Internet Electronic Journal of Molecular Design **2004**, *3*, 295–307 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

June 2004, Volume 3, Number 6, Pages 295–307

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday
Part 12

Guest Editor: Douglas J. Klein

Molecular Modeling of *c2h2* Zinc Finger Mutation of Putative Human Transcription Factor SALL4

Vladimír Frecer,[1,2] Jan Miertus,[3] Wiktor Borozdin,[4] Jürgen Kohlhase,[4] Antonio Amoroso,[3] and Stanislav Miertus [1]

[1] International Centre for Science and High Technology, UNIDO, AREA Science Park, Padriciano 99, Trieste I–34012, Italy
[2] Cancer Research Institute, Slovak Academy of Sciences, Vlárska 7, Bratislava SK–83391, Slovakia
[3] Pediatric Hospital, IRCCS Burlo Garofolo, Via dell'Istria 65/1, Trieste I–34137, Italy
[4] Institut für Humangenetik und Anthropologie, Universität Freiburg, Breisacher Str. 33, Freiburg D-79106, Germany

Received: October 1, 2003; Accepted: March 17, 2004; Published: June 30, 2004

Citation of the article:
V. Frecer, J. Miertus, W. Borozdin, J. Kohlhase, A. Amoroso, and S. Miertus, Molecular Modeling of *c2h2* Zinc Finger Mutation of Putative Human Transcription Factor SALL4, *Internet Electron. J. Mol. Des.* **2004**, *3*, 295–307, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Molecular Modeling of *c2h2* Zinc Finger Mutation of Putative Human Transcription Factor SALL4[#]

Vladimír Frecer,[1,2] Jan Miertus,[3] Wiktor Borozdin,[4] Jürgen Kohlhase,[4] Antonio Amoroso,[3] and Stanislav Miertus [1,*]

[1] International Centre for Science and High Technology, UNIDO, AREA Science Park, Padriciano 99, Trieste I–34012, Italy
[2] Cancer Research Institute, Slovak Academy of Sciences, Vlárska 7, Bratislava SK–83391, Slovakia
[3] Pediatric Hospital, IRCCS Burlo Garofolo, Via dell'Istria 65/1, Trieste I–34137, Italy
[4] Institut für Humangenetik und Anthropologie, Universität Freiburg, Breisacher Str. 33, Freiburg D-79106, Germany

Received: October 1, 2003; Accepted: March 17, 2004; Published: June 30, 2004

Internet Electron. J. Mol. Des. **2004**, *3* (6), 295–307

Abstract

Motivation. The SALL4 gene, Drosophila's region specific homeotic sal (spalt)–like gene family member, encodes for a zinc finger (ZF) transcription factor (TF). Mutations of the SALL4 have been demonstrated to cause the Okihiro syndrome a combination of Duane retraction syndrome and radial ray defects. We have studied a missense point mutation of the SALL4 (nucleotide substitution 2663A→G, residues substitution His:888→Arg:888) positioned within the first *c2h2* ZF of the C–terminal double ZF motif in the SALL4 gene identified in an Italian three–generation family, some members of which displayed the clinical features of the Okihiro syndrome. We performed a molecular modeling study on the wild type (*wt*) and mutated (*mt*) ZF domains of the SALL4 TF with the goal to propose a plausible hypothesis relating the modeled structural and energetic differences between the *wt* and *mt* forms to the defects connected with the observed mutation.

Method. Sequence alignment, homology protein modeling and molecular mechanics using CFF91 force field were utilized to build and refine the ZF models and to estimate their stability and DNA–binding affinity.

Results. We have modeled *wt* and *mt* ZF motifs of the SALL4 TF based on sequence homology with ZF domains of TFs with known crystal structures co–crystallized with a B–DNA segment. Secondary structure, zinc ion binding and DNA binding of the two static ZF models were analyzed in terms of mutual r.m.s. deviations and intramolecular and intermolecular interaction energies.

Conclusions. The modeled *wt* and *mt* forms of ZF motif of the SALL4 TF did not display significant structural differences cased by steric strain or charge of the bulkier Arg:888 and retained similar supersecondary structures and comparable strength of the zinc ion binding. However, more significant differences were predicted in their binding affinities to DNA. Calculated higher DNA binding affinity (and possibly also changed specificity) of the *mt* form of the ZF could be the reason for the altered activator/repressor function of the mutant form of the SALL4 TF at its natural target gene or the cause for erroneous targeting of a different DNA sequence of the same or another gene. We may thus hypothesize that the pathogenic effects of the mutation could be related to the altered regulation function by making the dissociation of the *mt* SALL4 TF–gene adducts more difficult.

[#] Dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday.

* Correspondence author; phone: 39–040–922–8114; fax: 39–040–922–8115; E–mail: Stanislav.Miertus@ics.trieste.it.

Keywords. SALL4; transcription factor; zinc finger; mutant form; Okihiro syndrome; molecular mechanics.

Abbreviations and notations	
H888R, mutation His:888 to Arg:888	TF, transcription factor
mt, mutant form	*wt*, wild type
SALL, sal (spalt)–like gene family member	ZF, zinc finger

1 INTRODUCTION

During the last decade or so it is becoming increasingly evident that methods of molecular modeling display great potential in the elucidation and interpretation of structures and properties of complex molecular and supramolecular systems and in the mechanisms of biochemical and biological processes. Attempts to utilize theoretical approaches appeared recently also in the field of human and molecular genetics [1] where information obtained from protein structure predictions were used to understand complex phenomena on the genetic level. It is evident that big gaps exist between clinical phenomena, biochemical processes and changes in the related macromolecular structures, nevertheless, putting together complementary pieces of specific information originating from various disciplines is useful for the formulation of hypotheses and design of experiments or molecular simulations, which may help to elucidate such complicated processes.

