coil design

The use of charged residues in coiled-coil design must satisfy two separate requirements: Residues with a high helix-forming propensity should be used, *i.e.* arginine, lysine and glutamic acid, but not aspartic acid. On the other hand, one of the configurations with the highest probability to form the ionic interaction is just one containing aspartic acid, namely the 4KD arrangement. But, indeed, this configuration appears relatively rarely. Figure 9 is most valuable to see how nature handles this problem.

¹ $r_s = 0.43$ and $p = 0.2992$

² $p = 0.1323$ for the data points containing glutamic acid. The p -value could not be calculated for the data points containing aspartic acid because of ties.

Salt-bridge networks are more stabilising than isolated ionic interactions (Musafia *et al.*, 1995; Olson *et al.*, 2001; Shi *et al.*, 2001). The residues should therefore be placed such that connected salt-bridge configurations arise which can form multiple interactions.

5.8.1 Parallel 2-stranded coiled coils

There are only two residue pairs which have an above-average probability to form the ionic interaction and a frequent occurrence in parallel 2-stranded coiled coils: 3ER and 4RE. All other intrahelical ionic interactions can probably be neglected in coiled-coil design (Figure 9c).

An i to $i+4$ salt bridge should, of course, not be positioned at a da or gd position. A ionic interaction at an ae position has an above-average probability to form (Figure 4g, Figure 4h and Figure 5a), but should only be used deliberately because it involves a core residue. An i to $i+3$ salt bridge should, obviously, not be positioned at an ad , dg or ea position. The 3ER configuration seems quite frequently to appear at a gc position (Figure 5a). We have seen that the data of the ionic interactions in parallel 2-stranded coiled coils prevalingly originate from solvent exposed coiled coils (see Section 4.4.). As such, they represent the situation usually encountered in coiled-coil design.

5.8.2 Antiparallel 2-stranded coiled coils

I am currently not aware of any attempted design of antiparallel 2-stranded coiled coils where ionic interactions have been used as a basis of the design. The configurations 4ER, 3ER, 4RE and 3RE have both, a high probability to form the ionic interaction and a numerous occurrence in antiparallel 2-stranded coiled coils (Figure 9b). They should therefore be appropriate for design. The 3DR ionic interaction has the highest probability to form in antiparallel 2-stranded coiled coils, but only in the interior of the protein (see Section 5.5.). It should therefore better be avoided in a design, even though it does not occur rarely.

Like in parallel 2-stranded coiled coils the configurations should not be put to positions involving core residues, except to the ae position in deliberate cases (Figure 3g and Figure 5b).

6. Conclusion

We have developed the program *SBSCC* to create a statistics of intrahelical ionic interactions in non-coiled coils, antiparallel and parallel 2-stranded coiled coils.

We have found that parallel 2-stranded coiled coils in the protein database are prevalingly solvent exposed, whereas antiparallel 2-stranded coiled coils are more frequently buried.

This has a big influence on the frequency and the probability of formation of ionic interactions. Salt-bridge configurations are more frequent in parallel 2-stranded coiled coils than in non-coiled coils, but the probability that the ionic interaction is formed is less. Antiparallel 2-stranded coiled coils do not differ from non-coiled coils in this respect.

The frequency of salt-bridge configurations containing aspartic acid is lower than those containing glutamic acid in α -helices and coiled coils, even when the numbers are corrected for the average frequencies of the amino acids in proteins. This is due to the low helix propensity of aspartic acid.

In non-coiled coils salt-bridge configurations where the first amino acid is negatively and the second amino acid positively charged are more frequent than the opposite arrangement. This can be understood as a compensation for the helix dipole. Antiparallel 2-stranded coiled coils require no such compensation and consequently no difference of frequency is observed.

We have discovered that not only i to $i+3$ and i to $i+4$ ionic interactions occur frequently, but also one i to $i+1$ interaction, namely 1RD. i to $i+1$ interactions were hitherto not believed to contribute to helix stability.

We have been able to detect a correlation between the probability of a salt-bridge configuration to form the ionic interaction and the frequency of occurrence of the configuration in non-coiled coils, antiparallel 2-stranded coiled coils and possibly also in parallel 2-stranded coiled coils. This is a clear indication that nature uses ionic intrahelical interactions to stabilise α -helices and coiled coils.

There are striking differences between the frequencies and probabilities of formation of the ionic interactions between non-coiled coils, parallel and antiparallel 2-stranded coiled coils, which is unexpected. The following salt-bridge configurations have been identified to have a large probability that the ionic interaction is formed: 4KD, 3DR, 3ER, 4ER and 4RE in non-coiled coils, 3DR, 4KD, 4DR, 4ER, 3ER and 4RE in antiparallel 2-stranded coiled coils and 4KD, 3ER and 4RE in parallel 2-stranded coiled coils. However, those configurations with the highest probability are less abundant than expected, especially if they contain aspartic acid. The probability to form the ionic interaction of the configurations 3DR and 4DR is greatly enhanced in antiparallel 2-stranded coiled coils compared to non-coiled coils. But this applies for the 3DR arrangement, at least, only if it is located in a buried environment with the possibility to form further polar interactions.

We have concluded that these salt-bridge configurations which have simultaneously a large probability to form the ionic interaction and a frequent occurrence are those which have the most stabilising effect. Thus they are recommended for design. These are 3ER,

3DR, 4ER and 4RE in non-coiled coils, 4ER, 3ER, 4RE and 3RE in antiparallel 2-stranded coiled coils and 3ER and 4RE in parallel 2-stranded coiled coils.

7. Future work

The following items remain to be done:

The analysis of side chain rotamers and side chain environments should be extended to all important intrahelical salt-bridge configurations to find an explanation for the different probabilities of formation.

The statistics should be extended not to investigate single salt bridges only, but complete networks, since networks are more stabilising than isolated ionic interactions (Musafia *et al.*, 1995; Olson *et al.*, 2001; Shi *et al.*, 2001).

The importance of the 1RD ionic interaction needs to be investigated.

8. Acknowledgements

I would like to thank Astrid Arion for the tedious sorting of the coiled coils extracted from the protein database and Armin Gemperli for his advice concerning statistical methods.

9. References

- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**, 235-242.
- Burkhard, P., Meier, M. and Lustig, A. (2000) Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci*, **9**, 2294-2301.
- Cameron, A.C. and Windmeijer, F.A.G. (1996) R-Squared Measures for Count Data Regression Models With Applications to Health Care Utilization. *Journal of Business and Economics Statistics*, **14**, 209-220.
- Chakrabarty, A. and Baldwin, R.L. (1995) Stability of alpha-helices. *Adv Protein Chem*, **46**, 141-176.
- Faraway, J.J. (2002) *Practical Regression And Anova using R*. Faraway, J. J., Michigan.
- Fernandez-Recio, J. and Sancho, J. (1998) Intrahelical side chain interactions in alpha-helices: poor correlation between energetics and frequency. *FEBS Lett*, **429**, 99-103.
- Finkelstein, A.V., Badretdinov, A.Y. and Ptitsyn, O.B. (1991) Physical reasons for secondary structure stability: alpha-helices in short peptides. *Proteins*, **10**, 287-299.
- Fisher, R.A. (1935) The logic of inductive inference. *Journal of the Royal Statistical Society Series A*, **98**, 39-54.
- Fujii, T., Sakai, H., Kawata, Y. and Hata, Y. (2003) Crystal structure of thermostable aspartase from *Bacillus* sp. YM55-1: structure-based exploration of functional sites in the aspartase family. *J Mol Biol*, **328**, 635-654.
- Gerstein, M. (1998) How representative are the known structures of the proteins in a complete genome? A comprehensive structural census. *Fold Des*, **3**, 497-512.
- Hol, W.G. (1985) Effects of the alpha-helix dipole upon the functioning and structure of proteins and peptides. *Adv Biophys*, **19**, 133-165.
- Hollander, M. and Wolfe, D.A. (1973) *Nonparametric statistical methods*. Wiley, New York etc.
- Hubbard, S.J. and Thornton, J.M. (1993) NACCESS. Department of Biochemistry and Molecular Biology, University College London, London.
- Hutchinson, E.G. and Thornton, J.M. (1996) PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci*, **5**, 212-220.
- Huyghues-Despointes, B.M., Scholtz, J.M. and Baldwin, R.L. (1993) Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci*, **2**, 80-85.
- Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

- King, D.A., Zhang, L., Guarente, L. and Marmorstein, R. (1999) Structure of a HAP1-DNA complex reveals dramatically asymmetric DNA binding by a homodimeric protein. *Nat Struct Biol*, **6**, 64-71.
- Klingler, T.M. and Brutlag, D.L. (1994) Discovering structural correlations in alpha-helices. *Protein Sci*, **3**, 1847-1857.
- Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, **55**, 379-400.
- Liu, J. and Lu, M. (2002) An alanine-zipper structure determined by long range intermolecular interactions. *J Biol Chem*, **277**, 48708-48713.
- Lorenz, R.J. (1992) *Grundbegriffe der Biometrie*. G. Fischer, Stuttgart etc.
- Lovell, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C. (2000) The penultimate rotamer library. *Proteins*, **40**, 389-408.
- Marqusee, S. and Baldwin, R.L. (1987) Helix stabilization by Glu...Lys+ salt bridges in short peptides of de novo design. *Proc Natl Acad Sci U S A*, **84**, 8898-8902.
- Musafia, B., Buchner, V. and Arad, D. (1995) Complex salt bridges in proteins: statistical analysis of structure and function. *J Mol Biol*, **254**, 761-770.
- Olson, C.A., Spek, E.J., Shi, Z., Vologodskii, A. and Kallenbach, N.R. (2001) Cooperative helix stabilization by complex Arg-Glu salt bridges. *Proteins*, **44**, 123-132.
- Pearson, E.S. (1966) *<<The>> selected papers of E. S. Pearson issued by the Biometrika Trustees to celebrate his 30 years as editor*. University of California Press, Berkeley.
- Philippsen, A. (2002) *DINO: Visualizing Structural Biology*. Basel.
- Shi, Z., Olson, C.A., Bell, A.J., Jr. and Kallenbach, N.R. (2001) Stabilization of alpha-helix structure by polar side-chain interactions: complex salt bridges, cation-pi interactions, and C-H...O H-bonds. *Biopolymers*, **60**, 366-380.
- Shortle, D. (2003) Propensities, probabilities, and the Boltzmann hypothesis. *Protein Sci*, **12**, 1298-1302.
- Smith, J.S. and Scholtz, J.M. (1998) Energetics of polar side-chain interactions in helical peptides: salt effects on ion pairs and hydrogen bonds. *Biochemistry*, **37**, 33-40.
- Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J. and Kammerer, R.A. (2000) Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. *Nat Struct Biol*, **7**, 772-776.
- Strelkov, S.V. and Burkhard, P. (2002) Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J Struct Biol*, **137**, 54-64.
- Team, R.D.C. (2003) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Wada, A. (1976) The alpha-helix as an electric macro-dipole. *Adv Biophys*, 1-63.
- Walshaw, J. and Woolfson, D.N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol*, **307**, 1427-1450.
- Walther, D. and Argos, P. (1996) Intrahelical side chain-side chain contacts: the consequences of restricted rotameric states and implications for helix engineering and design. *Protein Eng*, **9**, 471-478.
- Wang, G. and Dunbrack, R.L., Jr. (2003) PISCES: a protein sequence culling server. *Bioinformatics*, **19**, 1589-1591.
- Westbrook, J., Feng, Z., Jain, S., Bhat, T.N., Thanki, N., Ravichandran, V., Gilliland, G.L., Bluhm, W., Weissig, H., Greer, D.S., Bourne, P.E. and Berman, H.M. (2002) The Protein Data Bank: unifying the archive. *Nucleic Acids Res*, **30**, 245-248.

Tables

See 3. *Definitions* for an explanation of the abbreviations and conventions.

