



Structural and biophysical characterization of transcription factor HNF-1A as a tool to study MODY3 diabetes variants

Received for publication, December 20, 2021, and in revised form, February 15, 2022. Published, Papers in Press, March 4, 2022.
<https://doi.org/10.1016/j.jbc.2022.101803>

Laura Kind^{1,*}, Arne Raasakka¹, Janne Molnes^{2,3}, Ingvild Aukrust^{2,3}, Lise Bjørkhaug⁴, Pål Rasmus Njølstad^{2,5,*}, Petri Kursula^{1,6}, and Thomas Arnesen^{1,7,8,*}

From the ¹Department of Biomedicine, and ²Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway; ³Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway; ⁴Department of Safety, Chemistry, and Biomedical Laboratory Sciences, Western Norway University of Applied Sciences, Bergen, Norway; ⁵Section of Endocrinology and Metabolism, Children and Youth Clinic, Haukeland University Hospital, Bergen, Norway; ⁶Faculty of Biochemistry and Molecular Medicine & Biocenter Oulu, University of Oulu, Oulu, Finland; ⁷Department of Biological Sciences, University of Bergen, Bergen, Norway; ⁸Department of Surgery, Haukeland University Hospital, Bergen, Norway

Edited by Karen Fleming

Hepatocyte nuclear factor 1A (HNF-1A) is a transcription factor expressed in several embryonic and adult tissues, modulating the expression of numerous target genes. Pathogenic variants in the *HNF1A* gene are known to cause maturity-onset diabetes of the young 3 (MODY3 or *HNF1A* MODY), a disease characterized by dominant inheritance, age of onset before 25 to 35 years of age, and pancreatic β -cell dysfunction. A precise diagnosis can alter management of this disease, as insulin can be exchanged with sulfonylurea tablets and genetic counseling differs from polygenic forms of diabetes. Therefore, more knowledge on the mechanisms of HNF-1A function and the level of pathogenicity of the numerous *HNF1A* variants is required for precise diagnostics. Here, we structurally and biophysically characterized an HNF-1A protein containing both the DNA-binding domain and the dimerization domain, and determined the folding and DNA-binding capacity of two established MODY3 HNF-1A variant proteins (P112L, R263C) and one variant of unknown significance (N266S). All three variants showed reduced functionality compared to the WT protein. Furthermore, while the R263C and N266S variants displayed reduced binding to an HNF-1A target promoter, we found the P112L variant was unstable *in vitro* and in cells. Our results support and mechanistically explain disease causality for these investigated variants and present a novel approach for the dissection of structurally unstable and DNA-binding defective variants. This study indicates that structural and biochemical investigation of HNF-1A is a valuable tool in reliable variant classification needed for precision diabetes diagnostics and management.

Maturity-onset diabetes of the young (MODY) is a monogenic, autosomal-dominant form of diabetes mellitus, which is characterized by impaired β -cell function and insulin regulation. The disease is caused by mutations in one of 11 genes

crucial for pancreatic β -cell development and homeostasis (1). MODY3, which contributes to ~40 to 70% of all MODY cases, is associated with pathogenic variants of the hepatocyte nuclear factor 1A (*HNF1A*) gene (2). The encoded transcription factor HNF-1A is expressed both in the embryonic and adult stages of the liver, gall bladder, pancreas, gastro-intestinal tract, kidney, urinary bladder, as well as the adult bone marrow and immune system (3). This wide expression pattern of HNF-1A indicates diverse molecular functions in the human body. Specifically in the endocrine pancreas, HNF-1A regulates the expression of numerous target genes involved in glucose regulation and metabolism, such as those encoding the insulin receptor, HNF-4A, the glucose transporter GLUT2, and glucose-6-phosphatase (4–6). In a genome location experiment using human pancreatic islets, HNF-1A was found to occupy promoter regions of 106 islet genes (4) and HNF-1A deficiency in mice resulted in broad changes of islet gene expression patterns (5). Studies on the molecular structure and function of this important transcription factor is likely to improve our understanding of pancreatic islet biology and underlying mechanisms for the development of MODY3.

HNF-1A is a multi-domain protein (Fig. 1A), which contains an N-terminal dimerization domain (DD), a central DNA-binding domain (DBD), and a C-terminal transactivation domain (TAD) (3). Three isoforms (A, B, C) have been described, which differ in the length of the C-terminal TAD and are differentially expressed in fetal and adult tissues (7). The N-terminal DD is a 32-residue α -helical region, which can fold and dimerize independently from the residual domains (8, 9). The DD forms a heterotetramer with the pterin-4 alpha-carbinolamine dehydratase 1 (PCBD1), which stabilizes the dimeric form of HNF-1A (10, 11). A 50-residue long linker separates the DD from the DBD, which consists of a Pit-Oct-Unc-specific (POU_S) domain and a homeodomain (POU_H) and can bind to the inverted palindromic consensus DNA sequence GTTAATNATTAAC (12). The 200-residue long DBD contains a nuclear-localization signal preceding the POU_H domain, ensuring translocation of HNF-1A into the

* For correspondence: Laura Kind, laura.kind@uib.no; Pål Rasmus Njølstad, Pal.Njolstad@uib.no; Thomas Arnesen, Thomas.arnesen@uib.no.

Structural and biophysical characterization of HNF-1A

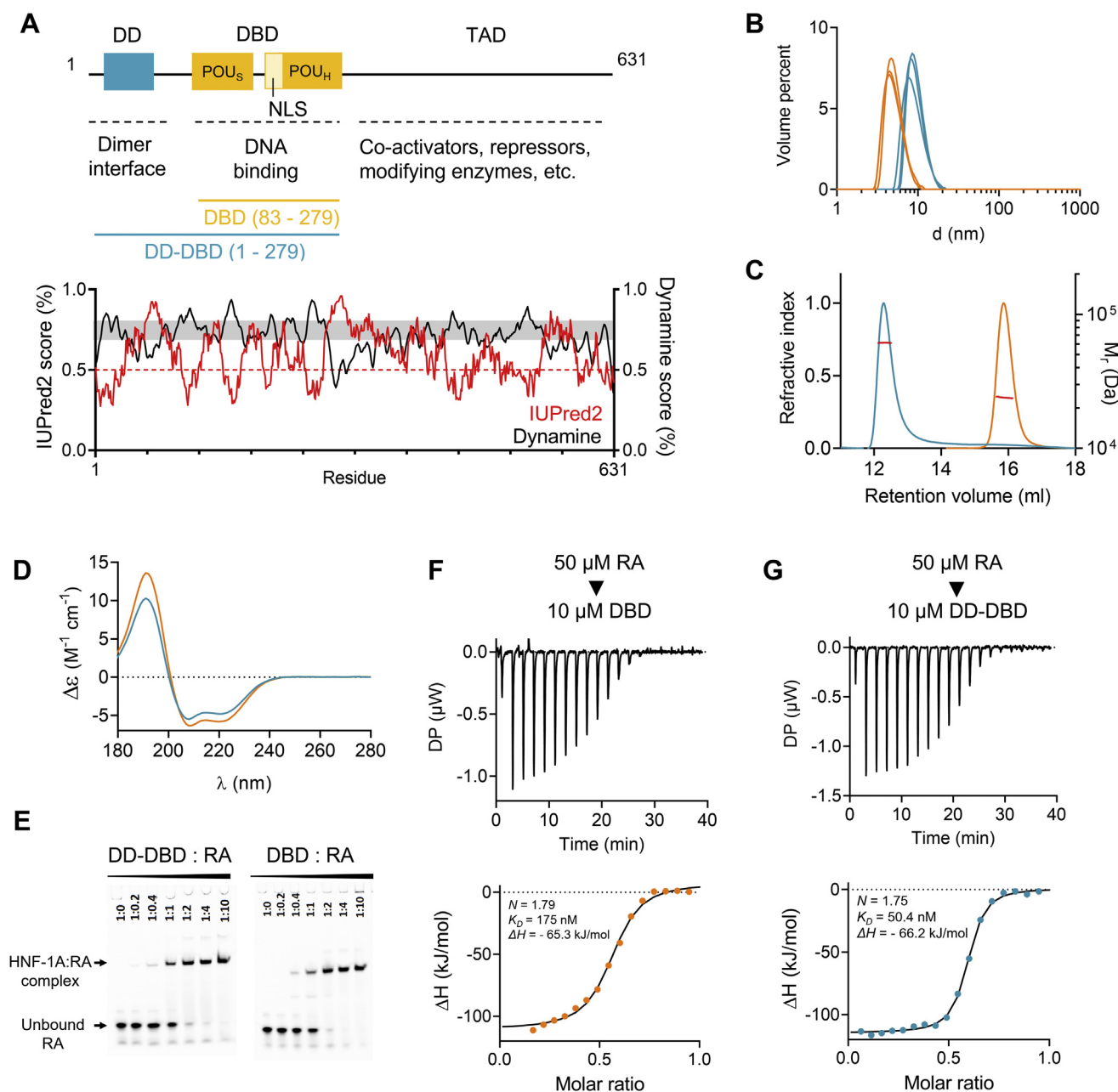


Figure 1. Biophysical and functional characterization of HNF-1A DBD and DD-DBD constructs. *A*, top: domain overview of HNF-1A with residue numbers and indicated domain functions. DD-DBD (residues 1–279) and DBD (residues 83–279) constructs used for *in vitro* protein characterization are shown in *turquoise* and *orange*, respectively. *Bottom*: IUPred2 and DynaMine prediction scores for full-length HNF-1A. IUPred2 scores above 0.5 indicate propensity for disorder, while scores below 0.5 predict ordered regions (54). DynaMine scores below 0.7 indicate disorder, while scores above 0.8 predict order. DynaMine prediction in the *gray zone* (scores between 0.7 and 0.8) indicate residues with highly context-dependent dynamics (55). *B*, DLS of DBD and DD-DBD. *C*, SEC-MALS profiles for DBD and DD-DBD. *Red line*: molecular weight based on MALS and refractive index. *D*, SRCD spectra for DBD and DD-DBD. *E*, EMSA titration native gels showing the complex formation of RA and respective HNF-1A construct. *F* and *G*, ITC titration curves for DBD and DD-DBD, respectively. Titrand: 50 μM double-stranded RA oligonucleotide. Titrand: 10 μM DBD (*F*) or DD-DBD (*G*). DBD, DNA-binding domain; DD, dimerization domain; DLS, dynamic light scattering; HNF-1A, Hepatocyte nuclear factor 1A; SEC-MALS, size-exclusion chromatography–multi-angle light scattering; SRCD, synchrotron radiation circular dichroism.

nucleus (13). The structure of the isolated POU_H domain from rat has been solved by X-ray crystallography and NMR spectroscopy, revealing an atypical homeodomain with an extended loop between the $\alpha 2$ - and $\alpha 3$ -helices (14, 15). Chi *et al.* reported the crystal structure of the entire DBD bound to a high-affinity promoter, corresponding to the consensus

sequence for optimal HNF-1A binding (16). The POU_S and POU_H domains form a stable interface when bound to the DNA, potentially increasing affinity and stability of the complex (16). In addition, HNF-1A contains a 450-residue long TAD in the C-terminus, which is involved in gene transcription activation and likely interacts with numerous

transcription factors, coactivators, and repressors (13, 17, 18). Apart from structural models for the isolated DD and DBD domains, the HNF-1A protein remains structurally and biophysically uncharacterized.