Recently, our group [2] followed a case of a missense point mutation in the SALL4 gene present in an Italian family. Three members of this three–generation family displayed clinical features of the Okihiro syndrome, a combination of Duane retraction syndrome and radial ray defects [3]. Moreover, two younger members (2^{nd} and 3^{rd} generation) display a midline cranial structures defects. The SALL4 mutations have been demonstrated to cause the Okihiro syndrome [3,4] and other related phenotypes [5]. The SALL4 gene, the Drosophila's region specific homeotic sal (spalt)–like gene family member, encodes for a putative zinc finger (ZF) transcription factor (TF). This gene is responsible for the regulation of the transcription of other target genes, however, exact function of the SALL4 need to be confirmed experimentally. Human SALL1, the best–studied spalt–like family member, has been demonstrated to act as a transcription repressor [6].

The observed SALL4 mutation (nucleotide substitution 2663A→G) corresponds to His:888 to Arg:888 (H888R) amino acid substitution positioned within the first *c2h2* ZF of the C–terminal double ZF motif and affects one of the four amino acids directly involved in the zinc binding. Pathogenic point mutations of the sal genes described to date belong to the truncation mutations or the splice site mutation types [7,8]. Recently, it was proposed that such mutations of the sal TFs cause birth defects [9]. To confirm the pathogenic effect of a specific missense mutation only the affected members of the family must bear this mutation. Moreover, fifty healthy control subjects must not display this mutation. The number of controls needed depends also on the level of conservation of the mutated amino acid in the primary structure of the protein. Functional studies (at the transcriptome/proteome levels) are difficult to perform, because of the unknown protein function and/or target genes of the SALL4 TF. Therefore, we decided to perform a molecular

modeling study on the wild type and mutated SALL4 TFs with the following goals: (*i*) to construct a 3D model of the relevant wild type (*wt*) ZF domain of the SALL4 TF in complex with a DNA segment and to evaluate its stability as well as important intramolecular and intermolecular interactions; (*ii*) to model the H888R mutant form (*mt*) of the ZF domain in the complex with DNA, to evaluate the structural and energetic changes caused by the mutation both on intra– and intermolecular levels; and (*iii*) to propose a plausible hypothesis relating the modeled structural and energetic differences to the pathogenic effects connected with the observed mutation.

2 MATERIALS AND METHODS

2.1 Sequence Alignment and Model Building

The alignment of ZF domain sequences of human TFs with known structures was carried out using the SIM alignment tool [10] and multiple sequence alignment tool DIALING [11]. The sequences and the 3D structures of the considered TFs were obtained from the Protein Data Bank (PDB) [12]. The ZF domains of selected homologous TFs were visualized and superimposed in the *Insight–II* molecular modeling package [13]. The first *c2h2* ZF of the C–terminal double ZF motif in the SALL4 TF containing the studied mutation H888R was modeled from crystal structures of the consensus sequences of homologous ZF domains of TFs [14–17] (PDB entry codes: 1SP2, 1ZNM, 1JK2 and 1UBD) using the human Yy1 ZF (1UBD) [17] as a template. In the models of the *wt* and H888R *mt* ZF the side chains of non–conserved residues were replaced using *Biopolymer* module of the *Insight–II* [13]. The initial conformation of the side chains was manually selected from the library of side chain rotamers [18] based of inter–residue interaction energies. The interaction energies were computed by molecular mechanics (MM) using class II consistent force field CFF91 and charge parameters [19]. Both the *wt* and *mt* models were considered to be at the pH of 6.5 with all protonable and ionizable residues charged. The ZF cysteine residues involved in the Zn^{2+} ion coordination were considered in ionized state with a net partial charge of –0.5 ē on the sulphur atom while the histidine residues coordinating the zinc were considered neutral [20]. The net atomic charges of these residues were derived by semiempirical AM1 calculation [21] of optimised ZF model complex composed of two deprotonated cysteine residues (Cys**:**872 and Cys**:**875) and two neutral histidine residues (His**:**888 and His**:**892) coordinating the zinc ion preserving the crystal structure geometry of the considered residues in the human Yy1 ZF (1UBD).

2.2 Molecular Mechanics Calculations

In MM calculations the force field parameters of the Zn^{2+} ion were approximated by the Ca^{2+} parameters due to non–availability of the CFF91 parameters for zinc. The mutated residue Arg**:**888 was considered in its protonated (cationic) form due to pK_a value of 12.5 of free guanidine group [12]. However, it is probable that the pK_a constant of arginine side chain will be significantly

lowered due to the vicinity of the Zn^{2+} ion [20,22]. An effective dielectric constant of 4 was used for all MM calculations in order to take into account the shielding effects in proteins. Minimizations of the *wt* and *mt* ZF models were carried out by slowly and gradually relaxing the structures (simulated annealing) starting with the side chains of the replaced residues, followed by full relaxation of the four residues coordinating the zinc ion and concluded by the relaxation of the whole peptide backbone. To simulate the flanking effects of the polypeptide chain, the backbone atoms of the N– and C–terminal residues of the model of the ZF motif were kept fixed in their original crystal structure positions. In the geometry optimization sufficient number of steepest descent and conjugate gradient iterative cycles with convergence criterion for the average gradient of 0.01 $kcal \cdot mol^{-1} \cdot Å^{-1}$ has been used.