Protein database entries containing coiled coils

1	2	3	4	5	6
Coiled-coil type	Number of RCSB entries (unfiltered)	Number of RCSB entries (filtered)	Number of coiled coils/coiled-coil residues (filtered)	Uniquified number of RCSB entries (filtered)	Uniquified number of coiled coils/coiled-coil-residues (filtered)
p2-cc	314	111	166/7980	61	66/3478
a2-cc	781	434	1132/48234	194	241/10070
p3-cc	141	92	115/9855	20	22/2181
a2-cc	83	52	104/8364	24	25/1689
p4-cc	10	7	11/1828	5	6/1008
a4-cc	23	14	19/1560	9	12/952
p5-cc	5	3	3/555	1	1/185
a5-cc	0	0	0	0	0

Table 1: Protein database entries containing coiled coils as of September 2003. Column 2 gives the number of all entries found to have knobs-into-holes interactions as defined by the *SOCKET* algorithm. Columns 3-6 are restricted to coiled coils with a minimum length of 15 residues, structures with a resolution > 3.0 Å and a free R-factor > 31.0 %, and *de novo sequences*, inverted sequences, right-handed coiled coils and chimeric coiled coils have been removed. In columns 5-6 sequence duplicates have been eliminated, including symmetry related chains.

Protein database entries used in the coiled-coil statistics

p2-cc	a2-cc	p3-cc	a3-cc	p4-cc	a4-cc	p5-cc
1JNM,1HLO, 1NKP,1GO4, 1YSA,1JOC, 1ZME,2CBL, 1GP2,1TBG, 1I7W,1M5N, 1IC2,1J1D, 2NCD,1D7M, 1GK4,1DEB, 1HCI,1OA0, 1GNT,1KNZ, 1LJ2,1NO4, 1LS3, 1DIP , 3ERT,1QO0, 2LIG,1MC0, 1KS8,1L1Y, 4TF4,1HX8, 1LS4 ,1M1J, 1GVN,1P5H, 1PT7,1AYX, 1HWW,1O8U, 1E2A,1FN9, 1FEW,1G6N, 1C16,1H89, 1NWQ,1HWT, 1QKK,1DH3, 1FT9,1AM9, 1GD2,1IK9, 1LR1 ,1OMI, 1LON,1DVG	1A22,1EXZ,1K4T,1BJT,1AB4, 1FP3,1N1B,1Q5N,1QJA,1HKU, 2ARC,1HZ4,1IXC,1FGK,1CXZ, 1IK7,1LRZ,1AIJ,1EYS,1JB0, 6PRC,1F99,1JBO,1HA7,1PHN, 1QNF,1AIK,2EBO,1EK9,1FIO, 2VSG,1VSG,1CIY,1QOY,1O9F, 1GYG,1E12,1J3U,1C3C,1K62, 1HY0,1I0A,1HXC,1JFA,1JQN, 1JQO,1FUP,1N7H,1F6K,1H12, 1EB6,1A17,1AYX,1LF6,1GAI, 1EUL,1FS0,1AQT,1GMJ,1H8E, 2PVI,1KS8,1G9Z,1M5X,1ENK, 1IS9,1L1Y,1TF4,1CPY,1IS8, 1G8M,1KP9,1E2A,1DCE,1LD8, 1JCR,1VR2,1B3Q,1K04,1KVK, 1M52,1K1F,1QPE,1CSN,1FVR, 1M48,1IAR,1I1R,1ALU,1F45, 1AX8,1B5L, 1ITF ,1CNT,1BGC, 1RHG,1EVS,1LKI,1BCF,1GU2, 1QLE,2OCC,1L0L,1EZV,1GWI, 2FRV,1LKO,1J14,1QGH,1JGC, 1EUM,1BFR,1NFV,2FHA,1H96, 1AEW,1RCD,1BG7,1MFR,1O9R, 1N1Q,1DPS,1A2L,1J55,1MHY, 1MTY,1R2F,1H0O,1JKV,1AFR, 1BSM,1KKC,1LUV,1KI1,1DG3, 1FCH,1HXI,1IHG,1G4Y,1I4T, 1E3M,1DTO,1A92, 2HMX ,5EAU, 1ECM,1IJ5,1EGW,1F4K,1L8D, 1FEW, 1I6Z ,1FXK,1KID,1ELR, 1QSD,1FNN,1ROP,1JBG,1JI5, 1JIG,1DOV,1H6G,1I7W,1B04, 1DGS,1JW9,1SES,1LI5,1IQ0, 1L8W,1G5Z,1HCI,2SPC,1HH8, 1IO1,1M5I,1P32,1WAS,1G1X, 1JJ2,1H5W, 1HYW ,1QQE,1HVV, 1LVF,1I1I,1XWL,1TGO,1D5A, 2KTQ,1I50,1T7P,1JMS,1A5T, 1IAS,1JAL,1NG6,1NOG,1O3U	1QR9, 1QBZ, 1KWT, 1HTN, 1PWB, 1CA9, 1AA0, 2EBO, 1G2C, 1SVF, 1QU1, 1JEK, 1AQ5, 1JY2, 1FZC, 1M1J, 1LWU, 1EQ7, 1OAH	2ASR, 1VLS, 1AYX, 1CHU, 1QR9, 1QBZ, 1AA0, 1JR3, 1K62, 1HY0, 1EZ3, 1FPO, 1HX1, 1FS7, 1QDB, 1GU6, 1HZ4, 1BG1, 1BF5, 1GQE, 1KEY, 1M1J, 1OR3, 1FIO	1G1J, 1GL2, 1JTH, 1N7S, 1EZJ	256B, 2CCY, 1F1M, 1GAX, 1TLF, 1RPR, 1NIG, 1QU7, 2TOH	1MZ9
Human lamin frag- ment, residues 313- 386, S.V. Strelkov						

Table 2: List of the coiled-coil structural models used for the salt-bridge statistics. The second row shows *ID* codes of the protein database. In the third row additional models are given which were not yet deposited in the database. Models obtained by *NMR* are marked bold. All other structures were solved by X-ray diffraction.

Configurations containing aspartic acids versus those containing glutamic acid

	n-cc		a2-cc		p2-cc	
	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled
Sum of configurations containing D	3937	4567.8	262	301.0	113	139.0
Sum of configurations containing E	7216	6585.2	558	519.0	338	312.0
Ratio	0.546	0.694	0.470	0.580	0.334	0.446
<i>p</i> -value	$3.108 \cdot 10^{-4}$	0.03792	$1.554 \cdot 10^{-4}$	$1.865 \cdot 10^{-3}$	$1.554 \cdot 10^{-4}$	$1.088 \cdot 10^{-3}$

Table 3: Comparison of the number of configurations which contain aspartic acid versus those that contain glutamic acid in non-coiled coils, antiparallel and parallel 2-stranded coiled coils. The first row shows the summed up number of configurations containing aspartic acid (3DK, 3KD, 3DR, 3RD, 4DK, 4KD, 4DR, 4RD). The second row shows the summed up number of configurations containing glutamic acid (3EK, 3KE, 3ER, 3RE, 4EK, 4KE, 4ER, 4RE). The third row contains the ratio of the number of the two types of configurations. The configurations containing aspartic acids and those containing glutamic acid form two separate distributions if the individual numbers of the configurations are considered instead of their sum. The fourth row shows the *p*-value of the Wilcoxon rank sum test which confirms that the medians of the two distributions differ. The columns with the label *Scaled* contain the number of configurations scaled against the average frequency of occurrence of the amino acids in proteins.

Distribution of negatively and positively charged amino acids

n-cc	- + configuration	3DK 687	3EK 1109	3DR 630	3ER 1015	4DK 577	4EK 1081	4DR 526	4ER 896
	+ - configuration	3KD 369	3KE 817	3RD 317	3RE 654	4KD 510	4KE 890	4RD 321	4RE 754
	Difference	318	292	313	361	67	191	205	142
	<i>p</i> -value	7.812·10⁻³							

a2-cc	- + configuration	3DK 44	3EK 81	3DR 36	3ER 63	4DK 32	4EK 71	4DR 21	4ER 65
	+ - configuration	3KD 34	3KE 78	3RD 18	3RE 67	4KD 48	4KE 77	4RD 29	4RE 56
	Difference	10	3	18	-4	-16	-6	-8	9
	<i>p</i> -value	0.8438							

p2-cc	- + configuration	3DK 14	3EK 56	3DR 21	3ER 31	4DK 21	4EK 41	4DR 10	4ER 40
	+ - configuration	3KD 7	3KE 69	3RD 8	3RE 25	4KD 18	4KE 39	4RD 14	4RE 37
	Difference	7	-13	13	6	3	2	-4	3
	<i>p</i> -value	0.4062							

p2-cc + p3-cc + p4-cc	- + configuration	3DK 24	3EK 80	3DR 26	3ER 47	4DK 35	4EK 55	4DR 30	4ER 64
	+ - configuration	3KD 24	3KE 98	3RD 15	3RE 47	4KD 29	4KE 53	4RD 29	4RE 52
	Difference	0	-18	11	0	6	2	1	12
	<i>p</i> -value	0.4375							

Table 4: Distribution of negatively and positively charged amino acids in configurations which can form ionic interactions. The first row of each table displays the configurations where the first amino acid position is negatively and the second position positively charged. The second row shows the configurations where the first amino acid position is positively and the second position negatively charged. The third row displays the pairwise differences between the values of the first and second row. The fourth row shows the *p*-value of the Wilcoxon signed rank test that the pairwise differences deviate from zero. Only the non-coiled coils have a significant preference for a negatively charged amino acid in the first position and a positively charged amino acid in the second position.

Probability of salt-bridge formation in antiparallel 2-stranded coiled coils compared to non-coiled coils

i to *i*+3 configurations:

		formed	not formed			formed	not formed
n-cc	3DK	47	640	3KD	14	355	
a2-cc		6	38		2	32	
<i>p</i> -value		0.1228			0.6353		
n-cc	3EK	197	912	3KE	122	695	
a2-cc		17	64		11	67	
<i>p</i> -value		0.4551			1.0		
n-cc	3DR	239	391	3RD	19	298	
a2-cc		22	14		0	18	
<i>p</i> -value		7.814·10⁻³			0.6114		
n-cc	3ER	385	630	3RE	131	523	
a2-cc		19	44		13	54	
<i>p</i> -value		0.2304			1.0		

i to *i*+4 configurations:

		formed	not formed			formed	not formed
n-cc	4DK	99	478	4KD	244	266	
a2-cc		5	27		24	24	
<i>p</i> -value		1.0			0.88		
n-cc	4EK	142	939	4KE	190	700	
a2-cc		6	65		14	63	
<i>p</i> -value		0.385			0.5637		
n-cc	4DR	118	408	4RD	63	258	
a2-cc		9	12		1	28	
<i>p</i> -value		0.03704			0.04062		
n-cc	4ER	271	625	4RE	223	531	
a2-cc		23	42		15	41	
<i>p</i> -value		0.404			0.7617		

Table 5: Statistical analysis if the probability that salt bridges are formed is different in non-coiled coils and antiparallel 2-stranded coiled coils using the two-sided Fisher's exact test (Fisher, 1935). This is the case for 3DR, 4DR and 4RD salt bridges (bold). Each boldly bordered rectangle contains a contingency table. In the column left to each table is shown the type of configuration which was analysed. The upper row of each contingency table contains number of configurations in non-coiled coils, the lower row the number of configuration in antiparallel 2-stranded coiled coils. The first column of each contingency table contains the number of formed salt bridges and the second row the number of ionic interactions which did not form. The bottom line of shows the *p*-value of the test.