According to the ClinVar database (19), 418 variants of the *HNF1A* gene have been reported, out of which 131 are recognized as pathogenic or likely pathogenic. Remaining variants are classified as likely benign, benign, or variants of uncertain significance (VUS). Diabetes-associated missense mutations occur at highest rates in the DBD, while the mutation rate is intermediate in the DD and low in the TAD (16). In this study, we aimed to structurally and functionally characterize the entire DD-DBD HNF-1A region, harboring most of the variants causing MODY3, and to establish the system as a tool to evaluate the dysfunctional characteristics of MODY3-associated variants *in vitro*. To this end, we compared recombinantly expressed and purified WT protein and two established pathogenic MODY3 variants: p.Pro112Leu and p.Arg263Cys, hereafter referred to as P112L and R263C (20, 21). P112L and R263C, respectively, contribute to approximately 5.7% and 1.4% of all MODY3 probands registered in the Norwegian MODY registry to date. Both variants have previously been characterized regarding cellular functions, but molecular mechanisms remain enigmatic (20, 22, 23). Our biophysical *in vitro* approach revealed defects regarding protein folding and DNA-binding ability, providing explanations for the pathogenicity of these established MODY3 variants. Finally, we used this model system to investigate the *HNF1A* p.Asn266Ser (N266S) variant, which is currently classified as a VUS. N266S is considered as a rare variant, with only one registered proband/family within the Norwegian MODY registry to date. This variant was not described functionally previously, and our biophysical characterizations revealed defects at the protein level, supporting disease causality for the N266S variant.

Results

Biophysical and functional characterization of two HNF-1A constructs

The individual DD and DBD of HNF-1A can be recombinantly expressed and purified for structural studies (8–10, 16). In contrast, the C-terminal TAD has not been studied using recombinant expression systems, which is likely due to the disordered nature typical for TADs (24, 25). Indeed, using disorder predictions, we confirmed that large regions in the TAD have a high tendency for disorder (Fig. 1A). To ensure soluble protein expression, we therefore omitted the TAD in this study and created two HNF-1A constructs: HNF-1A 83 to 279 contained the isolated DBD, while HNF-1A 1 to 279 contained the DD, a linker, and the DBD. These constructs are hereafter referred to as HNF-1A DBD and HNF-1A DD-DBD, respectively (Fig. 1A).

Both protein constructs were purified to homogeneity. DBD and DD-DBD exhibited a slower SDS-PAGE migration than expected based on theoretical molecular weight (Fig. S1); however, the protein identity could be confirmed by peptide

mass fingerprinting. Initial dynamic light scattering (DLS) experiments indicated a hydrodynamic radius of approximately 6.0 nm and 10.7 nm for construct DBD and DD-DBD, respectively (Fig. 1B). These observations were according to expectations, as construct DD-DBD had a higher theoretical mass than DBD and was expected to be dimeric in solution. Size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) confirmed this result (Fig. 1C), as construct DBD exhibited a molecular weight of 24.0 ± 2.4 kDa (monomeric theoretical weight: 22.8 kDa), and construct DD-DBD showed a molecular weight of 61.3 ± 3.7 kDa (monomeric theoretical weight: 31.2 kDa, dimeric theoretical weight: 62.4 kDa).

Synchrotron radiation circular dichroism (SRCD) measurements were conducted to investigate the secondary structure of DBD and DD-DBD, which adopted an α -helical fold. DBD gave a stronger α -helical signal, likely due to the absence of the disordered linker present in the DD-DBD construct (Fig. 1D). These findings were in line with previous reports regarding HNF-1A structure and were important quality-control steps, ensuring that both constructs were suitable for subsequent studies of MODY3 variants.

We qualitatively assessed DNA-binding ability of DBD and DD-DBD by EMSA, in which the purified proteins were added in different ratios to a double-stranded rat albumin promoter oligonucleotide (RA), and the migration behavior of RA was observed during native gel electrophoresis. As the free oligonucleotide probe migrates faster than when in a macromolecular complex, EMSA allows the separation of bound and unbound oligonucleotide and to approximate binding affinities between a protein and the specific promoter oligonucleotide. EMSA showed that both DBD and DD-DBD bound RA, and that the protein:RA complexes differed in size, accordingly to the described size differences of the purified HNF-1A constructs (Fig. 1E). Complex formation started to appear at a molar RA:protein ratio of 1:0.4 and seemed to be nearly complete at a molar ratio of 1:2.

We quantified DNA-binding properties by isothermal titration calorimetry (ITC), from which we extracted thermodynamic parameters of complex formation (Fig. 1, F and G and Table 1). The stoichiometry values from triplicate measurements were determined to 1.97 for construct DD-DBD and 1.89 for construct DBD (Table 1), which corresponds to a RA:protein binding stoichiometry of 1:2. This observation is in line with the crystal structure of an RA:DBD complex (PDB

Table 1
Protein:RA K_D and N values determined by ITC

HNF-1A construct	K_D (μ M)	N (HNF-1A:RA)
DBD	0.11 ± 0.05	1.89 ± 0.30
DBD/R263C	2.00 ± 0.14 ($p = 0.0076$, **)	1.74 ± 0.10
DBD/N266S	0.57 ± 0.05 ($p = 0.0008$, ***)	1.62 ± 0.18
DD-DBD	0.10 ± 0.05	1.97 ± 0.20
DD-DBD/R263C	8.00 ± 3.00 ($p < 0.0001$, ***)	1.73 ± 0.14
DD-DBD/N266S	0.74 ± 0.11 ($p = 0.0004$, ***)	1.55 ± 0.11

Parameters were extracted from binding experiments in which double-stranded RA oligonucleotide was titrated into DBD, DD-DBD, or corresponding HNF-1A variant constructs. p values indicate significant differences of K_D values between studied variants and respective WT constructs. (mean \pm SD; $N = 3$)

Structural and biophysical characterization of HNF-1A

ID: 1IC8; (16)), in which two DBD molecules bind to one double-stranded RA oligonucleotide. Enthalpy changes (ΔH) for both constructs were in the same range (Table 1), suggesting a similar binding mechanism. The affinity between the protein and RA was high, as the average equilibrium dissociation constant K_D was determined to 100 nM for DD-DBD and 114 nM for DBD (Table 1). EMSA and ITC measurements demonstrated that both constructs were able to bind DNA, establishing them as suitable models for *in vitro* characterization of disease-causing variants impacting DNA binding.

HNF-1A DBD and DD-DBD exhibit an extended conformation in solution

Since both DBD and DD-DBD were predicted to contain flexible linker regions (Fig. 1A), we performed small-angle X-ray scattering (SAXS) to obtain information about the conformation and flexibility of the proteins (Fig. 2). The scattering curves of DBD and DD-DBD gave a first indication that the proteins differ in molecular shape (Fig. 2A). The maximal molecular dimension (D_{max}) could be extracted from

the distance distribution function (Fig. 2B): 10.2 nm for DBD and 24.4 nm for DD-DBD. These results agree with our DLS and SEC-MALS data (Fig. 1, B and C). Kratky analysis of the SAXS data showed that both constructs adopt an extended, nonglobular conformation, which can be explained by the linkers between the POU_S and POU_H domain, as well as an unfolded region between the DD and DBD (Fig. 2C). Indeed, a CRY SOL analysis, comparing the DNA-bound state of DBD (PDB ID: 1IC8) and the free state in solution, revealed a difference in compactness and overall shape (Fig. 2A). An *ab initio* model was generated for DBD and superposed with the DBD crystal structure from the DBD:RA complex (16), which illustrates the difference in molecular conformation (Fig. 2D). It appears that the stable interface between the POU_S and POU_H domains reported by Chi *et al.* only forms upon DNA binding. A rigid body model with built linker regions also suggested that the POU_S and POU_H domains do not form intimate contacts in the unbound state (Fig. 2E). In order to probe for changes in secondary structure content upon DNA binding, we performed SRCD experiments for both DBD alone