Double helical 20 base–pair long DNA segment co–crystallized with the Yy1 TF was also added to the *wt* and *mt* ZF models to allow DNA–binding studies. The DNA segment was supplemented by 38 Na^+ counter ions coordinating the anionic phosphate groups and was kept fixed in its B–DNA crystal structure conformation during all ZF relaxations in both *wt* and *mt* models. To ensure equal positioning of the ZF helix towards the DNA segment in both models several levels of residue backbone fixing were applied during the gradual relaxations of the ZF models.

Relative changes in the zinc complexation energies and DNA binding energies between the *wt* and *mt* ZF models were considered as they are expected to lead to partial cancellation of errors caused by the approximate nature of the MM calculations and imperfection of the force field.

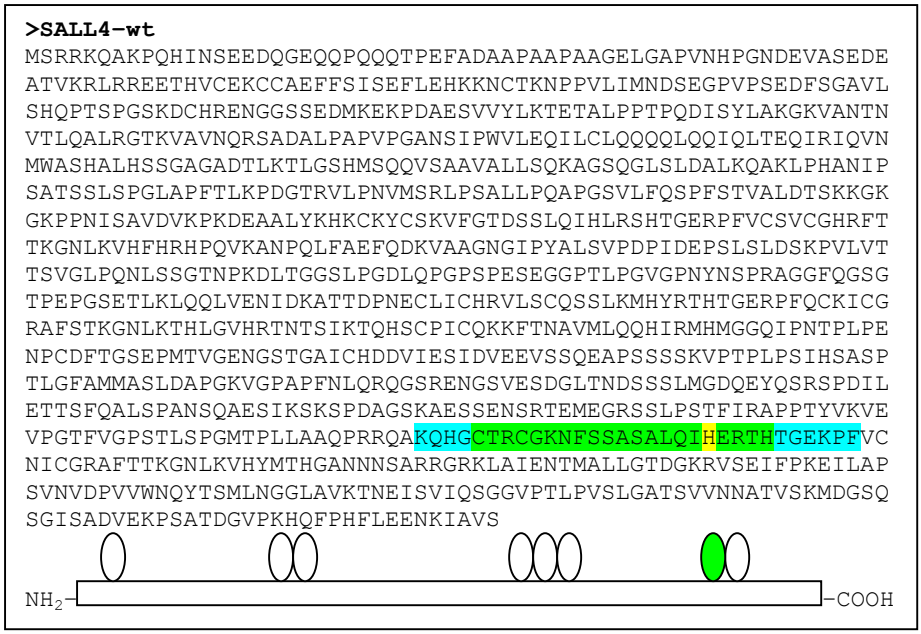


Figure 1. Sequence of the first *c2h2* ZF of wild type human SALL4 TF known to be important in the human malformation syndromes [4]. The modeled ZF domain is highlighted in green, the H888R mutation site in yellow and the flanking residues are shown in cyan. The position of the ZF domain in the SALL4 sequence (residues 872 – 892) is indicated in the scheme.

3 RESULTS AND DISCUSSION

3.1 Model Building of Zinc Finger Domain

The peptide sequence of the first *c2h2* ZF of the C–terminal double ZF motif of the SALL4 protein [4] (Figure 1), which contains the studied mutation H888R was aligned with sequences of TFs, the 3D crystal structures of which were available from the PDB.

The sequences of ZF domains of four TFs that showed the highest alignment score with the studied SALL4 ZF were aligned together and their crystal structures (1JKL, 1ZNM, 1SP2 and 1UBD [14–17]) were superimposed based on the sequence alignment (Figure 2).

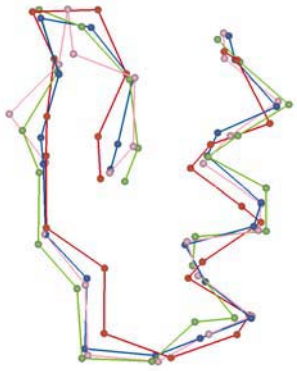
```
SALL4-ZF          1   -----CTR-- CGKNFSSASA LQIHERTH-- ---------- ----------
SP2               1   -RPFMCTWSY CGKRFTRSDE LQRHKRTHTG EK-------- ----------
ZNM               1   ---FQCTFXX CGKRFSLDFN LKTHVKIHTG ---------- ----------
JK2               1   -RPYACPVES CDRRFSRSAE LTRHIRIHTG QKPFQCri-- CMRNFSRSDH
UBD-ZF1          31   pRVHVCAE-- CGKAFVESSK LKRHQLVHTG EKPFQ----- ----------
UBD-ZF2          51   eKPFQCTFEG CGKRFSLDFN LRTHVRIHTG DRPYVCpfdg CNKKFAQSTN

                      ********* ********** ********** ******     **********
                      ********* ********** ********** ****
                          ***   ********** ********** **
                                ********** **********
                                ********** **********
```

A

```
55.2% identity in 29 residues overlap; Score: 92.0; Gap frequency: 0.0%
SALL4 ZF          1   HGCTRCGKNF SSASALQIHE RTHTGEKPF
UBD-ZF1          34   HVCAECGKAF VESSKLKRHQ LVHTGEKPF
                      | |  ||| |    | |  |    |||||||
```

B



C

Figure 2. A – DIALIGN [11] aligned sequences of ZF domains of four selected TFs, which showed high sequence homology with the *c2h2* ZF domain of SALL4 (1JK2, 1ZNM, 1SP2 and 1UBD [14–17]). **B** – SIM [10] aligned sequence of human Yy1 ZF domain [17] (1UBD), which showed the highest sequence homology to the studied ZF domain of SALL4 (Figure 1) and was used for the SALL4 ZF model building; **C** – consensus sequences of superimposed residues of aligned ZF domains, which show high conservation of the super–secondary structure of the ZF domains in TFs (r.m.s. deviations of superimposed backbone atoms of 22 residues (184 atoms) are: JKL – SP2 1.06 Å, UBD–ZF2 – SP2 0.85 Å, ZNM – SP2 1.50 Å). Only the traces of C_α atoms are shown for better clarity.