Probability of salt-bridge formation in parallel 2-stranded coiled coils and non-coiled coils

i to *i*+3 configurations:

		formed	not formed			formed	not formed
n-cc	3DK	47	640	3KD	14	355	
p2-cc		0	14		0	7	
<i>p</i> -value		0.6151			1.0		
n-cc	3EK	197	912	3KE	122	695	
p2-cc		6	50		9	60	
<i>p</i> -value		0.2081			0.8596		
n-cc	3DR	239	391	3RD	19	298	
p2-cc		3	18		0	8	
<i>p</i> -value		0.03644			1.0		
n-cc	3ER	385	630	3RE	131	523	
p2-cc		10	21		0	25	
<i>p</i> -value		0.5773			$7.723 \cdot 10^{-3}$		

i to *i*+4 configurations:

		formed	not formed			formed	not formed
n-cc	4DK	99	478	4KD	244	266	
p2-cc		0	21		7	11	
<i>p</i> -value		0.03417			0.483		
n-cc	4EK	142	939	4KE	190	700	
p2-cc		4	37		2	37	
<i>p</i> -value		0.6428			0.01356		
n-cc	4DR	118	408	4RD	63	258	
p2-cc		1	9		2	12	
<i>p</i> -value		0.6998			1.0		
n-cc	4ER	271	625	4RE	223	531	
p2-cc		5	35		7	30	
<i>p</i> -value		0.01975			0.1961		

Table 6a: Statistical analysis if the probability that salt bridges are formed is different in non-coiled coil and parallel 2-stranded coiled coils using two-sided Fisher's exact test (Fisher, 1935). This is the case for the 3DR, 3RE, 4DK, 4KE and 4ER salt bridges (bold). Each boldly bordered rectangle contains a contingency table. In the column left to each table is shown the type of configuration which was analysed. The upper row of each contingency table contains number of configurations in non-coiled coils, the lower row the number of configuration in parallel 2-stranded coiled coils. The first column of each contingency table contains the number of formed salt bridges and the second row the number of ionic interactions which did not form. The bottom line of shows the *p*-value of the test.

Probability of salt-bridge formation in antiparallel compared to parallel 2-stranded coiled coils

i to *i*+3 configurations:

		formed	not formed			formed	not formed
a2-cc	3DK	6	38	3KD	2	32	
p2-cc		0	14		0	7	
<i>p</i> -value		0.3192			1.0		
a2-cc	3EK	17	64	3KE	11	67	
p2-cc		6	50		9	60	
<i>p</i> -value		0.1626			1.0		
a2-cc	3DR	22	14	3RD	0	18	
p2-cc		3	18		0	8	
<i>p</i> -value		7.724·10⁻⁴			1.0		
a2-cc	3ER	19	44	3RE	13	54	
p2-cc		10	21		0	25	
<i>p</i> -value		1.0			0.01682		

i to *i*+4 configurations:

		formed	not formed			formed	not formed
a2-cc	4DK	5	27	4KD	24	24	
p2-cc		0	21		7	11	
<i>p</i> -value		0.144			0.5807		
a2-cc	4EK	6	65	4KE	14	63	
p2-cc		4	37		2	37	
<i>p</i> -value		1.0			0.08477		
a2-cc	4DR	9	12	4RD	1	28	
p2-cc		1	9		2	12	
<i>p</i> -value		0.1064			0.2433		
a2-cc	4ER	23	42	4RE	15	41	
p2-cc		5	35		7	30	
<i>p</i> -value		0.01219			0.4598		

Table 6b: Statistical analysis if the probability that salt bridges are formed is different in antiparallel and parallel 2-stranded coiled coils using the two-sided Fisher's exact test (Fisher, 1935). This is the case for the 3DR, 3RE and 4ER salt bridges (bold). If a significance level of 90 % is tolerated, then also the difference between the 4KE salt bridges is significant. Each boldly bordered rectangle contains a contingency table. In the column left to each table is shown the type of configuration which was analysed. The upper row of each contingency table contains number of configurations in antiparallel 2-stranded coiled coils and the lower row the number of configuration in parallel 2-stranded coiled coils. The first column of each contingency table contains the number of formed salt bridges and the second row the number of ionic interactions which did not form. The bottom line of shows the *p*-value of the test.

Probability of salt-bridge formation of XZ/ZX configurations in non-coiled coils

i to *i*+3 configurations:

	formed	not formed		formed	not formed
3DK	47	640	3EK	197	912
3KD	14	355	3KE	122	695
<i>p</i> -value	0.05201		<i>p</i> -value	0.1069	
	formed	not formed		formed	not formed
3DR	239	391	3ER	385	630
3RD	19	298	3RE	131	523
<i>p</i> -value	< 2.2·10⁻¹⁶		<i>p</i> -value	4.244·10⁻¹⁵	

i to *i*+4 configurations:

	formed	not formed		formed	not formed
4DK	99	478	4EK	142	939
4KD	244	266	4KE	190	700
<i>p</i> -value	< 2.2·10⁻¹⁶		<i>p</i> -value	1.624·10⁻⁶	
	formed	not formed		formed	not formed
4DR	118	408	4ER	271	625
4RD	63	258	4RE	223	531
<i>p</i> -value	0.3434		<i>p</i> -value	0.7874	

Table 7: Statistical analysis if the salt bridges in non-coiled coils form with different probability in XZ and ZX configurations using two-sided Fisher's exact test (Fisher, 1935). This is the case for the 3DR/3RD, 3ER/3RE, 4DK/4KD and 4EK/4KE configurations (bold). Each boldly bordered rectangle corresponds to a contingency table. The upper row of each contingency table contains the number of XZ configurations, the lower row the number of ZX configurations. The first column contains the number of formed salt bridges and the second row the number of salt bridges which did not form. The bottom line of shows the *p*-value of the test.

Probability of salt-bridge formation at different heptad positions in antiparallel 2-stranded coiled coils

i to *i*+3 configurations:

	formed	not formed		formed	not formed		formed	not formed
<i>be</i>	17	57	<i>cf</i>	15	92	<i>dg</i>	5	20
sum	84	333	sum	84	333	sum	84	333
<i>p</i> -value	0.6396		<i>p</i> -value	0.1675		<i>p</i> -value	1.0	
<i>ea</i>	7	24	<i>fb</i>	11	76	<i>gc</i>	29	62
sum	84	333	sum	84	333	sum	84	333
<i>p</i> -value	0.8167		<i>p</i> -value	0.1311		<i>p</i> -value	0.01808	

i to *i*+4 configurations:

	formed	not formed		formed	not formed		formed	not formed
<i>ae</i>	18	17	<i>bf</i>	25	83	<i>cg</i>	17	57
sum	97	298	sum	97	298	sum	97	298
<i>p</i> -value	$1.185 \cdot 10^{-3}$		<i>p</i> -value	0.8013		<i>p</i> -value	0.8828	
<i>eb</i>	12	65	<i>fc</i>	17	60	<i>gd</i>	8	12
sum	97	298	sum	97	298	sum	97	298
<i>p</i> -value	0.1038		<i>p</i> -value	0.7711		<i>p</i> -value	0.1836	

Table 8: Statistical analysis if the probability that salt bridges are formed is different from the mean at certain heptad positions in antiparallel 2-stranded coiled coils, using two-sided Fisher's exact test (Fisher, 1935). The ionic interactions at a specific heptad position are compared against the sum of the salt bridges at all heptad positions. The *i* to *i*+3 and *i* to *i*+4 salt bridges are treated separately. The *i* to *i*+3 salt bridges from a *g* to a *c* position and the *i* to *i*+4 salt bridges from an *a* to an *e* position are formed with a different probability than the average (bold). Each boldly bordered rectangle contains a contingency table. The upper row of each contingency table contains number of configurations at a specific heptad position, the lower row the sum of the salt bridges at all heptad positions. The first column of each contingency table contains the number of formed salt bridges and the second row the number of ionic interactions which did not form. The bottom line of shows the *p*-value of the test.

Probability of salt-bridge formation at different heptad positions in parallel 2-stranded coiled coils

i to *i*+3 configurations:

	formed	not formed		formed	not formed		formed	not formed
<i>be</i>	5	52	<i>cf</i>	2	30	<i>dg</i>	0	10
sum	26	203	sum	26	203	sum	26	203
<i>p</i> -value	0.8117		<i>p</i> -value	0.5471		<i>p</i> -value	0.6066	
<i>ea</i>	2	12	<i>fb</i>	5	54	<i>gc</i>	11	43
sum	26	203	sum	26	203	sum	26	203
<i>p</i> -value	0.6677		<i>p</i> -value	0.642		<i>p</i> -value	0.1131	

i to *i*+4 configurations:

	formed	not formed		formed	not formed		formed	not formed
<i>ae</i>	8	9	<i>bf</i>	3	44	<i>cg</i>	8	47
sum	27	192	sum	27	192	sum	27	192
<i>p</i> -value	$9.834 \cdot 10^{-4}$		<i>p</i> -value	0.3151		<i>p</i> -value	0.6542	
<i>eb</i>	4	43	<i>fc</i>	2	38	<i>gd</i>	2	4
sum	27	192	sum	27	192	sum	27	192
<i>p</i> -value	0.6183		<i>p</i> -value	0.2738		<i>p</i> -value	0.1733	

Table 9: Statistical analysis if the probability that salt bridges are formed is different from the mean at certain heptad positions in parallel 2-stranded coiled coils, using two-sided Fisher's exact test (Fisher, 1935). The ionic interactions at a specific heptad position are compared against the sum of the salt bridges at all heptad positions. The *i* to *i*+3 and *i* to *i*+4 salt bridges are treated separately, however. The *i* to *i*+4 salt bridges from an *a* to an *e* position is formed with a different probability than the average (bold). Each boldly bordered rectangle contains a contingency table. The upper row of each contingency table contains number of configurations at a specific heptad position, the lower row the sum of the salt bridges at all heptad positions. The first column of each contingency table contains the number of formed salt bridges and the second row the number of ionic interactions which did not form. The bottom line of shows the *p*-value of the test.

Side chain rotamers of the 3DR and 3RD salt bridges

3DR formed		D			
		$\alpha m-10^\circ$ 75%	$\alpha t60^\circ$ 19%	- 6%	Total
R	mtm-85° (13%)	5	1		6
	tpt85° (5%)	2	1		3
	tpt180° (5%)	1			1
	mtt85° (4%)	2		1	3
	tmm105° (2%)	2			2
	tpt-105° (1%)	1			1
	mtp-105° (0%)			1	1
	- (21%)	1	2	2	5
	Total	14	4	4	22

3RD formed		D			
		$\alpha m-10^\circ$ (75%)	$\alpha t60^\circ$ (19%)	- 6%	Total
R	tpt85° (3%)	1			1
	tpt180° (3%)			1	1
	tpp180° (1%)		3		3
	mmt180° (1%)	1			1
	mmm180° (< 1%)	1			1
	- (21%)	7	5		12
	Total	10	8	1	19

Table 10: Side chain rotamers in 3DR salt bridges in antiparallel 2-stranded coiled coils and 3RD ionic interactions in non-coiled coils. The rotamer nomenclature and frequencies are taken from (Lovell *et al.*, 2000). "-" describes side chain orientations which do not correspond to a defined rotamer. In the 3DR ionic interaction the side chains can both reside in the most favourable conformation (D: $\alpha m-10^\circ$, R: mtm-85°). In contrast, to form a 3RD electrostatic interaction it is not possible for both side chains to assume an energetically favoured conformation simultaneously.