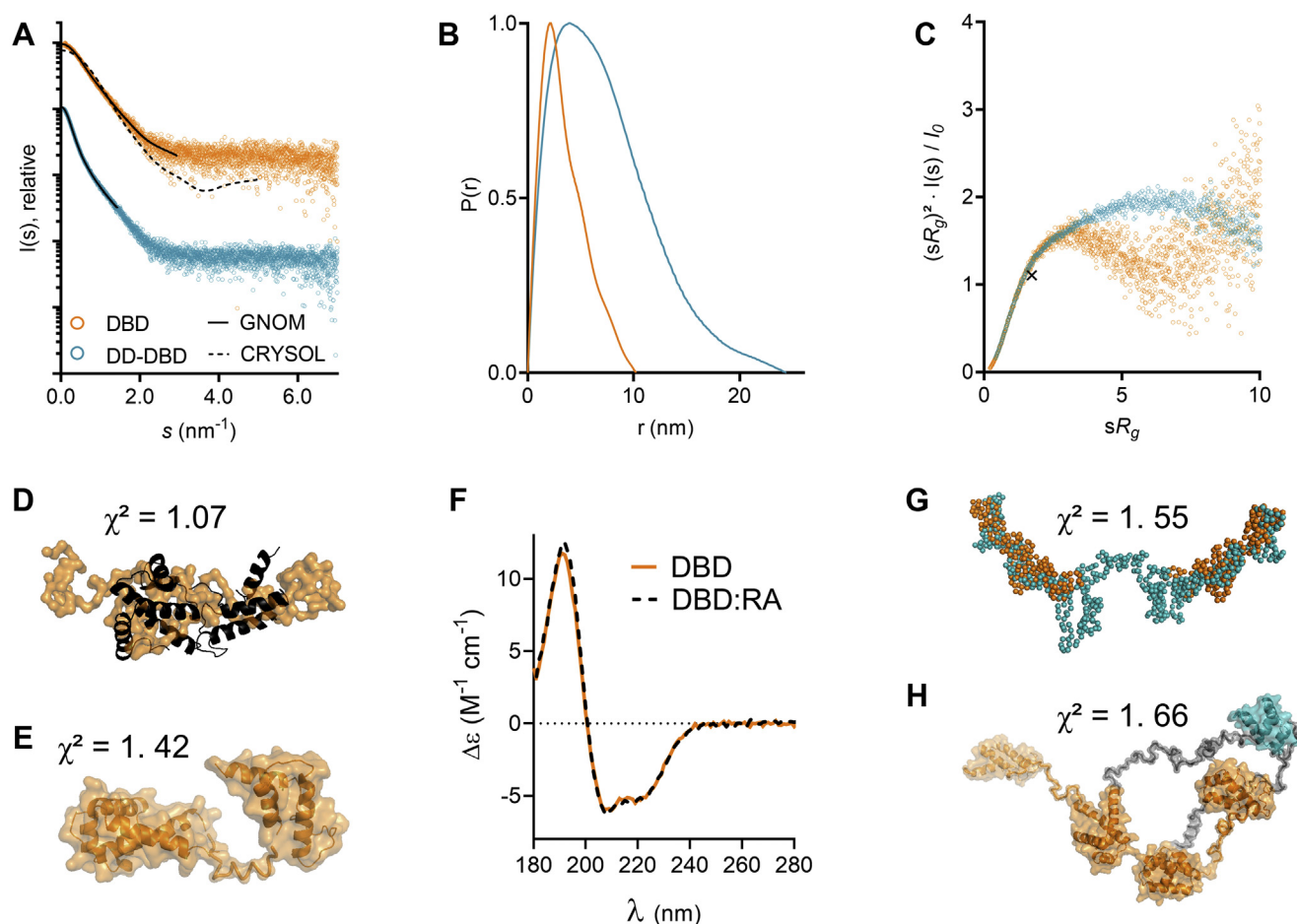


Figure 2. SAXS data for DBD and DD-DBD. A, scattering curves for DBD and DD-DBD. Solid line: GNOM fit. Dashed line: CRY SOL fit with crystal structure model from PDB 1IC8 (16) ($\chi^2 = 10.4$). B, distance distribution functions for DBD and DD-DBD. C, dimensionless Kratky plot for DBD and DD-DBD, with the cross indicating the expected maximum for a globular particle ($\sqrt{3}, 1.104$) (61). D, GASBOR model of DBD (orange), superposed with the oligo-bound model from DBD:RA crystal structure (16). E, CORAL model for DBD. F, SRCD spectra for DBD in the presence and absence of RA. G, GASBOR model for DD-DBD (turquoise), overlaid with GASBOR model of DBD (orange) at estimated positions. H, CORAL model of DD-DBD. DBD residues are shown in orange, linker residues in gray, and DD residues in turquoise. DBD, DNA-binding domain; DD, dimerization domain; RA, rat albumin promoter oligonucleotide; SAXS, small-angle X-ray scattering; SRCD, synchrotron radiation circular dichroism.

and a DBD:RA complex (Fig. 2F). We concluded from the highly similar spectra that there is no notable change in secondary structure content during complex formation, but that the 18-residue linker between the POU_S and POU_H domains becomes more compact upon DNA binding. An *ab initio* model (Fig. 2G) and a rigid body model (Fig. 2H) were generated for construct DD-DBD and revealed a similar behavior. The protein adopts an elongated, extended shape in solution.

HNF-1A P112L shows severely reduced solubility in vitro and decreased stability in cells

Following the biophysical and biochemical characterization of the two HNF-1A WT constructs, we aimed to utilize this system to study diabetes-causing variants of HNF-1A. We chose to explore two established pathogenic variants (P112L and R263C) and one VUS (N266S) to elucidate disease mechanisms of MODY3. All three mutation sites are located in the DBD of HNF-1A (Fig. 3, A–D). The alignment of HNF-1A sequences from model organisms (residues 1–279) showed that the three affected residues are highly conserved across species down to zebrafish and suggested low tolerance for sequence variance in regard to protein function (Fig. 3A). Pro112 is in the POU_S domain at the beginning of helix α 2 (Fig. 3, B and C). Considering the helix-terminating nature of proline residues, this position suggests that Pro112 might have an important role in the

secondary structure formation of POU_S. Both Arg263 and Asn266 are located in the POU_H domain (Fig. 3B). Both residues are found in the C-terminal helix α 9, which is inserted into the major groove of the DNA (Fig. 3D). Arg263 partakes in ionic interactions with the backbone phosphate group of the adenine in the 16th position of the RA oligonucleotide (o.A16). In addition, Arg263 forms polar contacts with the backbone carbonyl group of Leu258 and possibly with the side chain of Thr260. Asn266 is situated in proximity of the DNA molecule and may be involved in DNA recognition. The side chain of Asn266 does not partake in direct contacts with other residues, but the polar nature and chain length of asparagine may be important for energetically favorable protein folding.

Initial expression tests in *Escherichia coli* Rosetta(DE3) distinguished the P112L variant from the R263C and N266S variants. While the latter were equally soluble as the WT constructs, the P112L variant showed a severe reduction in protein solubility (Figs. 3E and S1, A and B). Different lysis buffers were tested, varying pH, NaCl concentration, and amounts of stabilizing additives, but no solubilizing condition was found (data not shown). An additional screen in *E. coli* Lemo21(DE3) was conducted, in which protein expression levels can be tuned by titration of L-rhamnose into the expression media, leading to slower and more controlled protein expression. However, the P112L variant appeared insoluble, independent of the titrant amount (Fig. S1C). This observed insolubility of P112L indicated

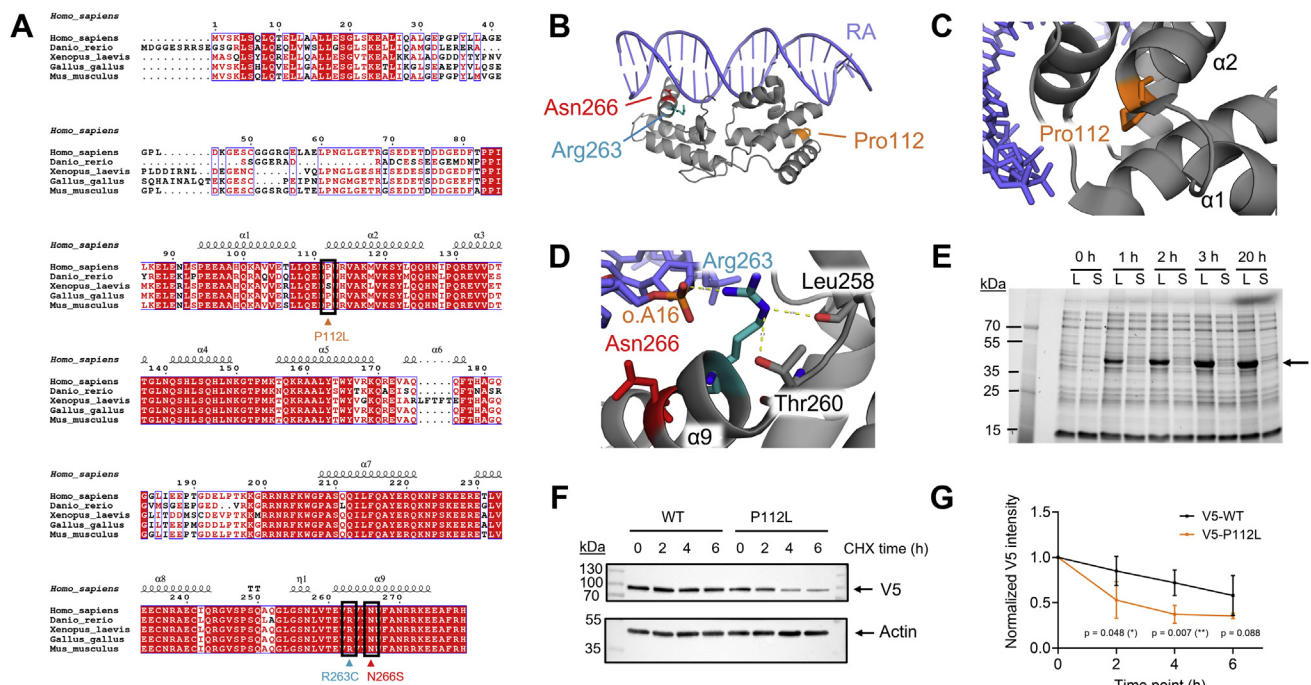


Figure 3. Overview of the studied variants and protein turnover dynamics of HNF-1A P112L. A, multiple sequence alignment of HNF-1A sequences from model organisms, with variants of interest indicated by *black boxes* and *arrows*. B, crystal structure of RA-bound DBD (PDB ID: 1IC8) with mutation sites highlighted. C, magnified view on Pro112 and arrangement of neighboring residues. D, magnified view on Arg263 and Asn266. o.A16 and neighboring residues Leu258 and Thr260 are indicated. E, expression and solubility test for DD-DBD P112L (31.2 kDa) in *Escherichia coli* Rosetta(DE3). Lysate (L) and soluble fraction (S) at different time points of expression (1–3 h: 37 °C, 20 h: 20 °C, 0 h: uninduced control). *Arrow* indicates over-expressed protein band with an apparent slow migration behavior, as observed for DD-DBD WT. F, representative Western blot membranes of CHX chase assay analysis of full-length WT and P112L protein turnover. Western blot membranes from all four conducted replicates are shown in Fig. S2. G, quantification of CHX chase assay for HNF-1A WT and P112L, based on normalization to whole protein amount from stain-free SDS-PAGE gel (N = 4, Fig. S2). CHX, cycloheximide; DBD, DNA-binding domain; DD, dimerization domain; HNF-1A, hepatocyte nuclear factor 1A; RA, rat albumin promoter oligonucleotide.

Structural and biophysical characterization of HNF-1A

misfolding of the protein, leading to aggregation or incorporation into bacterial inclusion bodies.

To test the hypothesis of protein instability further, we performed protein turnover experiments in HeLa cells, employing a cycloheximide (CHX) chase assay (Figs. 3, F and G and S2). The drug CHX inhibits protein synthesis in the cells, allowing the assessment of protein degradation dynamics. Using a specific anti-V5 antibody, the amount of full-length V5-WT and V5-P112L protein was detected at different time points of CHX treatment (Fig. 3F). Based on the V5-signal at time point 0 h, it was apparent that the P112L variant is expressed at lower levels than the WT protein at steady state conditions (Fig. 3F). Alternatively, the weaker V5-signals for P112L could arise from lower solubility of the P112L variant as compared to WT HNF-1A. The time-course of normalized protein levels indicated that the protein turnover of the P112L variant was elevated compared to the WT protein (Fig. 3G).