The consensus sequences of ZF domains of four TFs (Figure 2) showed highly conserved super–secondary structures of the ZF motif with an r.m.s. deviation of the backbone atoms not exceeding 1.5 Å. The *c2h2* ZF domains also displayed almost identical helix–β–hairpin motifs with the zinc ion holding together the hairpin loop, which contains the two Cys residues and the C–terminal end of the helix that accommodates the two His residues. The ZF helix is positioned in the major groove when bound to the DNA [17].

The crystal structure of human Yy1 ZF (1UBD), which showed the highest homology to the SALL4 ZF (Figure 2), was used as the template structure in the ZF homology model building (Figure 3).

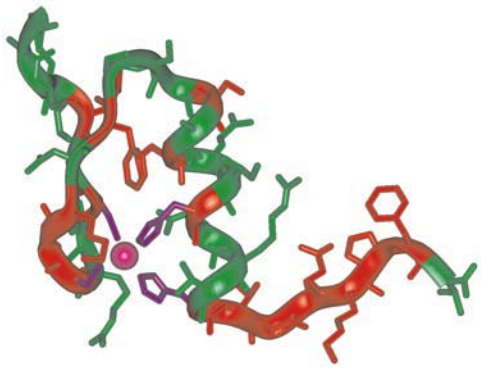


Figure 3. *c2h2* ZF domain of SALL4 model build by homology modeling using the Yy1 human TF [8] crystal structure (1UBD) as a template. The conserved residues are shown as red ribbon segments, replaced residues – green segments and the residues Cys**:**872, Cys**:**875, His**:**888 and His**:**892 coordinating the Zn^{2+} ion (purple sphere) are shown in dark purple color.

20–base–pair long segment of B–DNA double helix present in the crystal structure of human Yy1 ZF [17] was also added to the *wt* and *mt* ZF models to allow comparison of energetic aspects of DNA–binding. Two initial models containing *wt* SALL4 ZF and its mutant form with the His**:**888 residue coordinating the Zn^{2+} ion in the *wt* form replaced by bulkier and positively charged Arg residue, were built. Both models containing ZF bound to the frozen B–DNA segment were gradually relaxed using a simulated annealing protocol described in the Material and Methods section resulting in the initial (static) *wt* and *mt* SALL4 ZF models (Figure 4).

3.2 Analysis of Zinc Finger Models

3.2.1 Structural differences

The relaxed molecular models of the *wt* and *mt* ZF domain of SALL4 were analysed in terms of structural dissimilarities, energetic stabilities and differences in the intramolecular and intermolecular interactions of parts of the model with the zinc ion and the B–DNA segment. Calculated r.m.s. deviations (Table 1) that document the structural differences between the *wt* and

mt ZF models show that there is no significant difference in the overall fold of the ZF motifs between the two models since the r.m.s.d. of the backbone atoms of the superimposed models remains below 1.06 Å for the whole ZF as well as for its smaller parts (Figure 5).

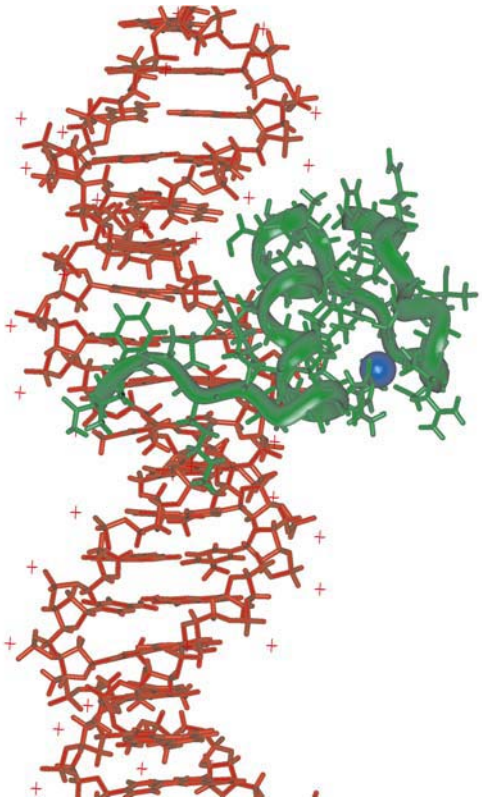


Figure 4. Initial model of the *c2h2* zinc ZF of *wt* human TF SALL4 bound to B–DNA segment. The B–DNA double helix with Na^+ counter ions is shown in red, and the backbone of the ZF is depicted as green solid ribbon. Hydrogen atoms were omitted for better clarity. The Zn^{2+} ion sphere is colored blue.

Table 1. Structural differences computed as r.m.s. deviations between the relaxed static models of *wt* and H888R *mt* SALL4 ZF domains.

Initial *wt* and H888R *mt* models of ZF domain of SALL4 human TF		Structural Difference [a]
wt c2h2 ZF	*mt c2h2* ZF, His:888 → Arg+	r.m.s.d. [Å]
β–hairpin, Ala:868 – Ser:880	β–hairpin, Ala:868 – Ser:880	0.98
α–helix, Ser:881 – Thr:893	α–helix, Ser:881 – Thr:893	0.62
tail, Gly:894 – Val:898	tail, Gly:894 – Val:898	0.78
Zn^{2+}	Zn^{2+}	0.91 [b]
whole ZF, Ala:868 – Val:898	whole ZF, Ala:868 – Val:898	1.06
Cys:872, Cys:875, His:888, His:892	Cys:872, Cys:875, Arg:888, His:892	0.78
wt residue, His:888	*mt* residue, Arg:888	0.03

[a] r.m.s.d. of peptide backbone atoms only;
[b] distance between Zn^{2+} ions in the superimposed *wt* and *mt* ZF models.