Environment of the 3DR configurations in parallel and antiparallel 2-stranded coiled coils

3DR side chains	buried		partly buried		exposed		other attractive polar interactions	
	D	R	D	R	D	R	D	R
p2-cc, salt bridge formed	1 33.3%	1 33.3%	1 33.3%	1 33.3%	1 33.3%	1 33.3%	2 66.6%	3 100%
p2-cc, salt bridge not formed	-	-	4 23.5%	6 35.3%	13 76.5%	11 64.7%	10 58.8%	11 64.7%
a2-cc, salt bridge formed	11 50.0%	9 40.9%	8 36.4%	9 40.9%	3 13.6%	4 18.2%	20 90.9%	13 59.1%
a2-cc, salt bridge not formed	5 45.5%	3 27.3%	1 9.0%	3 27.3%	5 45.5%	5 45.5%	7 63.6%	6 54.5%

Mean of side chain B -factors	p2-cc formed	p2-cc not formed	a2-cc formed	a2-cc not formed
D B -factor (standard deviation)	68.07 (18.50)	67.25 (28.25)	25.36 (12.27)	35.10 (19.02)
R B -factor (standard deviation)	60.13 (19.04)	70.75 (33.78)	27.45 (13.84)	36.85 (24.76)

Table 11: The environment of the side chains of residues in 3DR configurations in parallel and antiparallel 2-stranded coiled coils in the respective crystal structures. The static solvent accessibility of the side chains (C_{α} onwards) was calculated with the program *NACCESS* (Hubbard and Thornton, 1993) which uses the algorithm by (Lee and Richards, 1971). The calculated number is the percentage accessibility a relative to the extended tripeptide A-D-A or A-R-A. Crystal contacts were taken into account. The side chain was classified to be *buried* if $a < 10\%$, *partly buried* if $10\% \leq a \leq 40\%$ and *exposed* if $a > 40\%$. The last column in the upper table shows which percentage of side chains makes attractive polar interactions (ionic interactions and hydrogen bonds within a distance cut-off of 3.8 Å and no B -factor cut-off) in addition to the 3DR salt bridge, excluding side chain - main chain contacts. The lower table gives the mean of the temperature factors of the side chains (C_{β} onward).

α -helical peptides

a)

Assessed interaction	Mean ellipticity θ_{222} ($^{\circ}$ cm ² dmol ⁻¹) at pH 7.0, 0.1 M NaCl, 0°C	Peptide
3EK	-17 600	Ac-(AXAAX) ₃ A-NH ₂
3KE	-8 500	
3DR	-16 300	Ac-(AXAAX) ₃ Y-NH ₂
3RD	-2 200	
3ER	-22 400	
3RE	-11 800	
4EK	-29 000	Ac-A(XAAAX) ₃ A-NH ₂
4KE	-25 300	
4DR	-25 200	Ac-(AXAAX) ₃ Y-NH ₂
4RD	-14 400	
4ER	-31 300	
4RE	-22 500	

b)

Side chain-side chain interaction	ΔG_{sc} (cal/mol) at pH 7.0, 0.1 M NaCl, 0°C	Peptide
3DK	-120	Ac-AAQAAXAQXAAAQAAY-NH ₂
3KD	-400	
3EK	-290	
3KE	-280	
4DK	-240	Ac-AAQAAXAQAXAAQAAY-NH ₂
4KD	-580	
4EK	-415	
4KE	-400	

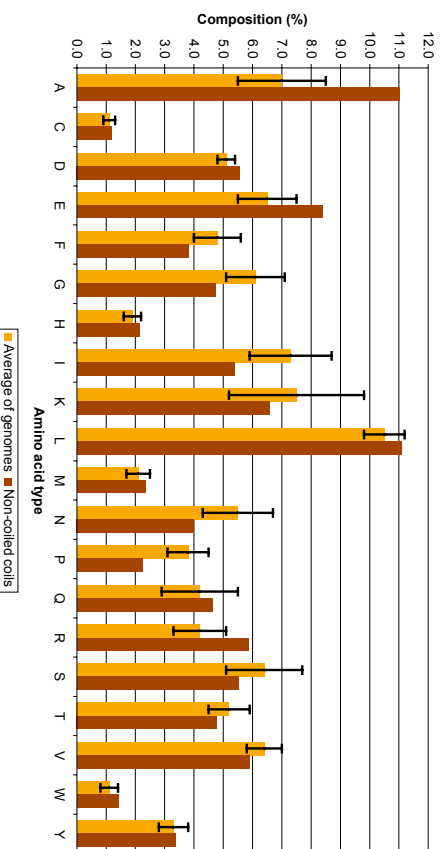
Table 12: a) Measured mean ellipticity of synthetic α -helical peptides at 222 nm wavelength in 1 mM phosphate buffer, pH 7.0, 0.1 M NaCl at 0°C. The peptides were designed and characterized by circular dichroism by (Huyghues-Despointes *et al.*, 1993; Marqusee and Baldwin, 1987). The X residues in the third column correspond to the interacting residues given in the first column. The pattern of stabilisation is 4AB > 4BA \approx 3AB > 3BA, where A denotes the acidic and B the basic residue of the assessed configurations. The peptide containing a C-terminal tyrosine cannot be directly compared to those containing alanine, because of the strong contribution of the helical tyrosine to the CD-signal.

b) Energetic contribution to peptide stability by side chain interactions (Smith and Scholtz, 1998). The error of the measurements is ± 80 cal mol⁻¹. Negative numbers indicate a stabilising interaction.

Figures

a)

Average composition of 8 microbial genomes and non-coiled coils from the protein database



b)

Proportion of charged amino acids

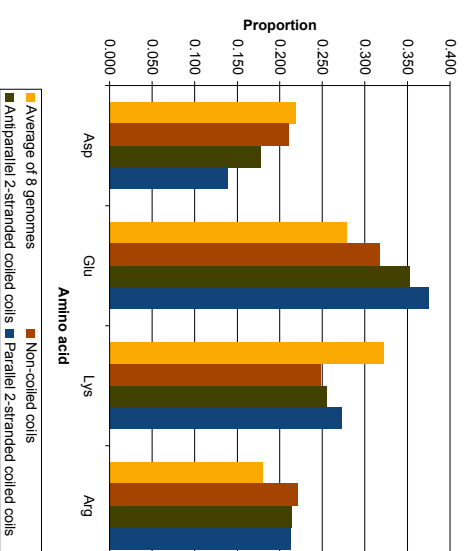


Figure 1a): The yellow columns show the average frequencies of the amino acids from the following 8 different microbial genomes: *Haemophilus influenzae*, *Mycoplasma genitalium*, *Methanococcus jannaschii*, *Synechocystis* sp., *Mycoplasma pneumoniae*, *Saccharomyces cerevisiae*, *Helicobacter pylori*, *Escherichia coli*. The error bars indicate the standard deviations. The data originate from (Gerstein, 1998). The brown columns show the amino acid composition of the non-coiled coils determined by the program SBSCC.

Figure 1b) The columns show the relative proportion of the charged amino acids in the average genome (yellow columns), non-coiled coils (brown columns), antiparallel 2-stranded coiled coils (green columns) and parallel 2-stranded coiled coils (blue columns). The data have been scaled such that the sum of the charged amino acids per data set equals 1.