HNF-1A R263C and N266S display normal protein folding and partially altered thermal stability

The established pathogenic variant R263C and the VUS N266S were expressed and purified like the WT constructs. We performed DLS experiments to investigate hydrodynamic

diameters of the proteins and found that the variant constructs were similar to the respective WT constructs (Fig. 4A). SEC-MALS measurements confirmed that DBD/R263C was monomeric in solution ($M_r = 23.9 \pm 2.4$ kDa), while DD-DBD/R263C formed a dimer in solution ($M_r = 60.2 \pm 3.4$ kDa) (Fig. 4B and Table S1). The corresponding N266S constructs were found to be in the same oligomeric state (DD-DBD/N266S: $M_r = 61.9 \pm 3.4$ kDa, DBD/N266S: $M_r = 24.1 \pm 2.1$ kDa). All variant elution profiles overlapped entirely with the respective WT SEC profile (Fig. 4B).

SRCD spectra of the DBD and DD-DBD variants overlaid with the corresponding WT spectra across the entire wavelength range, indicating identical secondary structure (Fig. 4C). We conducted SAXS experiments to investigate the solution structure of the studied HNF-1A variants (Fig. 4, D–F and Table S3). The R263C mutation did not introduce changes in molecular dimension, molecular shape, or flexibility. In contrast, the N266S variant appeared slightly less compact, as a minor increase in D_{max} (DBD/WT: 10.0 nm, DBD/N266S: 11.5 nm) became apparent from the distance distribution function (Fig. 4E and Table S3).

Thermal stability was assessed by differential scanning fluorimetry (DSF), from which the melting temperature (T_M)

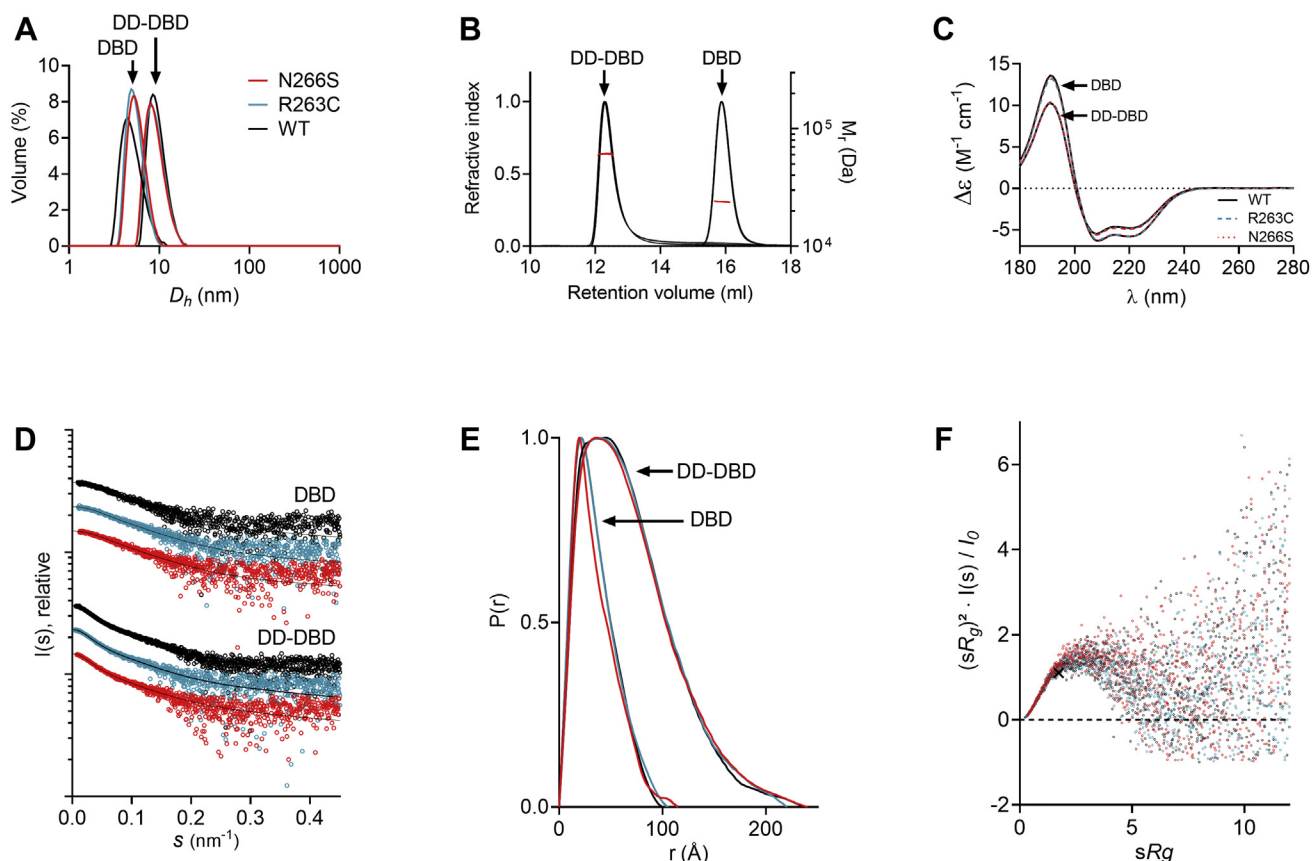


Figure 4. Biophysical characterization of HNF-1A R263C and N266S. A, DLS curves for DBD and DD-DBD variants. Red line: molecular weight based on MALS and refractive index. C, SRCD spectra for DBD and DD-DBD variants. D, SAXS scattering curves for DBD and DD-DBD variants. E, SAXS distance distribution functions for DBD and DD-DBD variants. F, dimensionless Kratky plot for DBD variants, with the cross indicating the expected maximum for a globular particle ($\sqrt{3}$, 1.104) (61). DBD, DNA-binding domain; DD, dimerization domain; DLS, dynamic light scattering; HNF-1A, hepatocyte nuclear factor 1A; SAXS, small-angle X-ray scattering; SEC-MALS, size-exclusion chromatography–multi-angle light scattering; SRCD, synchrotron radiation circular dichroism.

of each variant was extracted (Table 1 and Fig. S4). For the WT constructs, the T_M values were generally very low (DBD: 42.2 °C, DD-DBD: 35.8 °C), which likely results from the presence of the small, separated POU_H and POU_S domains, as well as the long, disordered linker in DD-DBD. The DSF protein stability assay also revealed that the (DD-)DBD/R263C variants were as stable as the respective WT constructs. In contrast, the N266S mutation was found to destabilize the protein during thermal denaturation, as the DBD/N266S variant exhibited a T_M value that was 4 °C lower than the T_M of the DBD WT construct (Table 1). The thermal instability of DBD/N266S may be connected to the less compact fold of the same construct, as observed in SAXS.

We compared the experimentally determined stability differences with computational predictions by using the CUPSAT prediction algorithm (Table 2). CUPSAT was used to assess possible changes in free energy of protein folding ($\Delta\Delta G$) upon point mutation, considering amino acid atom potentials and torsion angle distribution (26). All assessed variants (P112L, R263C, N266S) were predicted to introduce unfavorable torsion angles and to destabilize the protein structure ($\Delta\Delta G < 0$), which likely contributes to the observed defects.

HNF-1A R263C and N266S exhibit strongly reduced DNA-binding ability

While the overall folding of R263C and N266S constructs appeared normal, the DNA-binding ability was strongly affected (Fig. 5). We performed a dose-dependent EMSA to qualitatively measure DNA-binding affinity (Fig. 5, A and B). A faint complex band was observed at a 1:0.2 RA:protein molar ratio for WT DD-DBD, while a first complex band for DD-DBD/R263C was only visible at a molar RA:DD-DBD/R263C ratio of 1:0.4 (Fig. 5A). Full RA:DD-DBD binding was achieved at a ratio of 1:4, while the R263C variant did not reach complete binding even at a molar ratio of 1:10. The binding behavior for the R263C variant appeared shifted to the right side of the EMSA titration (Fig. 5A). The EMSA titration for DD-DBD/N266S also revealed a weaker protein DNA complex formation, although not as severe as for DD-DBD/R263C (Fig. 5, A and B). A similar behavior was observed in an EMSA titration for the corresponding DBD constructs (Fig. S3). ITC measurements allowed us to quantitatively analyze DNA-binding thermodynamics and to precisely compare WT and variant constructs (Fig. 5, C–E). Representative titration curves for DD-DBD/R263C and DD-DBD/N266S are shown in Figure 5, C and D. An average K_D value of 8.0 μM was obtained for DD-DBD/R263C, which is significantly higher than the average K_D value of 0.10 μM for the

DD-DBD WT construct (Fig. 1H and Table 1). DBD/R263C exhibited slightly stronger binding than the dimeric construct, with an average K_D value of 2.0 μM (Table 1). Similar to the WT proteins, the binding event for the R263C variants was an enthalpy-driven process (Fig. 5E). For the N266S variant, ITC experiments revealed average K_D values of 0.57 μM (DBD/N266S) and 0.74 μM (DD-DBD/N266S), which again pointed toward an impaired protein–DNA interaction compared to WT proteins (Fig. 5, D and E and Table 1). In addition, the stoichiometry dropped to an average number of approximately 1.6 (Table 1), which suggested a decrease in active protein concentration in the ITC cell. Indeed, a manual inspection of the sample after the ITC experiment confirmed that white precipitation had formed. As the ITC experiment was performed at 30 °C, this aggregation could be explained by the low thermal stability observed in DSF (Table 2).

Discussion

Considering the diverse expression patterns and major role of HNF-1A for a functional endocrine pancreas, remarkably little is known about the biochemical structure of HNF-1A and protein function on a molecular level. Notably, pathogenic HNF-1A variants cause the most common form of inherited diabetes, MODY3. Even though the genetic cause of MODY3 is known, there is limited molecular understanding of how a dysfunctional HNF-1A protein leads to decreased insulin secretion. We therefore aimed to extend the knowledge on the structural and biophysical characteristics of HNF-1A to study MODY3 variants in greater detail. In this work, we henceforth investigated two truncated versions of the HNF-1A protein and three MODY3-associated HNF-1A variants.

Our finding that DD-DBD can dimerize independently of the dimerization cofactors PCBD1 and PCBD2 was expected, as DD had been shown to form homodimers in isolation (8, 9). Complex formation with PCBD1/2 may stabilize this dimeric form, as demonstrated by Bayle *et al.* (11), and thereby modulate HNF-1A function. Even though the DNA-binding affinity was not significantly increased for the dimeric construct, the dimerization of HNF-1A may affect the protein's lifetime by protecting it from denaturation and proteolysis (27) or could participate in transcriptional activity as dimerization may increase the local concentration of the regulatory TAD (28).