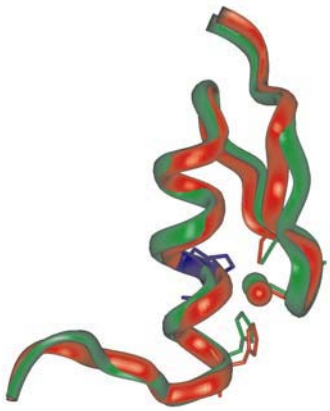


Figure. **5.** Overall fold of the superimposed relaxed models of *wt* and H888R *mt* forms of the SALL4 *c2h2* ZF domain. The backbone atoms are represented by green (*wt*) and red (*mt*) ribbons. Heavy atoms of side chains of the residues coordinating the zinc ion (green and red spheres) are depicted in stick representation. The *wt* His**:**888 and *mt* Arg**:**888 residues and their ribbon segments are colored blue.

The same conclusion was drawn for the *wt* and *mt* ZFs modeled in the absence of the B–DNA. These models resembled the super–secondary structures of their DNA–bound counterparts with r.m.s. deviations of all backbone atoms of 0.83 Å and 1.35 Å, respectively (Figure 6).

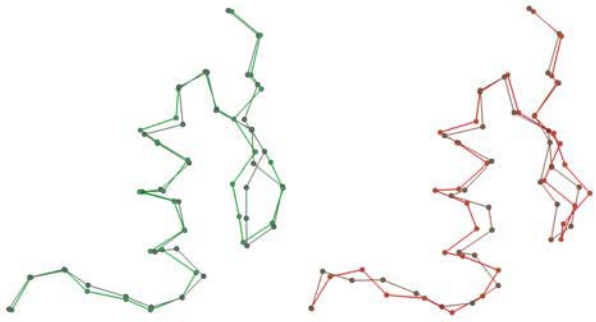


Figure 6. Superimposed *wt* (green) and *mt* (red) ZFs of the SALL4 TF modeled in the absence (lighter hue) and in the presence (darker hue) of the B–DNA segment. The r.m.s. deviation between backbone atoms of the *wt* models: *wt* – *wt*+DNA is equal to 0.83 Å and between the *mt* models: *mt* – *mt*+DNA is equal to 1.93 Å. Only the traces of C_α atoms traces are shown for better clarity.

Thus binding to DNA does not seem to change significantly the 3D structure of the ZF model. When comparing the binding of *wt* and *mt* ZF models to DNA main structural changes are observed on the level of individual residues, especially the mutated residue 888.

A closer look at the four residues coordinating the Zn^{2+} suggests that the cationic side chain of the Arg**:**888 residue in the *mt* ZF model points away from the zinc ion and is oriented towards the anionic sugar–phosphate backbone of the bound DNA segment (Figure 7).

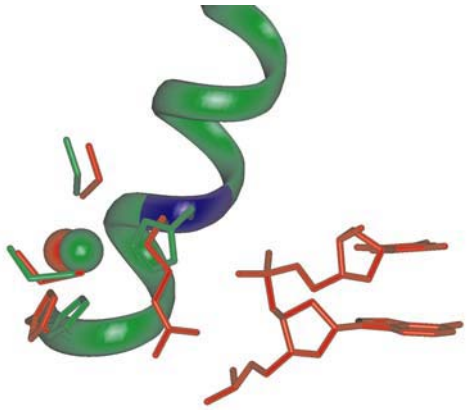


Figure 7. Superimposed residues coordinating the Zn^{2+} ion in the *wt* model of SALL4 ZF and its *mt* form. The side chains are represented as green (*wt*) and red (*mt*) sticks surrounding the Zn^{2+} ions (green and red spheres). The cationic side chain of the mutant Arg**:**888 residue points away from the zinc ion towards the anionic phosphates of the bound B–DNA segment. Only the ribbon of the helix of *wt* ZF is shown. Fragment of DNA sugar–phosphate backbone interacting with the Arg**:**888 is also shown in stick representation (red color).

The absence of significant structural differences between the static models of *wt* and *mt c2h2* ZFs of the SALL4 suggests that replacement of His:888 by the bulkier Arg residue may not significantly distort the ZF motif due to steric reasons.

3.2.2 Energetic differences in zinc complexation

Energetic aspects connected with the replacement of the His:888 side chain by most probably charged Arg side chain (pK_a = 12.5) may have more serious consequences upon the functions of the ZF motif of the SALL4 TF. Table 2 shows data relevant to zinc coordination and gives the interaction energy contributions of portions of the *wt* and *mt* ZF models with the zinc ion.

Table 2. CFF91 intramolecular interaction energies computed for the models of *wt* and *mt* SALL4 ZF domain and the zinc ion.