Formed intrahelical salt bridges in non-coiled coils

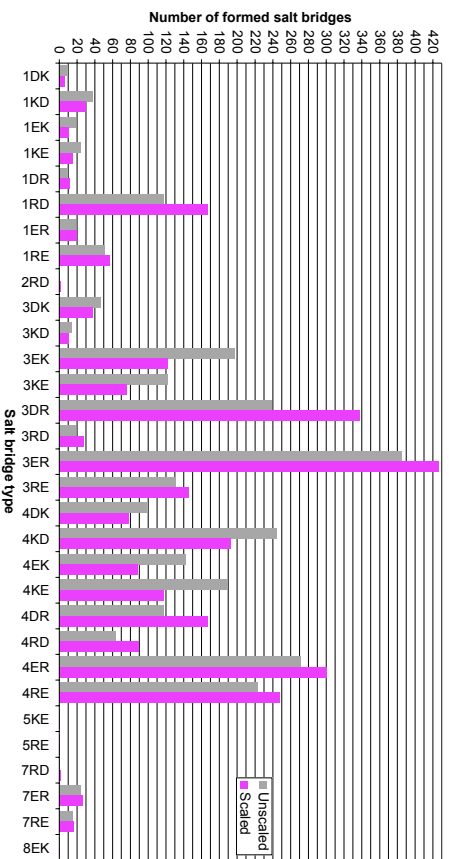


Figure 2a

Formed and possible intrahelical salt bridges in non-coiled coils

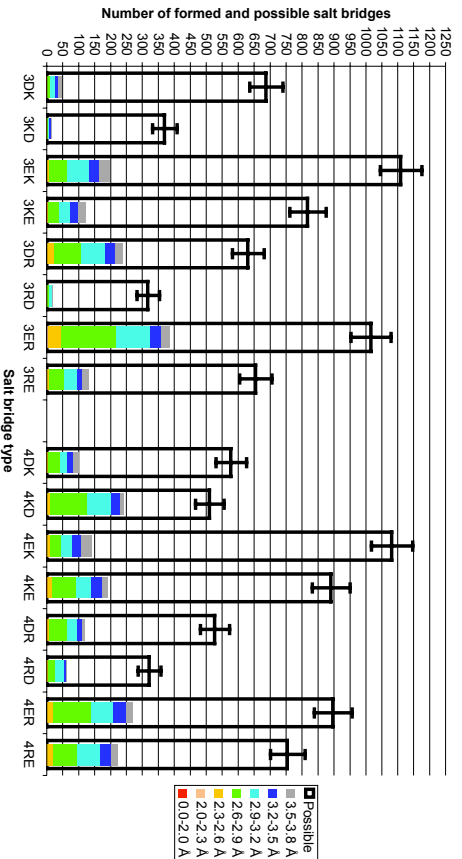


Figure 2c

Possible intrahelical salt bridges in non-coiled coils

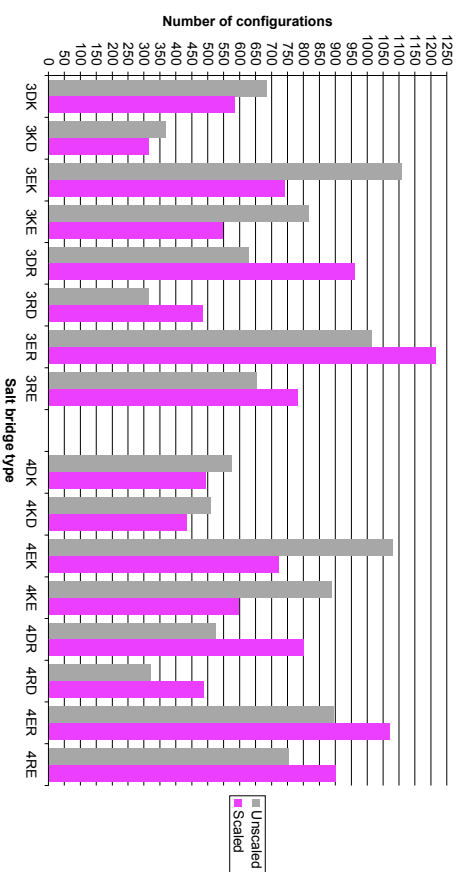


Figure 2b

Percentage of formed intrahelical salt bridges in non-coiled coils

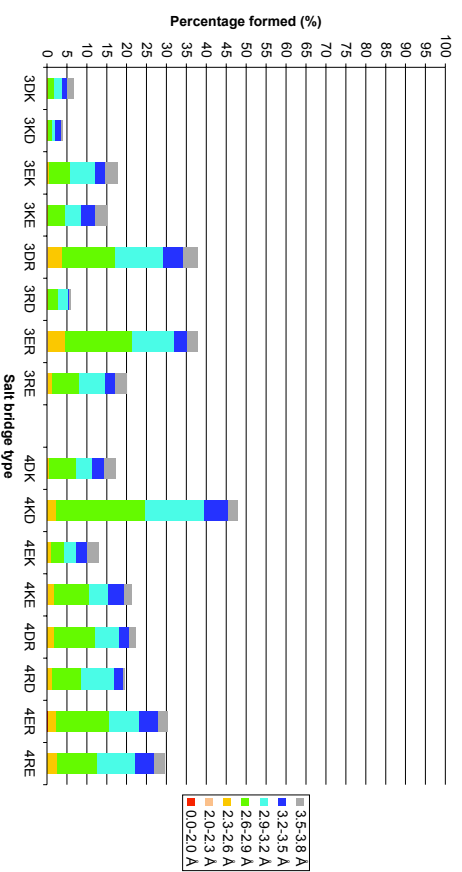


Figure 2d

Percentage of formed intrahelical salt bridges in non-coiled coils

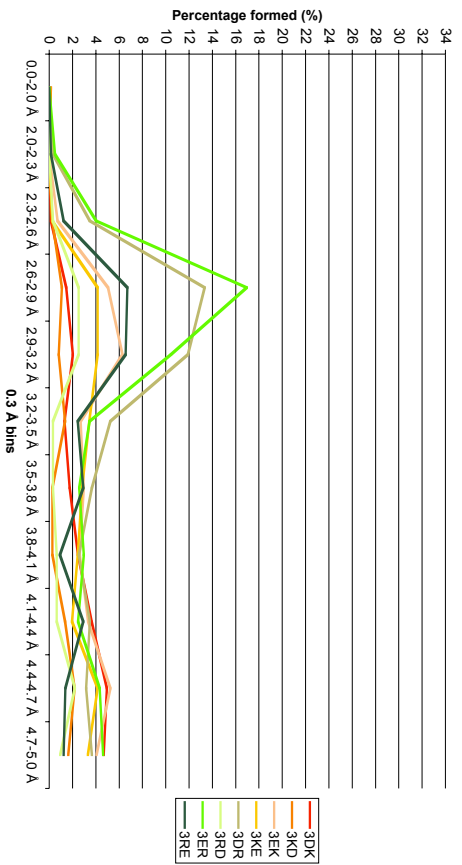


Figure 2e

Percentage of formed intrahelical salt bridges in non-coiled coils

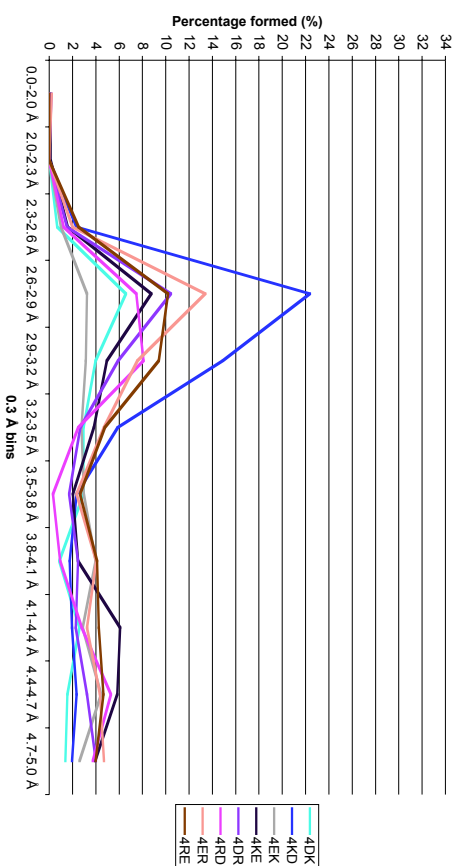


Figure 2f

Figure 2a: Number of formed intrahelical salt bridges identified by the program SBSSC in non-coiled coils. The grey columns show the counted numbers and the magenta columns show the numbers scaled by the average frequencies of amino acids in proteins.

Figure 2b: Number of possible intrahelical salt bridges in non-coiled coils. The grey columns show the counted numbers and the magenta columns the numbers scaled by the average frequencies of amino acids in proteins.

Figure 2c: Intrahelical salt bridges in non-coiled coils. The black margined columns show the counted number of configurations. The error bars mark the 95 % confidence limits of the counts if a Poisson distribution of the configurations is assumed. The filled columns show the number of formed salt bridges, colour coded according to the legend. Distances larger than 3.8 Å are not shown.

Figure 2d: Percentage of formed salt bridges in non-coiled coils. Only ionic interactions with a distance larger than 3.8 Å are shown. The salt bridges are divided into distance bins and the areas inside the columns are colour coded according to the legend.

Figure 2e: Relative frequencies of the i to $i+3$ intrahelical salt bridges in each distance bin in non-coiled coils.

Figure 2f: Relative frequencies of the i to $i+4$ intrahelical salt bridges in each distance bin in non-coiled coils.

Formed intrahelical salt bridges in a2 coiled coils

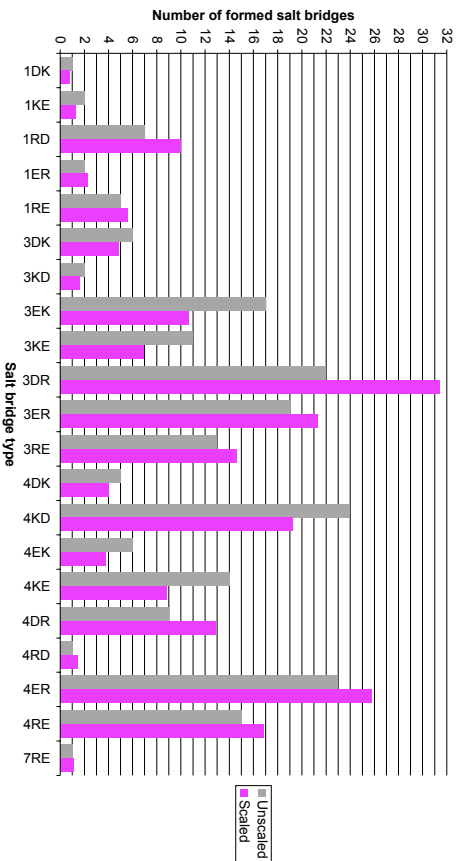


Figure 3a

Formed and possible intrahelical salt bridges in a2 coiled coils

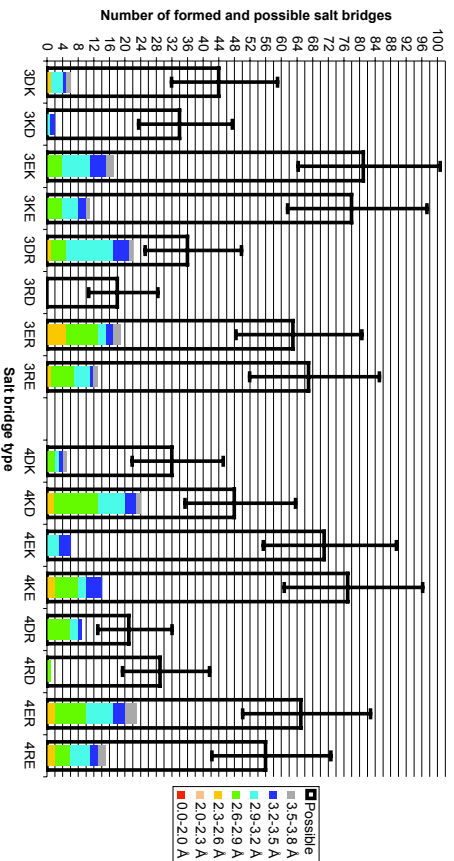


Figure 3c

Possible intrahelical salt bridges in a2 coiled coils

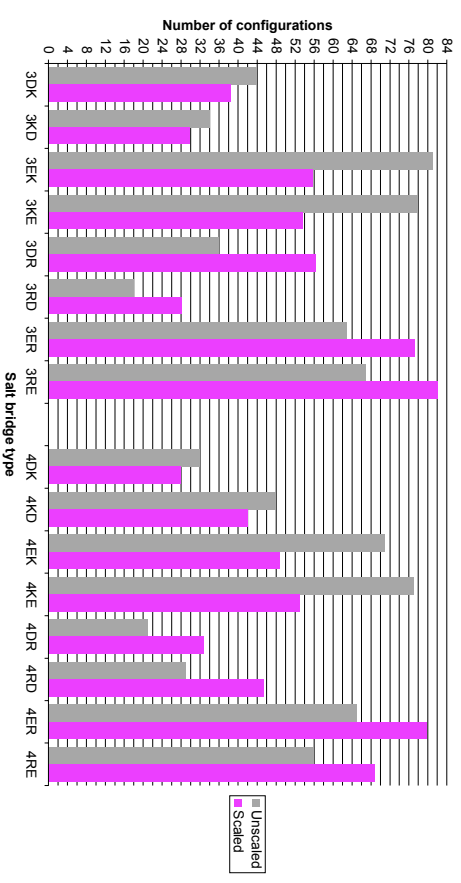


Figure 3b

Percentage of formed intrahelical salt bridges in a2 coiled coils

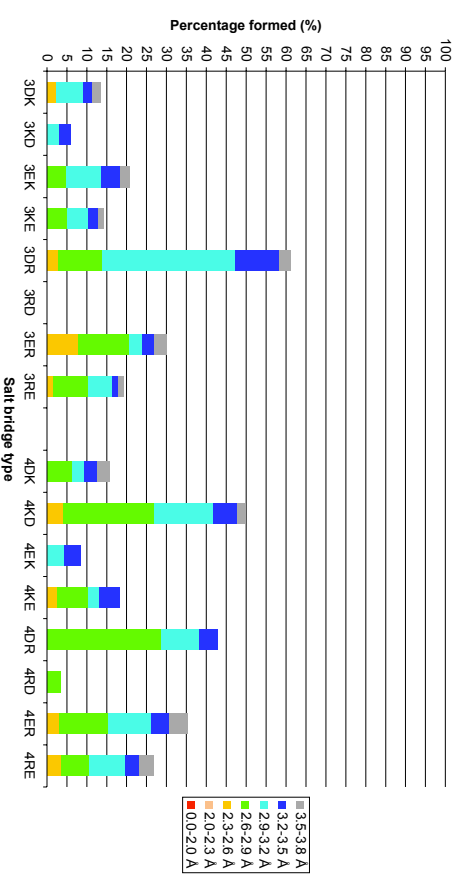


Figure 3d

Percentage of formed intrahelical salt bridges in a2 coiled coils

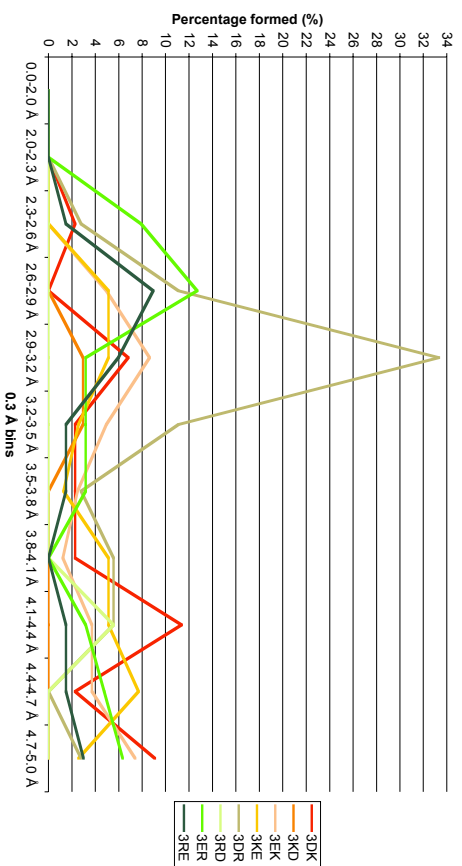


Figure 3e

Formed and possible intrahelical salt bridges along the heptad repeat in antiparallel 2-stranded coiled coils