Moreover, we found that our HNF-1A constructs exhibit a high degree of flexibility. The 30-residue POU_S–POU_H linker tethers the two homeodomains together and allows for distance variation between them. This intrinsic conformational flexibility is a general feature of the POU transcription factor

Table 2
Thermal protein stability predictions by CUPSAT and experimental T_M values determined by DSF analysis (mean \pm SD; N = 4)

Variant	$\Delta\Delta G_{\text{CUPSAT}}$	Torsion _{CUPSAT}	T_M DD-DBD (DSF)	T_M DBD (DSF)
WT	–	–	35.8 \pm 0.3 °C	42.2 \pm 0.1 °C
P112L	– 1.77 kcal/mol	Unfavorable	n.d.	n.d.
R263C	– 2.5 kcal/mol	Unfavorable	36.6 \pm 0.2 °C	41.3 \pm 0.1 °C
N266S	– 0.45 kcal/mol	Unfavorable	35.1 \pm 0.3 °C	38.4 \pm 0.1 °C

Abbreviations: n.d., not determined.

Structural and biophysical characterization of HNF-1A

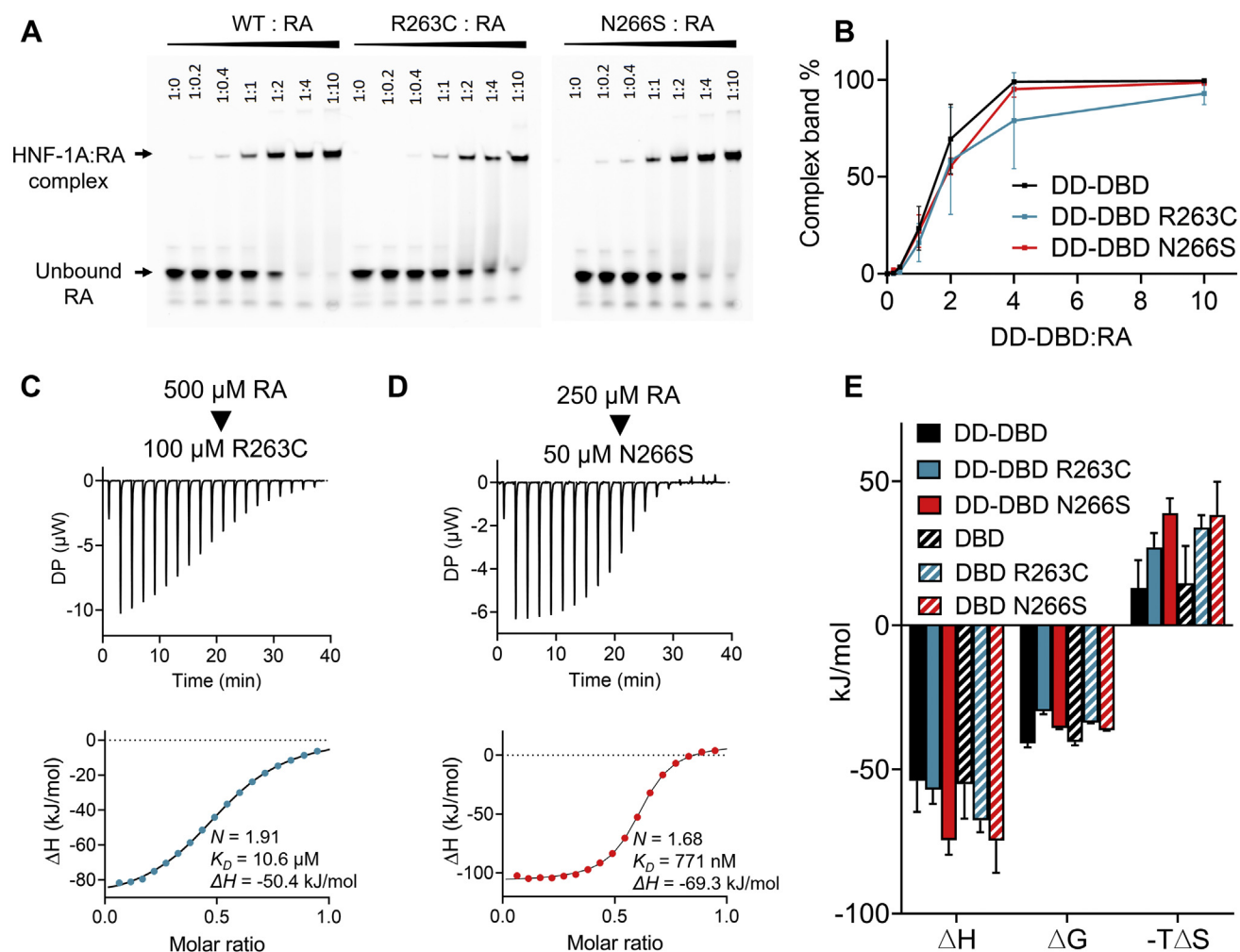


Figure 5. DNA-binding parameters for HNF-1A R263C and N266S. A, EMSA gel for DD-DBD WT, R263C, and N266S variants. B, quantification of EMSA gels for DD-DBD variants ($N = 3$). C and D, representative ITC titration curve for DD-DBD/R263C (C) and DD-DBD/N266S (D). Titrant: 500 μ M (C) or 250 μ M (D) double-stranded RA. Titrand: 100 μ M DD-DBD/R263C (C) or 50 μ M DD-DBD/N266S (D). E, quantification of thermodynamic energies from ITC analysis (mean \pm SD; $N = 3$). DBD, DNA-binding domain; DD, dimerization domain; ITC, isothermal titration calorimetry.

family, enabling the proteins to adopt diverse quaternary structures during DNA recognition (29–31). Two well-characterized exemplary POU transcription factors are Oct-1 and Pit-1, which are both able to bind to DNA motifs in different arrangements, depending on the nature of the DNA response element (32). Oct-1 binds to both the *Palindromic Oct-factor response element* (PORE) as well as to *More of PORE* (MORE), albeit in different conformations (33) (Fig. 6). In the PORE-binding mode, the POU_S and POU_H domains of one Oct-1 protomer are arranged on opposite sides of the DNA molecule, with the linker being extended across the double helix axis. In contrast, the POU_S and POU_H domains are in a compact formation on the same side of the DNA molecule when Oct-1 is bound to a MORE DNA motif (33). The DNA recognition mode of Oct-1 influences its ability to interact with the transcriptional enhancer OBF-1, which is only able to bind and activate Oct-1 transcriptional activity in the PORE-bound state (32, 34, 35). Thus, the conformational flexibility of the Oct-1 transcription factor is crucial for variable arrangements and gene activation modes. A similar

conformational variability was shown for the POU transcription factor Pit-1 (36), where different DNA protein arrangements alter the accessibility for transcription modulators, such as the corepressor N-CoR (32). These examples of protein–DNA recognition modes illustrate the importance of the flexible linker and allow speculations about HNF-1A not only recognizing the established promoters in a MORE-like fashion but perhaps also binding to undescribed response elements with distinct HNF-1A promoter recognition modes.

The P112L variant has previously been functionally studied in cells (20, 22). While the nuclear localization appears normal, HNF-1A P112L exhibits strongly impaired DNA-binding ability *in vitro* and transactivation potential on the RA promoter in HeLa cells (20, 22). Surprisingly, the P112L constructs were insoluble in our expression conditions. This insolubility could arise from structural instability, which can be explained by the substitution of Pro112 at the critical position in the beginning of helix α 2 (Fig. 3, B and C). Proline residues play important roles in modulating hydrogen bond formation in α -helix backbone structures (37–39), and the P112L

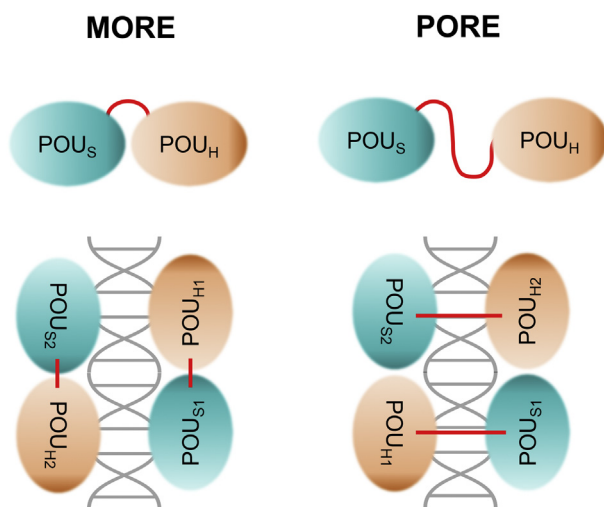


Figure 6. MORE and PORE DNA-binding modes of POU transcription factors (29–33). Transcription factors in MORE arrangement are bound with both POU_S and POU_H domains on the same side of the DNA molecule. The homeodomains are in close proximity and the linker is compact. In PORE-binding mode, the POU_S and POU_H domains are located on opposite sides of the DNA, with the linker being extended across the DNA symmetry axis. MORE, More of PORE; PORE, palindromic Oct-factor response element; POU_S, Pit-Oct-Unc-specific domain; POU_H, Pit-Oct-Unc homeodomain.

substitution may lead to an N-terminally elongated $\alpha 2$ helix with impact on overall protein folding. Moreover, the rigidity and bulkiness of Pro112 may also contribute to protein folding by providing an optimal local backbone conformation.

The conducted protein turnover experiments in HeLa cells support our hypothesis of enhanced protein degradation dynamics. One way to further delineate the mechanism is the application of specific drugs that inhibit lysosomal and proteasomal degradation pathways. Our findings agree with a recent large-scale multidimensional analysis of MODY3 variants, where HNF-1A P112L expression in HeLa cells was lower than WT HNF-1A (40). Our data supports an alternative cause of disease for HNF-1A P112L, with P112L being a structurally destabilizing mutation, rather than a pure DNA-binding mutation as suggested in earlier reports (20, 22).

HNF-1A R263C proved to be a loss-of-function mutation regarding DNA binding. Our results are in accordance with earlier studies showing that *in vitro* translated HNF-1A R263C is unable to bind DNA and that this HNF-1A variant has a significantly decreased transactivation potential in HeLa cells (22, 23). This effect can be explained by the positioning of Arg263 in HNF-1A (Fig. 3D). Arg263 undergoes direct contacts with o.A16 of the promoter motif and appears to form ion-dipole interactions with Leu258 and Thr260. Interestingly, a genetic variance in the latter (T260M) was identified in a MODY3 patient. It compromises DNA binding and causes impaired β -cell function (41). Based on these data, the Arg263-Thr260-Leu258 microenvironment and the direct interaction between Arg263 and o.A16 appear to be crucial for adequate HNF-1A DNA interaction (Fig. 3D).