Interaction of *wt* and *mt* ZF models of human SALL4 TF with Zn^{2+}		Interaction Energy, CFF91 [kcal·mol^{-1}]		
wt ZF (Q_{wt}=4 ē)	System	E_{coul} (ε = 4) [a]	E_{vdW} [b]	E_{tot} [c]
His:888	Zn^{2+}	–15.0	1.8	–13.2
Cys:872, Cys:875, His:888, His:892	Zn^{2+}	–85.3	8.2	–77.1
whole ZF Ala:868 – Val:898	Zn^{2+}	–38.0	7.3	–30.7
mt ZF His:888→ Arg+ (Q_{mt}=5 ē)	System	E_{coul} (ε = 4)	E_{vdW}	E_{tot}
Arg+:888	Zn^{2+}	25.1	–0.2	24.9
Cys:872, Cys:875, Arg+:888, His:892	Zn^{2+}	–49.2	7.2	–42.0
whole ZF Ala:868 – Val:898	Zn^{2+}	–11.3	6.5	–4.8

[a] E_{coul} (ε = 4) is the coulombic contribution to the interaction energy computed with a dielectric constant ε=4;
[b] E_{vdW} is the van der Waals (dispersion and repulsion) contribution to the interaction energy;
[c] $E_{tot} = E_{coul}$ (ε = 4) + E_{vdW} is the total interaction energy.

Comparison between the *wt* and *mt* ZF models suggests that the stability of the ZF super–secondary structure (which can be related to the interaction energy of the zinc coordinating residues or the whole ZF with the Zn^{2+} ion) is significantly reduced in the case of the H888R mutant, c.f. E_{tot}[*wt*] = –30.7 kcal·mol^{-1}*vs*. E_{tot}[*mt*] = –4.8 kcal·mol^{-1}, respectively, for the whole ZFs (Table 2).

A similar ratio of the Zn^{2+} ion stabilizations in the ZFs was computed also for the *wt* and *mt* ZFs modeled in the absence of the DNA. It means that the stability of the ZF motif in the protein is weakened by the His residue substitution (and possibly also the resulting function of the TF is altered). However, also the *mt* ZF remains sufficiently stable due to a favourable complexation of Zn^{2+} ion by the remaining coordinating residues in Cys:872, Cys:875, His:892 (E_{tot} = –17.1 kcal·mol^{-1}).

3.2.3 Energetic differences in DNA binding

Binding affinity of TFs to specific DNA sequences of target genes is essential to the gene regulation function of these proteins. Even though it has not been demonstrated that the first *c2h2* ZF of the C–terminal double ZF motif plays a crucial role in DNA binding of the SALL4 TF an altered capacity of the *mt* ZF to interact with DNA can have serious consequences for the SALL4 TF function. Table 3 gives the interaction energy contributions of ZF residues or larger secondary structure motifs with the bound DNA segment for the static *wt* and *mt* ZF models.

Tabel 3. Computed CFF91 intermolecular interaction energies between the *wt* and *mt* models of SALL4 ZF domains and DNA.

Interaction of *wt* and *mt* ZF domain models of human SALL4 TF with B–DNA segment [a]		Interaction Energy, CFF91 [kcal·mol^{-1}]		
wt ZF (Q_{wt}=4 ē)	System	E_{coul} (ε = 4) [b]	E_{vdW} [c]	E_{tot} [d]
β–hairpin, Ala:868 – Ser:880	B–DNA	–83.6	–9.6	–93.2
helix, Ser:881 – Thr:893	B–DNA	–13.2	–34.5	–47.4
tail, Gly:894 – Val:898	B–DNA	–25.7	–2.7	–28.4
Zn^{2+}	B–DNA	–48.5	0.0	–48.5
whole ZF, Ala:868 – Val:898	B–DNA	–170.8	–46.7	–217.5
Cys:872, Cys:875, His:888, His:892, Zn^{2+}	B–DNA	–34.3	–3.8	–38.1
wt residue, His:888	B–DNA	–5.7	–2.6	–8.3
mt ZF His:888→ Arg+ (Q_{mut}=5 ē)	System	E_{coul} (ε = 4)	E_{vdW}	E_{tot}
β–hairpin, Ala:868 – Ser:880	B–DNA	–81.8	–14.1	–95.9
α–helix, Ser:881 – Thr:893	B–DNA	–42.7	–29.7	–72.4
tail, Gly:894 – Val:898	B–DNA	–29.6	–3.6	–33.2
Zn^{2+}	B–DNA	–45.6	0.0	–45.6
whole ZF, Ala:868 – Val:898	B–DNA	–199.7	–47.4	–247.1
Cys:872, Cys:875, Arg:888, His:892, Zn^{2+}	B–DNA	–65.7	–3.9	–69.6
mt residue, Arg:888	B–DNA	–40.0	–3.1	–43.1

[a] B–DNA double helix with Na^+ counter ions:
AGGGTCTCCATTTTGAAGCG
CGCTTCAAAATGGAGACCCT
[b] E_{coul} (ε = 4) is the coulombic contribution to the interaction energy computed with a dielectric constant ε=4;
[c] E_{vdW} is the van der Waals (dispersion and repulsion) contribution to the interaction energy;
[d] $E_{tot} = E_{coul}$ (ε = 4) + E_{vdW} is the total interaction energy.

The binding affinity of the ZF to the B–DNA (which can be related to the total interaction energy of the whole ZF with the B–DNA segment) changes when passing from the *wt* form to the *mt* form. However, in this case the interaction energies computed for the two models predict that the H888R mutant of SALL4 ZF is able to bind stronger to the B–DNA segment through its α–helix (containing the mutated residue) than the *wt* ZF, c.f. E_{tot}[*wt*] = –217.5 kcal·mol^{-1} *vs*. E_{tot}[*mt*] = –

247.1 kcal·mol^{-1}, respectively. The main difference in the interaction with DNA between the *wt* and *mt* ZF models can be attributed to coulombic (electrostatic) interactions of the neutral His:888 and mutated cationic Arg:888 with the DNA counterparts.