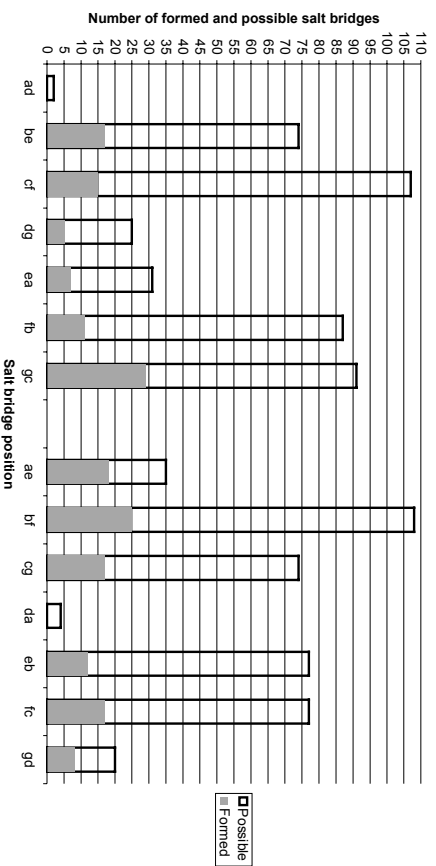


Figure 3g

Percentage of formed intrahelical salt bridges in a2 coiled coils

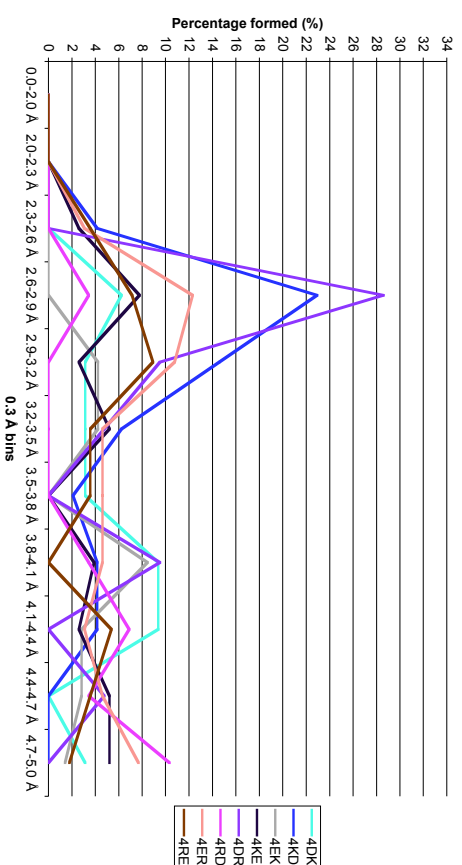


Figure 3f

Percentage of formed intrahelical salt bridges along the heptad repeat in antiparallel 2-stranded coiled coils

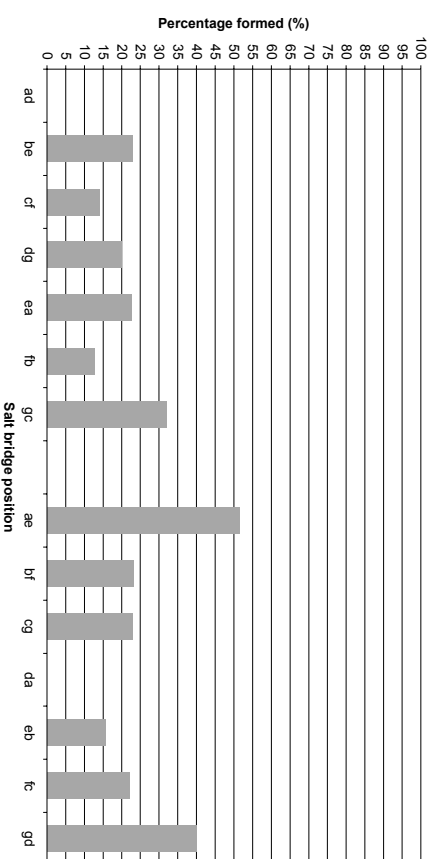


Figure 3h

Figure 3a: Number of formed intrahelical salt bridges identified by the program SBSSCC in antiparallel 2-stranded coiled coils. The grey columns show the counted numbers and the magenta columns show the numbers scaled by the average frequencies of amino acids in proteins.

Figure 3b: Number of possible intrahelical salt bridges in antiparallel 2-stranded coiled coils. The grey columns show the counted numbers and the magenta columns the numbers scaled by the average frequencies of amino acids in proteins.

Figure 3c: Intrahelical salt bridges in antiparallel 2-stranded coiled coils. The black margined columns show the counted number of configurations. The error bars mark the 95 % confidence limits of the counts if a Poisson distribution of the configurations is assumed. The filled columns show the number of formed salt bridges, colour coded according to the legend. Distances larger than 3.8 Å are not shown.

Figure 3d: Percentage of formed salt bridges in antiparallel 2-stranded coiled coils. Only ionic interactions with a distance larger than 3.8 Å are shown. The salt bridges are divided into distance bins and the areas inside the columns are colour coded according to the legend.

Figure 3e: Relative frequencies of the i to $i+3$ intrahelical salt bridges in each distance bin in antiparallel 2-stranded coiled coils.

Figure 3f: Relative frequencies of the i to $i+4$ intrahelical salt bridges in each distance bin in antiparallel 2-stranded coiled coils.

Figure 3g: Distribution of the intrahelical ionic interactions along the heptad repeat in antiparallel 2-stranded coiled coils. The columns left from the gap are the i to $i+3$ salt bridges and those right from the gap are the i to $i+4$ electrostatic interactions. The black margined columns show the counted number of configurations and the filled columns the number of formed salt bridges. There are apparent differences to the data in Figure 3c because intrahelical salt bridges can simultaneously be part of several coiled coils. Such ionic interactions occur several times in this figure, once for each coiled coil they are part of. And some salt-bridge configurations could not be assigned to a particular heptad position because they reside in coiled-coil stutters or stammers.

Figure 3h: Distribution of the intrahelical ionic interactions along the heptad repeat in antiparallel 2-stranded coiled coils. The columns show the percentage of formed electrostatic interactions. There are apparent differences to the data in Figure 3d because intrahelical salt bridges can simultaneously be part of several coiled coils. Such ionic interactions occur several times in this figure, once for each coiled coil they are part of. And some salt-bridge configurations could not be assigned to a particular heptad position because they reside in coiled-coil stutters or stammers.

Percentage of formed intrahelical salt bridges in p2 coiled coils

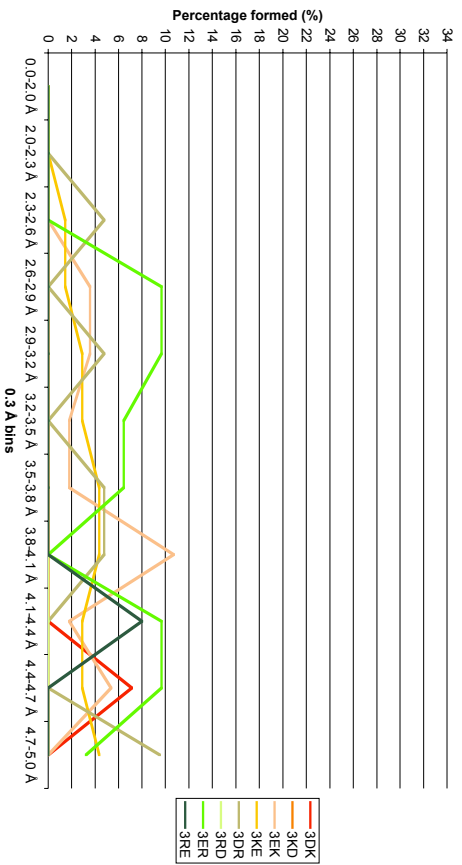


Figure 4e

Formed and possible intrahelical salt bridges along the heptad repeat in parallel 2-stranded coiled coils

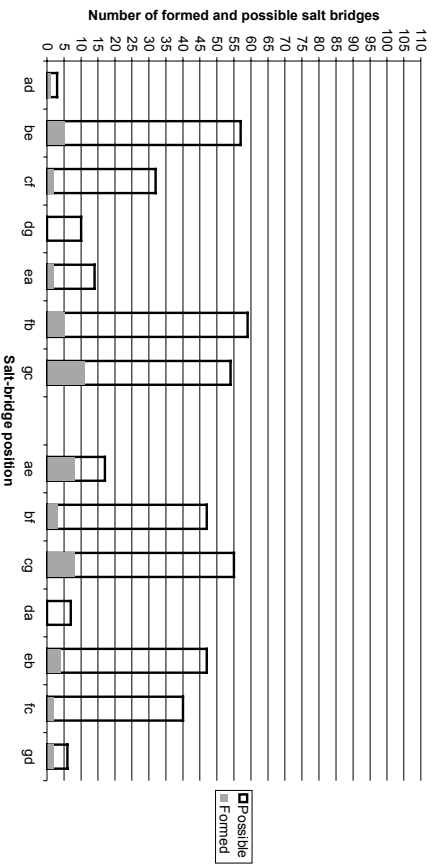


Figure 4g

Percentage of formed intrahelical salt bridges in p2 coiled coils

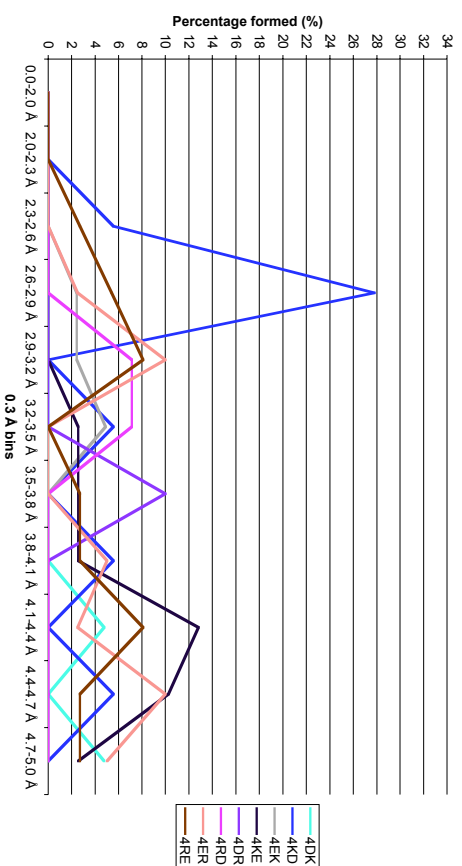


Figure 4f

Percentage of formed intrahelical salt bridges along the heptad repeat in parallel 2-stranded coiled coils

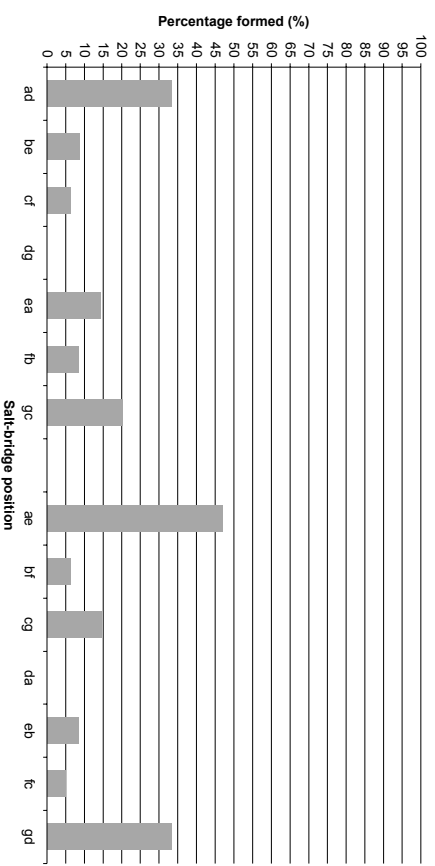


Figure 4h

Figure 4a: Number of formed intrahelical salt bridges identified by the program SBSCC in parallel 2-stranded coiled coils. The grey columns show the counted numbers and the magenta columns show the numbers scaled by the average frequencies of amino acids in proteins.

Figure 4b: Number of possible intrahelical salt bridges in parallel 2-stranded coiled coils. The grey columns show the counted numbers and the magenta columns the numbers scaled by the average frequencies of amino acids in proteins.

Figure 4c: Intrahelical salt bridges in parallel 2-stranded coiled coils. The black margined columns show the counted number of configurations. The error bars mark the 95 % confidence limits of the counts if a Poisson distribution of the configurations is assumed. The filled columns show the number of formed salt bridges, colour coded according to the legend. Distances larger than 3.8 Å are not shown.

Figure 4d: Percentage of formed salt bridges in parallel 2-stranded coiled coils. Only ionic interactions with a distance larger than 3.8 Å are shown. The salt bridges are divided into distance bins and the areas inside the columns are colour coded according to the legend.

Figure 4e: Relative frequencies of the i to $i+3$ intrahelical salt bridges in each distance bin in parallel 2-stranded coiled coils.