Finally, we utilized this system to investigate an HNF-1A VUS (N266S), for which previously unknown mechanisms

were uncovered. The mutant exhibits altered thermal stability, minor changes in folding, and a decrease in protein:DNA affinity. A combination of all effects may lead to impaired HNF-1A function in the pancreatic β -cell. Bioinformatic analyses of the protein region harboring the Asn266 residue may assist in explaining possible molecular mechanisms (Fig. S5). By performing a local alignment search for HNF-1A, we identified a homologous protein, homeobox-containing protein 1 (HMBOX1) (42), which shares 28% sequence identity and exhibits a conserved homeobox domain (Fig. S5A). A structure alignment between the HNF-1A DBD crystal structure (16) and the crystal structure of the HMBOX1 homeodomain (42) showed high structural similarity between the compared domains (Fig. S5B). Especially the amino acid sequence and structural arrangement of helix $\alpha 9$ is highly conserved between the two proteins (Fig. S5, A and B), indicating that residues in this region are critical for HNF-1A and HMBOX1 function. Asn266 in HNF-1A may form polar interactions important for the stabilization of the α -helical backbone structure of $\alpha 9$ or might indirectly bind to the DNA. Hence, the N266S substitution may compromise helix stability or DNA binding. Alternatively, the substitution of Asn266 with Ser may potentially introduce a phosphorylation site, which could destabilize protein structure and hinder DNA binding. Interestingly, HMBOX1 is strongly expressed in pancreatic tissue, where it may act as a transcriptional repressor (43). The conservation of the homeodomain and the $\alpha 9$ helix in HNF-1A and HMBOX1 allow speculations that both proteins are able to bind to the same target promoter sequence; HNF-1A as a transcriptional activator and HMBOX1 as a transcriptional repressor. Due to compromised $\alpha 9$ integrity and DNA binding, the N266S variant may be outcompeted by HMBOX1, leading to an altered regulation of gene transcription in β -cells. Further studies in human cell lines will likely complete our understanding of disease causality, possibly leading to a reclassification of HNF-1A N266S from VUS to (likely) pathogenic.

In summary, we established a biophysical characterization platform for HNF-1A, which proved to be a valuable tool to investigate MODY3 mutations in the DD and DBD of HNF-1A. We verified pathogenic effects, as well as provided further understanding for structural and functional dysregulation of the variants. In particular, primary effects caused by structurally destabilizing mutations may be uncovered in a direct manner, providing explanations for secondary effects in gene transactivation activity.

Conclusions

Structural and biochemical studies is an underrepresented approach in the variant interpretation process of transcription factor-associated diseases, such as MODY. Our presented methodology provides new insights into WT HNF-1A protein function and demonstrates that structure-based investigations may lead to increased certainty in variant-phenotype correlation and can play a valuable part in MODY precision diagnostics and treatment.

Structural and biophysical characterization of HNF-1A

Experimental procedures

Plasmid generation

The human *HNF1A* gene in a pcDNA3.1/HisC plasmid (22) was used in order to subclone two HNF-1A constructs (DD-DBD: residues 1–279, DBD: residues 83–279) into the Gateway donor vector pDONR221 (Invitrogen). An N-terminal Tobacco-Etch Virus protease site (ENLYFQG) was included. In addition, attB1 and attB2 sites were introduced N- and C-terminally of the respective gene fragments, allowing for Gateway cloning strategy. Final expression clones in the destination vector pTH27 (44) harbored an N-terminal His₆-tag (His₆-(DD-)DBD). MODY3-associated variants (P112L, R263C, N266S) were introduced into (DD-)DBD pTH27 expression plasmids by Q5 site-directed mutagenesis (New England Biolabs). These constructs were used for recombinant protein expression and purification in order to biophysically and functionally characterize HNF-1A variants.

Using the same strategy, the full-length *HNF1A* gene sequence (residues 1–631) was subcloned into the pcDNA3.1/nV5-DEST mammalian expression vector (Invitrogen), harboring an N-terminal V5-tag. The P112L variant was generated by Q5 site-directed mutagenesis (New England Biolabs). These constructs were used to assess protein stability in a CHX assay.

Primers used in Gateway cloning, Q5 site-directed mutagenesis, and plasmid sequencing are listed in Table S4.

Recombinant protein expression and purification

His₆-(DD-)DBD constructs were expressed in the *E. coli* Rosetta (DE3). Bacteria were grown in LB medium at 37 °C until an A₆₀₀ value of 0.6 to 0.8 was reached. Protein expression, induced by the addition of 1 mM of IPTG, was performed at 20 °C for 20 h.

Bacterial cell pellets were resuspended in 50 mM Hepes (pH 7.5), 500 mM NaCl, 1 mM DTT, 20 mM imidazole, 1x cOmplete EDTA-free protease inhibitor cocktail (Roche), 1 mM PMSE, and lysed by ultrasonication (7 min, 1 s on/off cycles, 25 W). Cell lysates were cleared by centrifugation (16,000g, 30 min, 4 °C) and 6xHis-(DD-)DBD was purified using Ni-NTA affinity chromatography. Elution was done using an imidazole gradient (40–300 mM), and protein-containing fractions were dialyzed overnight at 4 °C (20 mM Hepes pH 7.5, 500 mM NaCl, 1 mM DTT). The N-terminal His₆-tags were proteolytically removed by the addition of Tobacco-Etch Virus protease during the dialysis step. A second Ni-NTA affinity chromatography step was used to separate cleaved from uncleaved proteins, where a gradient in a low-imidazole range was applied (0 mM – 80 mM). Relevant fractions were concentrated and subjected to SEC, using a Superdex 75 pg 16/60 column (GE Healthcare). The running buffer contained 20 mM Hepes pH 7.5, 500 mM NaCl, and 1 mM Tris(2-carboxyethyl)phosphine (TCEP). Fractions containing pure (DD-)DBD were concentrated, snap-frozen in liquid N₂, and stored at –80 °C.

The expression and solubility of mutant constructs was assessed in *E. coli* Rosetta(DE3) and *E. coli* Lemo21(DE3)

strains, varying expression temperatures, and lysis buffer composition. HNF-1 A variants R263C and N266S were recombinantly expressed and purified in the same way as WT proteins. The HNF-1A P112L variant was not expressed and purified due to insolubility in the tested conditions.

Protein identity was confirmed by peptide mass fingerprinting using a Bruker Ultra fleXtreme MALDI-TOF mass analyzer.

Dynamic light scattering

Monodispersity and R_h of (DD-)DBD variants was assessed using DLS. Measurements of 1.5 mg/ml purified variants (20 mM Hepes pH 7.5, 500 mM NaCl, 1 mM TCEP) were performed using a Malvern Zetasizer Nano-ZS instrument at 25 °C. Each variant was measured in triplicate.

Size exclusion chromatography–multi-angle light scattering

Molecular weights of (DD-)DBD and respective variants were determined using SEC-MALS. Chromatography was performed at room temperature on a Superdex 200 Increase 10/300 G1 (GE Healthcare) column in SEC buffer (20 mM Hepes pH 7.5, 500 mM NaCl, 1 mM TCEP). Hundred microgram sample was subjected to gel electrophoresis at a flow rate of 0.5 ml/min, and MALS was recorded using a miniDAWN TREOS instrument. Concentration determination was based on refractive index, recorded with an on-line ERC RefractoMax 520 refractometer. The monomeric third fibronectin III domain of rat β4-integrin (MW = 12.76 kDa) was used as internal molecular weight standard (45). The data were analyzed using the ASTRA software (Wyatt).

Differential scanning fluorimetry

DSF measurements were performed using a Lightcycler 480 II (Roche) thermocycler. 0.25 mg/ml DD-DBD or 0.5 mg/ml DBD variants in 2 mM Hepes pH 7.5, 50 mM NaCl, 0.1 mM TCEP, 15x SyproOrange (Invitrogen) were subjected to thermal denaturation (20–95 °C, 0.2 °C data pitch). SyproOrange fluorescence at 465 to 610 nm was followed. All measurements were performed as technical triplicate. The fluorescence signal was normalized and plotted against temperature. Graphpad Prism 8.3.0 (GraphPad Software) was used to fit the data to a sigmoidal function and to extract T_M values from the inflection point of the curve.

Small-angle X-ray scattering

Initial SAXS experiments on WT proteins were performed on the P12 synchrotron beamline (46), EMBL/DESY, as mail-in service. Batch measurements were done in 20 mM Hepes (pH 8.0), 500 mM NaCl, 1 mM TCEP at 10 °C. A dilution series (0.5 mg/ml–2.0 mg/ml) was prepared in order to investigate potential concentration-dependent effects. Bovine serum albumin was used as a standard protein for molecular weight determination.

SAXS measurements of all variants were performed as mail-in service on the SWING beamline (47) at the SOLEIL synchrotron. All measurements were performed in SEC-SAXS mode using an Agilent Biosec3-300 column at 14 °C (20 mM Hepes pH 7.5, 500 mM NaCl, 2 mM TCEP). Fifty microliters of protein sample at 4 to 7 mg/ml was injected, and X-ray scattering data were collected as the sample eluted from the column.

SAXS data were reduced and processed using the ATSAS 3.0.3 package (48). CRY SOL (49) was employed to generate a theoretical scattering curve of the DNA-bound DBD from the published DBD:DNA crystal structure (PDB: 1IC8) (16), which was subsequently compared to the experimental scattering curve of the DNA-free DBD in solution. *Ab initio* models for (DD-)DBD were generated by using GASBOR (50). Rigid body models of multi-domain proteins were generated by using CORAL (51). Here, the isolated POU_S and POU_H domains (PDB: 1IC8) and the dimeric DD (PDB: 1F93, (10)) were used as rigid bodies, and linker regions were modeled as flexible loops. Further modeling details are listed in Tables S2 and S3. Visualization of SAXS models was done in PyMOL (52).

Electrophoretic mobility shift assay

EMSA was used to qualitatively assess the DNA-binding ability of HNF-1A variants. Cy5-labeled oligonucleotide probes, corresponding to the RA promoter as found in PDB 1IC8 (16), were commercially synthesized (Merck KGaA) and dissolved in 10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA to 100 μM final concentration. 5' oligo ([Cy5]-CTTGGTTAA-TAATTCACCAGA) and 3' oligo ([Cy5]-TCTGGTGAAT-TATTAACCAAG) were mixed in an equimolar ratio. The probes were annealed into double-stranded DNA by incubation at 95 °C for 5 min, before being allowed to cool down to room temperature over 45 min.

EMSA was performed using the LI-COR EMSA buffer kit (LI-COR Biosciences). The EMSA reaction mix contained 1x Binding buffer (10 mM Tris pH 7.5, 50 mM KCl, 1 mM DTT), 0.5 μM [Cy5]-RA, 2 mM DTT, and varying amounts of protein sample. The DNA probe was mixed in different ratios with purified (DD-)DBD variants, sampling DNA:protein molar ratios from 1:0 to 1:10. EMSA was performed according to protocol instructions and Cy5 signal was detected from the EMSA gel using a ChemiDoc MP imaging system.