4 CONCLUSIONS

The modeled *wt* and *mt* forms of the first *c2h2* ZF of the C–terminal double ZF motif of SALL4 TF did not display significant structural differences between the models due to the steric strain or charge of the bulkier Arg:888 side chain. Both these static models retained similar supersecondary structures and comparable strength of the zinc ion binding. However, more significant differences between the *wt* and *mt* ZF models were predicted for their binding to DNA. Calculated higher DNA binding affinity (and possibly also altered specificity) of the *mt* form of the ZF (increased by approx. 14 %) could be the reason for the altered activator/repressor function of the H888R mutant form of the SALL4 TF at its natural target gene or the cause for erroneous targeting of a different DNA sequence of the same or another gene. The energetic analysis indicated that different DNA–binding properties of the two models are a direct consequence of the conservative His:888 → Arg mutation. We may thus hypothesize that the pathogenic effect of the studied H888R mutation of the SALL4 TF is related to the altered transcription regulation function by making the dissociation of the SALL4 TF–DNA adduct more difficult. This prediction suggests that the function of the mutant SALL4 TF could be elucidated by DNA–binding experiments. Additional pieces of useful information may come from more sophisticated simulations on the *wt* and *mt* ZF models that will probe their dynamic properties.

5 REFERENCES

[1] M. Tartaglia, E. L. Mehler, R. Goldberg, G. Zampino, H. G. Brunner, H. Kremer, I. van der Burgt, A. H. Crosby, A. Ion, S. Jeffery, K. Kalidas, M. A. Patton, R. S. Kucherlapati and B. D. Gelb, Mutations in PTPN11, Encoding the Protein Tyrosine Phosphatase SHP–2, Cause Noonan Syndrome, *Nature Genetics* **2001**, *29*, 465–468.

[2] J. Miertus, W. Borozdin, M. Liebers, V. Frecer, A. Amoroso, S. Miertus and J. Kohlhase, A *SALL4* Zinc Finger Mutation Predicted to Result in Increased DNA Binding Affinity is Associated with a Combination of Duane Anomaly, Cranial Midline Deffects, Growth Retardation and Radial Defercts, *J. Med. Gen.* **2004** (submitted).

[3] R. Al–Baradie, K. Yamada, C. St. Hilaire, W.–M. Chan, C. Andrews, N. McIntosh, M. Nakano, E. J. Martonyi, W. R. Raymond, S. Okumura, M. M. Okihiro and E. C. Engle, Duane Radial Ray Syndrome (Okihiro Syndrome) Maps to 20q13 and Results from Mutations in SALL4, a New Member of the SAL Family, *Am. J. Hum. Genet.* **2002**, *71*, 1195–1199.

[4] J. Kohlhase, M. Heinrich, L. Schubert, M. Liebers, A. Kispert, F. Laccone, P. Turnpenny, R. M. Winter and W. Reardon, Okihiro Syndrome is Caused by SALL4 Mutations, *Hum. Molec. Genet.* **2002**, *11*, 2979–2987.

[5] J. Kohlhase, L. Schubert, M. Liebers, A. Rauch, K. Becker, S. N. Mohammed, R. Newbury–Ecob and W. Reardon, Mutations at the SALL4 Locus on Chromosome 20 Result in a Range of Clinically Overlapping Phenotypes, Including Okihiro Syndrome, Holt–Oram Syndrome, Acro–Renal–Ocular Syndrome, and Patients Previously Reported to Represent Thalidomide Embryopathy, *J. Med. Genet.* **2003**, *40*, 473–478.

[6] C. Netzer, L. Rieger, A. Brero, C.–D. Zhang, M. Hinzke, J. Kohlhase and S. K. Bohlander, *SALL1*, the Gene Mutated in Townes–Brocks Syndrome, Encodes a Transcriptional Repressor which Interacts with TRF1/PIN2 and Localizes to Pericentromeric Heterochromatin, *Hum. Mol. Genet.* **2001**, *10*, 3017–3024.

[7] http://archive.uwcm.ac.uk/uwcm/mg/search/4216161.html.

[8] http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=607343.
[9] S. McLeskey Kiefer, K. K. Ohlemiller, J. Yang, B. W. McDill, J. Kohlhase and M. Rauchman, Expression of a Truncated Sall1 Transcriptional Repressor is Responsible for Townes–Brocks Syndrome Birth Defects, *Hum. Mol. Genet.* **2003**, *12*, 2221–2227.
[10] X. Huang and W. Miller, A Time–Efficient, Linear–Space Local Similarity Algorithm, *Adv. Appl. Math.***1991**, *12*, 337–357.
[11] B. Morgenstern, O. Rinner, S. Abdeddaïm, D. Haase, K. Mayer, A. Dress and H.–W. Mewes, Exon Discovery by Genomic Sequence Alignment, *Bioinformat.* **2002**, *18*, 777–787.
[12] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein Data Bank, *Nucl. Acids Res.* **2000**, *28*, 235–242.
[13] *Biopolymer* Module of *Insight–II 2000* Molecular Modelling Package and *Discover 2.98* Simulation Package, Accelrys Inc., San Diego, CA.
[14] V. A. Narayan, R. W. Kriwacki and J. P. Caradonna, Structures of Zinc Finger Domains from Transcription Factor Sp1. Insights into Sequence–Specific Protein–DNA Recognition, *J. Biol. Chem.***1997**, *272*, 7801–7808.
[15] J. H. Viles, S. U. Patel, J. B. Mitchell, C. M. Moody, D. E. Justice, J. Uppenbrink, P. M. Doyle, C. J. Harris, P. J. Sadler and J. M. Thornton, Design, Synthesis and Structure of a Zinc Finger with an Artificial Beta–Turn, *J. Mol. Biol.***1998**, *279*, 973–986.
[16] J. C. Miller and C. O. Pabo, Rearrangement of Side–Chains in a Zif268 Mutant Highlights the Complexities of Zinc Finger–DNA Recognition., *J. Mol. Biol.* **2001**, *313*, 309–315.
[17] H. B. Houbaviy, A. Usheva, T. Shenk and S. K. Burley, Cocrystal Structure of YY1 Bound to the Adeno–Associated Virus P5 Initiator, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13577–13582.
[18] J. W. Ponder and F. M. Richards, Tertiary Templates for Proteins. Use of Packing Criteria in the Enumeration of Allowed Sequences for Different Structural Classes, *J. Mol. Biol.* **1987**, *193*, 775–791.
[19] J. R. Maple, M. J. Hwang, T. P. Stockfish, U. Dinur, M. Waldman, C. S. Ewing and A. T. Hagler, Derivation of Class II Force Fields. 1. Methodology and Quantum Force Field for the Alkyl Functional Group and Alkane Molecules, *J. Comput. Chem.* **1994**, *15*, 162–182.
[20] U. Ryde, The Coordination of the Catalytic Zinc in Alcohol Dehydrogenase Studied by Combined Quantum–Chemical and Molecular Mechanics Calculations, *J. Comput. Aided Mol. Des.* **1996**, *10*, 153–164.
[21] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, AM1: A New General Purpose Quantum Mechanical Molecular Model, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[22] P. Andersson, J. Kvassman, A. Lindstroem, B. Olden and G. Pettersson, Effect of NADH on the pKa of Zinc–Bound Water in Liver Alcohol Dehydrogenase, *Eur. J. Biochem.***1981**, *113*, 425–433.