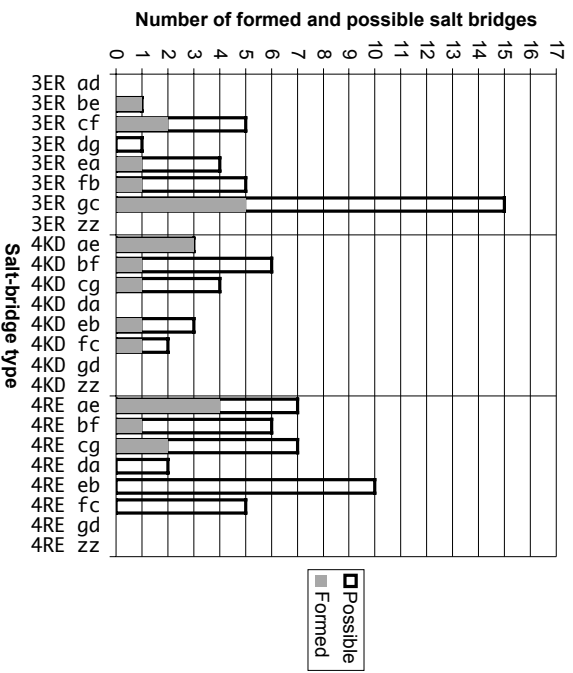
Figure 4f: Relative frequencies of the i to $i+4$ intrahelical salt bridges in each distance bin in parallel 2-stranded coiled coils.

Figure 4g: Distribution of the intrahelical ionic interactions along the heptad repeat in parallel 2-stranded coiled coils. The columns left from the gap are the i to $i+3$ salt bridges and those right from the gap are the i to $i+4$ electrostatic interactions. The black margined columns show the counted number of configurations and the filled columns the number of formed salt bridges. There are apparent differences to the data in Figure 4c because intrahelical salt bridges can simultaneously be part of several coiled coils. Such ionic interactions occur several times in this figure, once for each coiled coil they are part of. And some salt-bridge configurations could not be assigned to a particular heptad position because they reside in coiled-coil stutters or stammers

Figure 4h: Distribution of the intrahelical ionic interactions along the heptad repeat in parallel 2-stranded coiled coils. The columns show the percentage of formed salt bridges. There are apparent differences to the data in Figure 4d because intrahelical salt bridges can simultaneously be part of several coiled coils. Such ionic interactions occur several times in this figure, once for each coiled coil they are part of. And some salt-bridge configurations could not be assigned to a particular heptad position because they reside in coiled-coil stutters or stammers.

a)

Important intrahelical salt bridges in parallel 2-stranded coiled coils



b)

Important intrahelical salt bridges in antiparallel 2-stranded coiled coils

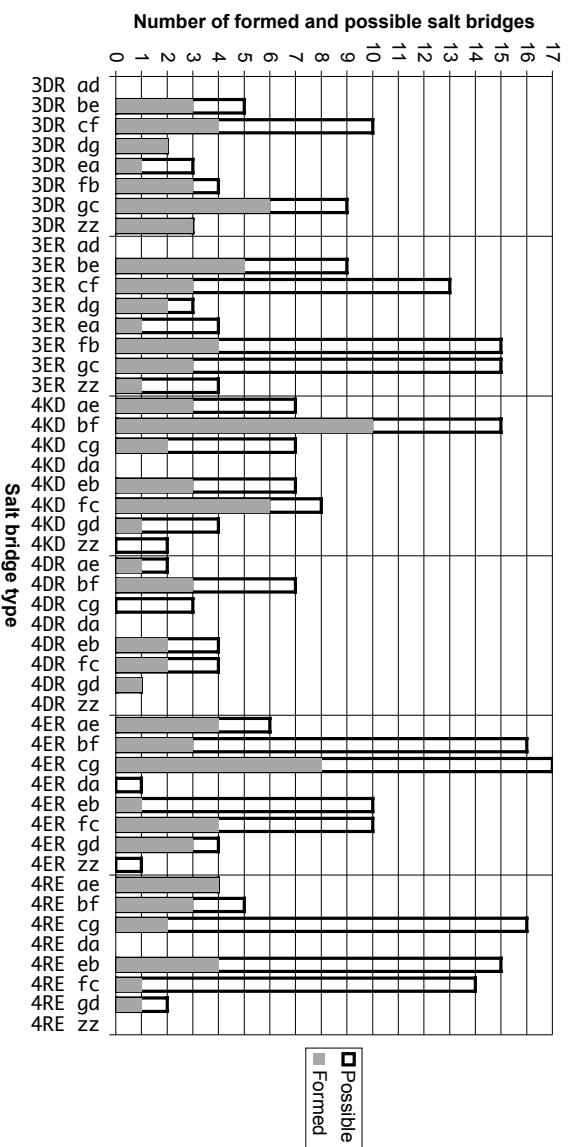


Figure 5a: Distribution of important intrahelical salt bridges along the heptad repeat in parallel 2-stranded coiled coils.

Figure 5b: Distribution of important intrahelical salt bridges along the heptad repeat in antiparallel 2-stranded coiled coils.

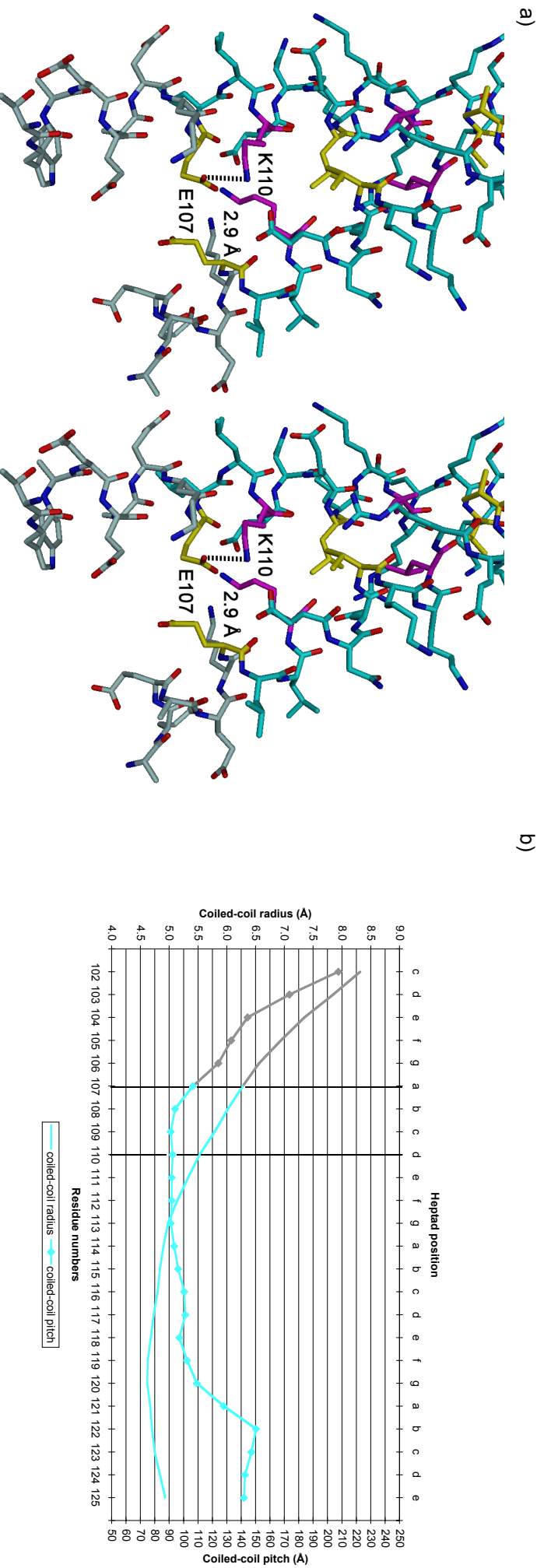


Figure 6a: Stereo view on the intrahelical 3EK salt bridge from an *a* to *d* position in the parallel 2-stranded coiled coil of a HAP1/DNA complex. The picture was generated with the program DINO (Philippson, 2002) from the coordinates 1HWT (King *et al.*, 1999). Residues at an *a* position are coloured yellow and residues at a *d* position magenta. Other residues which are part of the coiled coil are dyed cyan.

Figure 6b: Residual radius and pitch of the coiled coil shown in Figure 6a as calculated by the program TWISTER (Strekov and Burkhard, 2002). The cyan part of the curve denotes the extent of the knobs-into-holes interactions as identified by SBSCC. The grey part of the curve marks the remaining part of the α -helices. The location of the 3EK salt bridge from an *a* to a *d* position is marked by two perpendicular lines. The salt bridge resides at the *N*-terminal end of the coiled coil and induces separation of the α -helices, thus terminating the coiled coil.

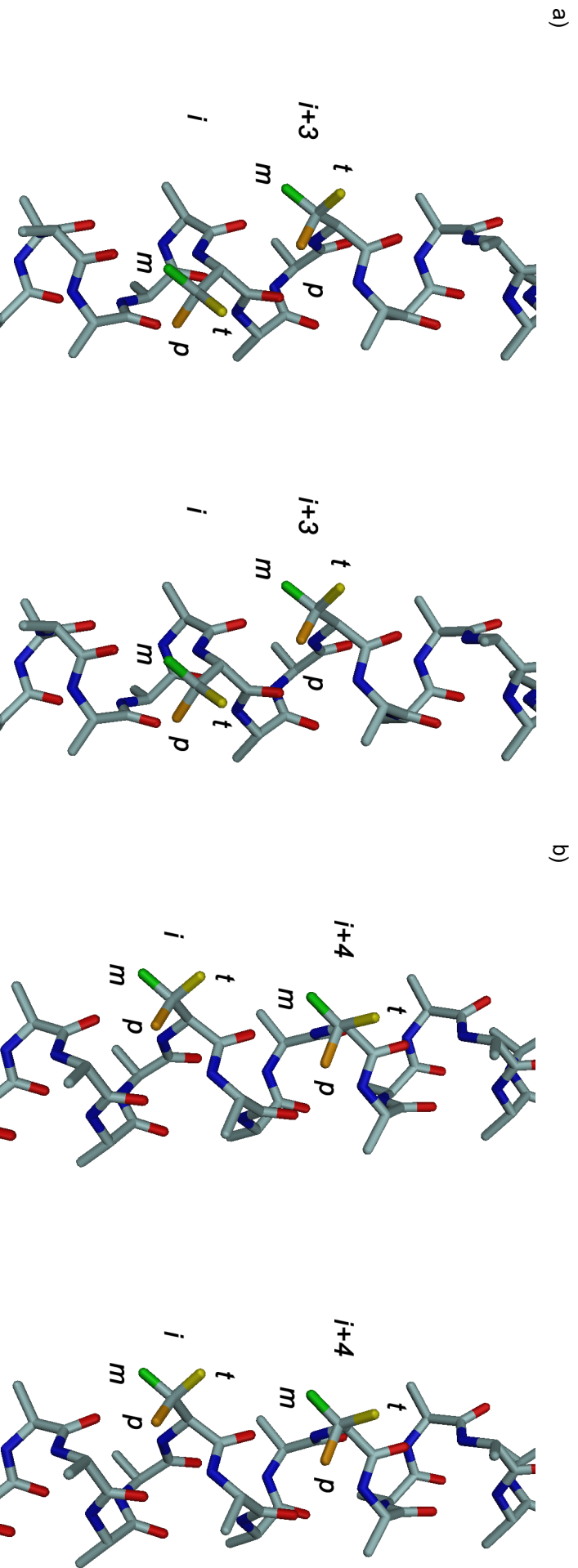


Figure 7: Stereo diagrams of the χ_1 angles in i to $i+3$ and i to $i+4$ side chains interactions. The m angle is coloured green, the t angle yellow and the forbidden p angle orange. The picture was generated with the program DINO (Philippsen, 2002) and is an adaptation of a figure in (Klingler and Brüttag, 1994).

a) To form an optimal i to $i+3$ interactions both χ_1 angles should be in the m orientation. If t/t , m/t and t/m angle orientations are combined, the side chains point away from each other. Because the $i+3$ position is shifted to the left compared to the i position when viewed from the N- to C-terminus, the side chain at the $i+3$ position should be longer than the side chain at the i position in order to make an optimal interaction.

b) To form the best possible i to $i+4$ side chain interaction the χ_1 angle at position i should be in t and the angle at $i+4$ in m orientation. m/m and t/t combinations can also form interactions. Side chain with m/t χ_1 combination point in opposite directions and can therefore not interact. In contrast to the i to $i+3$ interaction, the position $i+4$ is only slightly shifted to right compared to the i position when viewed from the N- to C-terminus. Thus the two positions are basically located above each other. For an optimal t/m interaction, the residues should have approximately the same length.