Isothermal titration calorimetry

ITC measurements were performed using a MicroCal iTC200 instrument (Malvern Panalytical). Unlabeled oligonucleotides (TAG Copenhagen) were annealed into double-stranded RA as described above. (DD-)DBD variants and the DNA were dialyzed against 20 mM Hepes pH 7.5, 200 mM NaCl, 1 mM TCEP. Typically, 50 μM RA was titrated into 10 μM (DD-)DBD protein. Higher concentrations were required for the R263C and N266S variants, where 500 μM or

250 μM RA, respectively, was titrated into 50 μM protein. The injection program contained 19 injections with 2 μl injection volume, with a two-minute spacing interval. All measurements were performed at 30 °C in triplicate from the same protein batch. Titrations were analyzed using the MicroCal PEAQ-ITC analysis software (Malvern Panalytical). ΔH , ΔG , $T\Delta S$, K_D , and N were extracted from the fits. Statistical analysis was applied to compare K_D values of different variants. p values were calculated by performing unpaired t-tests for each individual variant-WT pair.

Synchrotron radiation circular dichroism

SRCD measurements were performed on the AU-CD beamline, ASTRID2 synchrotron. Purified proteins and double-stranded RA were dialyzed into CD buffer (20 mM Naphosphate buffer pH 8.0, 150 mM NaF, 1 mM TCEP). Initial measurements were conducted at 25 °C for WT constructs and HNF-1A:RA complex formation studies. SRCD measurements for the comparison of WT HNF-1A and MODY3 variants were performed at 10 °C. 0.1-mm circular cells (Hellma Analytics) were used. Initial SRCD spectra on WT constructs and HNF-1A:RA complexes were recorded in the wavelength range of 350 to 170 nm, while spectra for WT-variant comparisons were recorded in a wavelength range of 280 to 170 nm. Wavelength scans were performed in 1 nm increments. Three scans per sample were performed and averaged. Generally, protein constructs were measured at a concentration of 1 mg/ml. For measurements of protein:DNA complexes, double-stranded RA oligonucleotide and (DD-)DBD constructs were mixed in a 1:2 M ratio. Data processing was done using CDtoolX (53) or using Microsoft Excel.

CHX chase assay

HeLa cells were grown in DMEM growth medium (Sigma-Aldrich), supplemented with 15% fetal bovine serum, 1% penicillin-streptomycin, and 4 mM L-glutamine, in a humidity incubator keeping constant conditions of 37 °C and 5% atmospheric CO₂. HeLa cells were seeded in 6-well plates at a density of 200,000 cells per well. 4 × 4 wells were prepared for each variant (WT/P112L HNF-1A). Transfection with full-length nV5-WT or nV5-P112L (0.8 μg DNA/well) was performed 24 h post-seeding using X-tremeGENE 9 DNA Transfection Reagent (Roche). Twenty-four hours post-transfection, the growth medium was exchanged with growth medium supplemented with 50 μg/ml CHX (Merck KGaA). Sample acquisition was done at different time points of CHX treatment (0 h, 2 h, 4 h, 6 h). HeLa cells were washed twice with ice-cold PBS and harvested in 1 ml ice-cold PBS using a cell scraper. Samples were centrifuged (16,000g, 4 °C, 2 × 15 s), and the dry cell pellets were stored at -80 °C. Cells were resuspended in 40 μl lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 1x cComplete EDTA-free protease inhibitor cocktail) and incubated on ice for 30 min. Cell lysates were cleared by centrifugation (16,000g, 10 min, 4 °C), and concentrations of the soluble fraction were

Structural and biophysical characterization of HNF-1A

determined by using the Pierce bicinchoninic acid protein assay (Thermo Fisher Scientific). Twenty microgram protein samples were analyzed by SDS-PAGE and Western blotting. Following protein transfer, the membrane was blocked in 5% milk-PBS-Tween (milk-PBS-T). Primary antibodies anti-V5 (Invitrogen, R960–25) and anti-pan-actin (Abcam, ab200658) were applied at 1:2000 in 1% milk-PBS-T and incubated overnight at 4 °C. Membranes were washed three times with PBS-T. Secondary antibody was applied at a 1:5000 dilution in 3% milk-PBS-T and incubated for 2 h at 4 °C. Membranes were washed three times in PBS-T before horseradish peroxidase signals were developed with SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific). Detection of horseradish peroxidase signal was conducted using a ChemiDoc XRS+ imaging system. Band intensities were quantified using the ImageLab v3.0 software (Bio-Rad Laboratories). V5-bands, representing protein levels of V5-WT and V5-P112L variants, were normalized to total protein amount based on stain-free gel imaging prior to Western blot transfer (Fig. S2). Signals for V5-WT and V5-P112L at different time points of CHX treatment were further normalized to the respective V5-signal at 0 h treatment. Antiactin bands were used to visualize equal protein loading. Statistical analysis was done in GraphPad Prism version 8.3.0., using an unpaired *t* test per time-point without assuming a consistent SD. Statistical significance was determined using the Holm-Sidak method, with a significance threshold of 0.05.

Bioinformatic analysis

Protein disorder prediction was done using the IUPred2 prediction server (54). Prediction score curves produced by IUPred2A and DynaMine (55) algorithms were plotted in Graphpad Prism 8.3.0.

Sequence conservation of the HNF-1A DBD was investigated by the generation of a global alignment. HNF-1A protein sequences from several model organisms were extracted from the UniProt (56) database (*Homo sapiens*: P20823, *Mus musculus*: P22361, *Danio rerio*: Q4L1M9, *Xenopus laevis*: Q05041, *Gallus gallus*: Q90867). Sequences were aligned using Clustal Omega (57) and a graphical presentation was generated using ESPript 3.0 (58).

Thermal stability of HNF-1A variants was predicted by using the CUPSAT algorithm (26), considering torsion angle distributions and amino acid atom potentials. As convention, a negative $\Delta\Delta G$ corresponds to a destabilizing point mutation.

A local alignment search for HNF-1A was conducted using BLAST (59). A pair-wise alignment of human HNF-1A (UniProt ID: P20823) and human HMBOX1 (UniProt ID: Q6NT76) was generated using Clustal Omega (57) and a graphical presentation was generated using ESPript 3.0 (58). A structure alignment of the DBD of HNF-1A (PDB: 1IC8, chain A, (16)) and the homeobox domain of HMBOX1 (PDB: 4J19, chain A, (42)) was generated with the RCSB PDB pair-wise structure alignment tool using the jFATCAT rigid algorithm

(60). Protein structures were analyzed and visualized using PyMOL (52).

Data availability

All obtained data are contained within the manuscript and the supplemental data.

Supporting information—This article contains supporting information.

Acknowledgments—We acknowledge Rezan Erman and Marcus Langeland Larsen Nygård for technical assistance. We are thankful for guidance with SEC-MALS measurements by Dr Anne Baumann. ITC, DSF, DLS, and SEC-MALS experiments were performed at the Biophysics, Structural Biology and Screening (BiSS) facility at the University of Bergen. We thank Ulrich Bergmann (Biocenter Oulu Proteomics Core Facility) for conducting mass spectrometry analysis. We are grateful for beamtime and supportive synchrotron beamline staff at PETRAIII EMBL P12, SOLEIL Synchrotron SWING, and ASTRID2 ISA AU-CD beamlines.

Author contributions—L. K., A. R., P. R. N., P. K., and T. A. conceptualization; L. K. and A. R. methodology; L. K., A. R., P. K., and T. A. validation; L. K. and A. R. formal analysis; L. K. and A. R. investigation; L. K. and A. R. data curation; L. K. writing—original draft; L. K., A. R., J. M., I. A., L. B., P. R. N., P. K., and T. A. writing—review and editing; L. K. visualization; L. K., P. R. N., P. K., and T. A. funding acquisition; J. M., I. A., L. B., P. R. N., P. K., and T. A. resources; P. R. N. and T. A. project administration; P. K. and T. A. supervision.

Funding and additional information—This work was funded with a PhD fellowship by the Medical Faculty, University of Bergen, UiB, Norway (to L. K.) and by a UiB Meltzer foundation project grant (to L. K.). L. K. received a research visit grant from the Norwegian Graduate School in Biocatalysis. T. A. was supported by the Norwegian Cancer Society (Project 171752—PR-2009-0222). P. R. N. received funding from the European Research Council (AdG SELECTionPREDISPOSED #293574), the Research Council of Norway (FRIPRO grant #240413), the Western Norway Regional Health Authority (Strategic Fund “Personalized Medicine for Children and Adults”), the Novo Nordisk Foundation (grant #54741), and the Norwegian Diabetes Association. This work has been supported by the project CALIPSOplus under the Grant Agreement 730872 from the EU Framework Programme for Research and Innovation HORIZON 2020 (to P. K.).

Conflicts of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CHX, cycloheximide; DBD, DNA-binding domain; DD, dimerization domain; DLS, dynamic light scattering; DSF, differential scanning fluorimetry; HMBOX1, homeobox-containing protein 1; HNF-1A, hepatocyte nuclear factor 1A; ITC, isothermal titration calorimetry; milk-PBS-T, milk-PBS-Tween; MODY, maturity-onset diabetes of the young; MORE, More of PORE; PCBD, pterin-4 alpha-carbinolamine dehydratase; PORE, palindromic Oct-factor response element; POU_H, Pit-Oct-Unc homeodomain; POU_S, Pit-Oct-Unc-specific domain; RA, rat albumin promoter oligonucleotide; SAXS,

small-angle X-ray scattering; SEC-MALS, size-exclusion chromatography–multi-angle light scattering; SRCD, synchrotron radiation circular dichroism; TAD, transactivation domain; VUS, variants of uncertain significance.