Biographies

Vladimír Frecer is computational chemist and scientific advisor at the International Centre for Science and High Technology, UNIDO, Trieste, Italy. After obtaining a Ph.D. degree in physical chemistry from the Slovak Technical University, Dr. Frecer undertook postdoctoral research at several universities and research institutions, such as Mount Sinai School of Medicine, City University of New York, New York, NY, USA; National Cancer Institute, NIH, Bethesda, MD, USA; University of Trieste, Trieste, Italy; University of Innsbruck, Innsbruck, Austria; National University of Singapore, Singapore and others. More recently, Dr. Frecer has collaborated on numerous projects in the field of structure–based drug design and computer–assisted combinatorial chemistry. Dr. Frecer holds the position of senior scientist at Cancer Research Institute, Slovak Academy of Sciences, Bratislava, Slovakia.

Jan Miertus is a Medical Doctor, being specialized in clinical genetics and genetic testing. He is a post graduate fellow of the University of Trieste at the Pediatric Hospital IRCCS Burlo Garofolo, Trieste, Italy. He published 5 papers on clinical and molecular genetics.

Wiktor Borozdin is a biologist and currently PhD student in the group of **Jürgen Kohlhase**. He obtained his degree from the University of Torun, Poland.

Jürgen Kohlhase is a Medical Doctor, Board certified specialist in Human Genetics and currently head of the Institute of Human Genetics and Anthropology, University of Freiburg, Germany. His main research focus is on the family of *SALL* genes and their involvement in human genetic disease, especially Townes-Brocks and Okihiro syndromes.

Antonio Amoroso did his MD degree in 1979 at the University of Torino and PhD in immunohaematology in 1983 at the University of Ferrara. A. Amoroso is a Full Professor of medical genetics at the University of Trieste and the Director of the Medical Genetics Department at IRCCS “Burlo Garofolo” Children Hospital in Trieste, Italy. Among his professional experiences he was: Associated Professor of Medical Genetics at University of Torino, Italy, Associated and Full Professor of Medical Genetics at University of Trieste, Director of The School of Specialization in Medical Genetics at University of Trieste, Coordinator of the Interdepartmental Center of Molecular Medicine at University of Trieste. He is also the coordinator of numerous international and national research projects. Prof. Amoroso acts as a

member of Editorial Board of the Journal of Nephrology and as a peer reviewer of international journals (Lancet, Arthritis and Rheumatism, Europen Journal of Immunogenetics). His research activities have been focused on immunogenetics, transplantation medicine, medical genetics, genetics of complex diseases. He co-authored approximately 150 papers and is a co-author and co-editor of more than 50 scientific books, books of proceedings and technical compendia.

Stanislav Miertus did his MSc. in Physical Chemistry in 1971 at Slovak Technical University – STU, Bratislava, Slovakia; his PhD. in Physical Chemistry in 1975. He is Deputy Managing Director of the International Centre for Science and High Technology of the United Nations Industrial Development Organization (ICS–UNIDO), in Trieste, Italy, and Area Director for the ICS–UNIDO Area of Pure and Applied Chemistry, covering different field of chemistry and technology, among which Molecular Modeling, Combinatorial Chemistry and Combinatorial Technologies. Among his professional experiences he was: Full Professor of Physical Chemistry and Deputy Dean of the Faculty of Chemistry of STU, Bratislava, Slovakia, project leader at Poly Tech, AREA Science Park, Trieste, Italy; Coordinator of various projects national and international project, Professor at Faculty of Pharmacy, University of Trieste; Visiting scientist at the University of Pisa; Visiting professor at the Mount Sinai School of Medicine, City Univ. of New York, USA; Researcher at the Institut de Chimie Biophysique, Paris, France, and others. His research activities have been focused on computational, physical and analytical chemistry, especially elucidation of structure and properties of bioactive compounds and biomacromolecules, theory of solvent effects, design and development of anticancer and anti AIDS drugs, biosensors. He co–authored approximately 200 papers and 3 patents. He is co–author and co–editor of more than 20 scientific books, books of proceedings and technical compendia.