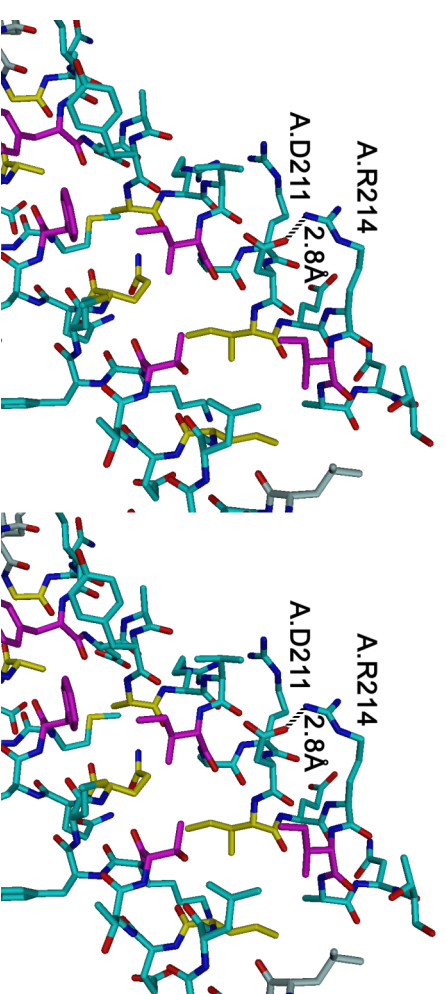
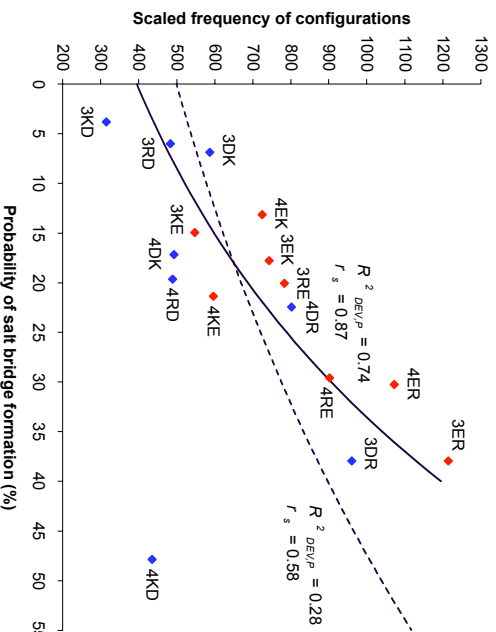


Figure 8: Stereo view on a typical 3DR salt bridge from an *g* to *c* position in the antiparallel 2-stranded coiled coil of aspartase of *Bacillus* sp. YM55-1. Both residues are in the most frequent rotamer position ($\alpha\text{m}-10^\circ$ and $\text{mtm}-85^\circ$) and a hydrogen bond with optimal geometry is formed. The coiled coil is embedded in the protein context. The arginine of the salt bridge is partly solvent accessible, and it makes also a ionic interaction with Asp272 of a neighboured molecule (not shown). The aspartate of the salt bridge is completely buried. The picture was generated with the program DINO (Philippsen, 2002) from the coordinates 1J3U (Fuji *et al.*, 2003). Residues at an *a* position are coloured yellow and residues at a *d* position magenta. Other residues which are part of the coiled coil are dyed cyan.

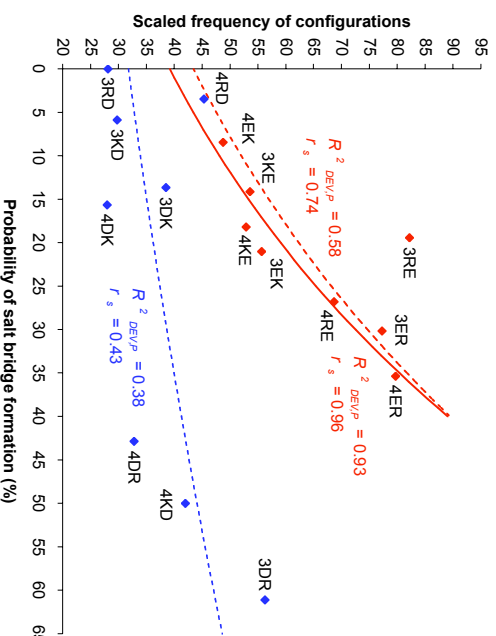
a)

Scaled frequencies vs. probabilities in non-coiled coils



b)

Scaled frequencies vs. probabilities in antiparallel 2-stranded coiled coils



c)

Scaled frequencies vs. probabilities in parallel 2-stranded coiled coils

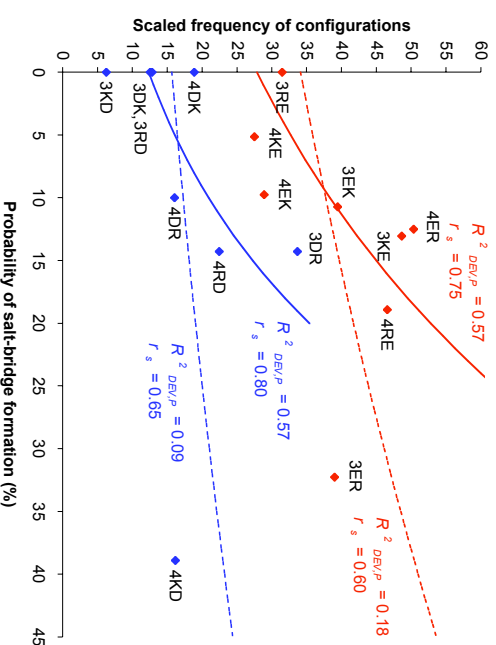


Figure 9: Scaled frequencies of configurations versus probabilities of salt bridge formation in a) non-coiled coils, b) antiparallel 2-stranded coiled coils and c) parallel 2-stranded coiled coils. A standard Poisson linear regression with a log-link function was calculated. A black curve includes all data, whereas a red curve includes only configurations containing glutamic acid and a blue curve only configurations containing aspartic acid. If a fit was calculated excluding an outlier, it is drawn as only configurations. A fit which was calculated including all corresponding data points is drawn as a solid curve. The following data points were treated as outliers: 4KD in non-coiled coils, 3RE in antiparallel 2-stranded coiled coils and 3ER and 4KD in parallel 2-stranded coiled coils. The goodness of fit $R^2_{DEV/P}$ and Spearman's rank correlation coefficient r_s are indicated. The number of configurations where scaled by the average amino acid frequencies in genomes. The application of a Poisson model for the fit is therefore not really appropriate.

Curriculum vitae

Name: Markus Meier

Present work address: The Institute of Materials Science
University of Connecticut
97 North Eagleville Road
Storrs, CT 06269-3136, USA
Phone: +1 860 486 3830
E-mail: markus.meier@uconn.edu

Former work Address: M. E. Müller Institut
Biozentrum der Universität Basel
Klingelbergstrasse 70
CH-4056 Basel, Switzerland
Phone: +41 61 267 20 91
E-mail: m.meier@unibas.ch

Personal: Date of birth: December 26, 1974
Place of origin: Flaach (ZH)
Nationality: Swiss

Education: 1982-1986 Primary school, Erlensträsschen, Riehen, BS
1986-1987 Secondary school, Bäumlhof, Basel, BS
1987-1991 Secondary school, Spiegelfeld, Binningen, BL
1991-1994 High school, Oberwil, BL
1995 Military training, Brugg, AG
1995-2000 University of Basel, Biologie II (Molecular Biology)

Degrees: December 1994 Matura
March 2000 University diploma in Biology II
June 2004 Ph.D. in Biophysics

Awards: October 2004 One-year Fellowship for Prospective Researchers of the Swiss National Science Foundation

Research experience: 1999-2000 Diploma thesis in protein crystallography:
Protein X-ray Crystallography Studies
(Advisors Dr. P. Burkhard and Prof. U. Aebi)
2000-2004 M.E. Müller Institut, University of Basel
Ph.D. project:
Implications of molecular interactions for protein structure, function and design
(Advisors PD Dr. P Burkhard and Prof. U. Aebi)
2004-2005 Postdoctoral fellow at The Institute of Materials Science, University of Connecticut, USA

Scientific training: 8-10/1998 Industrial training, F. Hoffmann-La Roche AG, division PRPV / Thrombosis
(directed by Dr. M. Riederer)
3-14/5/1999 Training in Protein Purification and Expression, Deutsches Krebsforschungszentrum, Heidelberg, Germany
(directed by PD Dr. H.Herrmann)
13-17/5/2001 6th International School on the Crystallography of Biological Macromolecules, Como, Italy
7-14/4/2002 Phasing and Refinement of Macromolecular Structures, York, United Kingdom

Conferences: 2000 Invited speaker at the 14th REGIO meeting, Braunwald, Switzerland

	2001	Invited speaker at the Alpbach Workshop "Coiled-coils, Collagen & Co-Proteins: III", Alpbach, Austria
	2002	Invited speaker at the 5 th International Symposium on Vitamin B6, PQQ, Carbonyl Catalysis and Quinoproteins, Southampton, United Kingdom
	2003	Invited speaker at the 4 th International Homocysteine Conference, Basel, Switzerland
Work experience:	1995-1996	Security employee, Securitas AG, Basel
	1998-2004	Application programming project in the field of geographical information systems (GIS) and virtual reality (VR): <i>VR-GI Modeler</i> (<i>p.p.</i> Vermessungsamt Basel-Stadt and Vermessungsamt Bern)
	2001-2004	System and software administrator, Department of Structural Biology, University of Basel
Teaching experience:	2001, 2002	Tutorial instructor in biology, University of Basel
Language skills:		German (mother tongue), English (fluent), French

List of publications

- Meier, M., Bider, M.D., Malashkevich, V.N., Spiess, M. and Burkhard, P. (2000) Crystal structure of the carbohydrate recognition domain of the H1 subunit of the asialoglycoprotein receptor. *J Mol Biol*, **300**, 857-865.
- Burkhard, P., Meier, M. and Lustig, A. (2000) Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci*, **9**, 2294-2301.
- Janosik, M., Meier, M., Kery, V., Oliveriusova, J., Burkhard, P. and Kraus, J.P. (2001) Crystallization and preliminary X-ray diffraction analysis of the active core of human recombinant cystathionine beta-synthase: an enzyme involved in vascular disease. *Acta crystallographica Sect D*, **57**, 289-291.
- Meier, M., Janosik, M., Kery, V., Kraus, J.P. and Burkhard, P. (2001) Structure of human cystathionine beta-synthase: a unique pyridoxal 5'-phosphate-dependent heme protein. *Embo J*, **20**, 3910-3916.
- Meier, M., Lustig, A., Aebi, U. and Burkhard, P. (2002) Removing an interhelical salt bridge abolishes coiled-coil formation in a de novo designed peptide. *J Struct Biol*, **137**, 65-72.
- Meier, M., Oliveriusova, J., Kraus, J.P. and Burkhard, P. (2003) Structural insights into mutations of cystathionine beta-synthase. *Biochim Biophys Acta*, **1647**, 206-213.