References

- Zhang, H., Colclough, K., Gloyn, A. L., and Pollin, T. I. (2021) Monogenic diabetes: A gateway to precision medicine in diabetes. *J. Clin. Invest.* **131**, e142244
- Yahaya, T. O., and Ufuoma, S. B. (2020) Genetics and pathophysiology of maturity-onset diabetes of the young (MODY): A Review of current trends. *Oman Med. J.* **35**, e126
- Lau, H. H., Ng, N. H. J., Loo, L. S. W., Jasmen, J. B., and Teo, A. K. K. (2018) The molecular functions of hepatocyte nuclear factors - in and beyond the liver. *J. Hepatol.* **68**, 1033–1048
- Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., Fraenkel, E., Bell, G. I., and Young, R. A. (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381
- Servitja, J. M., Pignatelli, M., Maestro, M. A., Cardalda, C., Boj, S. F., Lozano, J., Blanco, E., Lafuente, A., McCarthy, M. I., Sumoy, L., Guigo, R., and Ferrer, J. (2009) Hnf1alpha (MODY3) controls tissue-specific transcriptional programs and exerts opposed effects on cell growth in pancreatic islets and liver. *Mol. Cell Biol.* **29**, 2945–2959
- Wang, H., Hagenfeldt-Johansson, K., Otten, L. A., Gauthier, B. R., Herrera, P. L., and Wollheim, C. B. (2002) Experimental models of transcription factor-associated maturity-onset diabetes of the young. *Diabetes* **51**, S333–S342
- Bach, I., and Yaniv, M. (1993) More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing. *EMBO J.* **12**, 4229–4242
- De Francesco, R., Pastore, A., Vecchio, G., and Cortese, R. (1991) Circular dichroism study on the conformational stability of the dimerization domain of transcription factor LFB1. *Biochemistry* **30**, 143–147
- Narayana, N., Phillips, N. B., Hua, Q. X., Jia, W., and Weiss, M. A. (2006) Diabetes mellitus due to misfolding of a beta-cell transcription factor: Stereospecific frustration of a schellman motif in HNF-1alpha. *J. Mol. Biol.* **362**, 414–429
- Rose, R. B., Bayle, J. H., Endrizzi, J. A., Cronk, J. D., Crabtree, G. R., and Alber, T. (2000) Structural basis of dimerization, coactivator recognition and MODY3 mutations in HNF-1alpha. *Nat. Struct. Biol.* **7**, 744–748
- Bayle, J. H., Randazzo, F., Johnen, G., Kaufman, S., Nagy, A., Rossant, J., and Crabtree, G. R. (2002) Hyperphenylalaninemia and impaired glucose tolerance in mice lacking the bifunctional DCoH gene. *J. Biol. Chem.* **277**, 28884–28891
- Courtois, G., Baumhueter, S., and Crabtree, G. R. (1988) Purified hepatocyte nuclear factor 1 interacts with a family of hepatocyte-specific promoters. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7937–7941
- Bjorkhaug, L., Bratland, A., Njolstad, P. R., and Molven, A. (2005) Functional dissection of the HNF-1alpha transcription factor: A study on nuclear localization and transcriptional activation. *DNA Cell Biol.* **24**, 661–669
- Ceska, T. A., Lamers, M., Monaci, P., Nicosia, A., Cortese, R., and Suck, D. (1993) The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding. *EMBO J.* **12**, 1805–1810
- Schott, O., Billeter, M., Leiting, B., Wider, G., and Wuthrich, K. (1997) The NMR solution structure of the non-classical homeodomain from the rat liver LFB1/HNF1 transcription factor. *J. Mol. Biol.* **267**, 673–683
- Chi, Y. I., Frantz, J. D., Oh, B. C., Hansen, L., Dhe-Paganon, S., and Shoelson, S. E. (2002) Diabetes mutations delineate an atypical POU domain in HNF-1alpha. *Mol. Cell* **10**, 1129–1137
- Rosenfeld, M. G. (1991) POU-domain transcription factors: Pou-er-ful developmental regulators. *Genes Dev.* **5**, 897–907
- Yu, M., Wang, J., Li, W., Yuan, Y. Z., Li, C. Y., Qian, X. H., Xu, W. X., Zhan, Y. Q., and Yang, X. M. (2008) Proteomic screen defines the hepatocyte nuclear factor 1alpha-binding partners and identifies HMGB1 as a new cofactor of HNF1alpha. *Nucleic Acids Res.* **36**, 1209–1219
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipati, Z., Malheiro, A., et al. (2018) ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067
- Bjorkhaug, L., Ye, H., Horikawa, Y., Sovik, O., Molven, A., and Njolstad, P. R. (2000) MODY associated with two novel hepatocyte nuclear factor-1alpha loss-of-function mutations (P112L and Q466X). *Biochem. Biophys. Res. Commun.* **279**, 792–798
- Iwasaki, N., Oda, N., Ogata, M., Hara, M., Hinokio, Y., Oda, Y., Yamagata, K., Kanematsu, S., Ohgawara, H., Omori, Y., and Bell, G. I. (1997) Mutations in the hepatocyte nuclear factor-1alpha/MODY3 gene in Japanese subjects with early- and late-onset NIDDM. *Diabetes* **46**, 1504–1508
- Bjorkhaug, L., Sagen, J. V., Thorsby, P., Sovik, O., Molven, A., and Njolstad, P. R. (2003) Hepatocyte nuclear factor-1 alpha gene mutations and diabetes in Norway. *J. Clin. Endocrinol. Metab.* **88**, 920–931
- Yang, Q., Yamagata, K., Yamamoto, K., Miyagawa, J., Takeda, J., Iwasaki, N., Iwahashi, H., Yoshiuchi, I., Namba, M., Miyazaki, J., Hanafusa, T., and Matsuzawa, Y. (1999) Structure/function studies of hepatocyte nuclear factor-1alpha, a diabetes-associated transcription factor. *Biochem. Biophys. Res. Commun.* **266**, 196–202
- Staby, L., O’Shea, C., Willemoes, M., Theisen, F., Kragelund, B. B., and Skriver, K. (2017) Eukaryotic transcription factors: Paradigms of protein intrinsic disorder. *Biochem. J.* **474**, 2509–2532
- Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888
- Parthiban, V., Gromiha, M. M., and Schomburg, D. (2006) Cupsat: Prediction of protein stability upon point mutations. *Nucleic Acids Res.* **34**, W239–W242
- Hashimoto, K., and Panchenko, A. R. (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20352–20357
- Frietze, S., and Farnham, P. J. (2011) Transcription factor effector domains. *Subcell Biochem.* **52**, 261–277
- Phillips, K., and Luisi, B. (2000) The virtuoso of versatility: POU proteins that flex to fit. *J. Mol. Biol.* **302**, 1023–1039
- Malik, V., Zimmer, D., and Jauch, R. (2018) Diversity among POU transcription factors in chromatin recognition and cell fate reprogramming. *Cell Mol. Life Sci.* **75**, 1587–1612
- Ryan, A. K., and Rosenfeld, M. G. (1997) POU domain family values: Flexibility, partnerships, and developmental codes. *Genes Dev.* **11**, 1207–1225
- Remenyi, A., Tomilin, A., Scholer, H. R., and Wilmanns, M. (2002) Differential activity by DNA-induced quarternary structures of POU transcription factors. *Biochem. Pharmacol.* **64**, 979–984
- Remenyi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Scholer, H. R., and Wilmanns, M. (2001) Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Mol. Cell* **8**, 569–580
- Chasman, D., Cepek, K., Sharp, P. A., and Pabo, C. O. (1999) Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: Specific recognition of a protein-DNA interface. *Genes Dev.* **13**, 2650–2657
- Sauter, P., and Matthias, P. (1998) Coactivator OBF-1 makes selective contacts with both the POU-specific domain and the POU homeodomain and acts as a molecular clamp on DNA. *Mol. Cell Biol.* **18**, 7397–7409
- Scully, K. M., Jacobson, E. M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D. W., Hooshmand, F., Aggarwal, A. K., and Rosenfeld, M. G. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* **290**, 1127–1131
- Richardson, J. S., and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science* **240**, 1648–1652

Structural and biophysical characterization of HNF-1A

38. Williams, K. A., and Deber, C. M. (1991) Proline residues in transmembrane helices: Structural or dynamic role? *Biochemistry* **30**, 8919–8923
39. Hurley, J. H., Mason, D. A., and Matthews, B. W. (1992) Flexible-geometry conformational energy maps for the amino acid residue preceding a proline. *Biopolymers* **32**, 1443–1446
40. Althari, S., Najmi, L. A., Bennett, A. J., Aukrust, I., Rundle, J. K., Colclough, K., Molnes, J., Kaci, A., Nawaz, S., van der Lugt, T., Hassanali, N., Mahajan, A., Molven, A., Ellard, S., McCarthy, M. I., *et al.* (2020) Unsupervised clustering of missense variants in HNF1A using multidimensional functional data aids clinical interpretation. *Am. J. Hum. Genet.* **107**, 670–682
41. Haliyur, R., Tong, X., Sanyoura, M., Shrestha, S., Lindner, J., Saunders, D. C., Aramandla, R., Poffenberger, G., Redick, S. D., Bottino, R., Prasad, N., Levy, S. E., Blind, R. D., Harlan, D. M., Philipson, L. H., *et al.* (2019) Human islets expressing HNF1A variant have defective beta cell transcriptional regulatory networks. *J. Clin. Invest.* **129**, 246–251
42. Kappei, D., Butter, F., Benda, C., Scheibe, M., Draskovic, I., Stevense, M., Novo, C. L., Basquin, C., Araki, M., Araki, K., Krastev, D. B., Kittler, R., Jessberger, R., Londono-Vallejo, J. A., Mann, M., *et al.* (2013) HOT1 is a mammalian direct telomere repeat-binding protein contributing to telomerase recruitment. *EMBO J.* **32**, 1681–1701
43. Chen, S., Saiyin, H., Zeng, X., Xi, J., Liu, X., Li, X., and Yu, L. (2006) Isolation and functional analysis of human HMBOX1, a homeobox containing protein with transcriptional repressor activity. *Cytogenet. Genome Res.* **114**, 131–136
44. Hammarstrom, M., Woestenenk, E. A., Hellgren, N., Hard, T., and Berglund, H. (2006) Effect of N-terminal solubility enhancing fusion proteins on yield of purified target protein. *J. Struct. Funct. Genomics* **7**, 1–14
45. Raasakka, A., Linxweiler, H., Brophy, P. J., Sherman, D. L., and Kursula, P. (2019) Direct binding of the flexible C-terminal segment of periaxin to beta4 integrin suggests a molecular basis for CMT4F. *Front. Mol. Neurosci.* **12**, 84
46. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, J., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M., and Svergun, D. I. (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443
47. Thureau, A., R. P., and Pérez, J. (2021) BioSAXS on the SWING beamline at synchrotron SOLEIL. *J. Appl. Crystallogr.* **54**, 1698–1710
48. Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., and Franke, D. (2021) Atsas 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355
49. Svergun, D. I. B., C., and Koch, M. H. J. (1995) Crysol - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773
50. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953
51. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350
52. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
53. Miles, A. J., and Wallace, B. A. (2018) CDtoolX, a downloadable software package for processing and analyses of circular dichroism spectroscopic data. *Protein Sci.* **27**, 1717–1722
54. Erdos, G., and Dosztanyi, Z. (2020) Analyzing protein disorder with IUPred2A. *Curr. Protoc. Bioinformatics* **70**, e99
55. Cilia, E., Pancsa, R., Tompa, P., Lenaerts, T., and Vranken, W. F. (2014) The DynaMine webserver: Predicting protein dynamics from sequence. *Nucleic Acids Res.* **42**, W264–W270
56. UniProt, C. (2021) UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489
57. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641
58. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324
59. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
60. Prlic, A., Bliven, S., Rose, P. W., Bluhm, W. F., Bizon, C., Godzik, A., and Bourne, P. E. (2010) Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* **26**, 2983–2985
61. Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53