



LUND UNIVERSITY
Faculty of Medicine

LUP

Lund University Publications

Institutional Repository of Lund University

This is an author produced version of a paper published in *Journal of Molecular Biology*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:
Madhumati Sevana, Kristin Kassler,
Josefin Ahnström, Sigrid Weiler, Björn Dahlbäck,
Heinrich Sticht, Yves A. Muller

"Mouse ApoM Displays an Unprecedented
Seven-Stranded Lipocalin Fold: Folding Decoy or
Alternative Native Fold?"

Journal of Molecular Biology
2010 404(3), 363 - 371

<http://dx.doi.org/10.1016/j.jmb.2010.09.062>

Access to the published version may require journal
subscription.

Published with permission from: Academic Press Ltd-
Elsevier Science Ltd

Mouse apoM displays an unprecedented 7-stranded lipocalin fold - Folding decoy or alternative native fold?

Madhumati Sevvana¹, Kristin Kessler², Ahnström Josefin³, Sigrid Weiler¹, Björn Dahlbäck³, Heinrich Sticht², & Yves A. Muller^{1*}

¹ Lehrstuhl für Biotechnik, Department Biologie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

² Institut für Biochemie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

³ Department of Laboratory Medicine, Clinical Chemistry, University of Lund, University Hospital, Malmö, Sweden

*Corresponding author:

Yves A. Muller

Lehrstuhl für Biotechnik, Department of Biology,
Friedrich-Alexander University Erlangen-Nuremberg

Im IZMP, Henkestr. 91, D-91052 Erlangen

Phone: +49-(0)9131-8523081

Fax: +49-(0)9131-8523080

E-mail: ymuller@biologie.uni-erlangen.de

Running title: Unprecedented 7-stranded lipocalin fold

Summary

Mouse apoM (m-apoM) displays 79 % sequence identity to human apoM (h-apoM). Both proteins are HDL-associated apolipoproteins with similar anticipated biological functions. The structure of h-apoM has recently been determined by X-ray crystallography and revealed that h-apoM displays, as expected, a lipocalin-like fold characterized by an 8-stranded β -barrel that encloses an internal fatty acid-binding site. Surprisingly, this is not true for m-apoM. After refolding from inclusion bodies, the crystal structure of m-apoM reported here at 2.5 Å resolution displays a novel yet unprecedented 7-stranded β -barrel structure. The fold difference is not caused by a mere deletion of a single β -strand; instead, β -strands E and F are removed and replaced by a single β -strand A' formed from residues from the N-terminus. Molecular dynamics simulations suggest that m-apoM is able to adopt both a 7-stranded and an 8-stranded barrel structure in solution and that both folds are comparably stable. Thermal unfolding simulations identify the position where the β -strand exchange occurs as the weak point of the β -barrel. We wonder whether the switch in topology could have a biological function and facilitate ligand release since it goes in hand with a narrowing of the barrel diameter. Possibly also, the observed conformation represents an on-pathway or off-pathway folding intermediate of apoM. The difference in fold topology is quite remarkable, and the fold promiscuity observed for m-apoM might possibly provide a glimpse at potential crosspoints during the evolution of β -barrels.

Key words: Misfolding, refolding, alternative conformations, apolipoprotein, lipocalin

Abbreviations used: apoM, apolipoprotein M; m-apoM, mouse apoM; h-apoM, human apoM; MD, molecular dynamics; RMSD, root mean square deviation; SCR, structurally conserved regions

Introduction

The breadth of the conformational space that can be sampled by a protein has important consequences for protein function. Anfinsen demonstrated that protein folding is energy driven and that the primary sequence of a polypeptide chain contains all the information required for the correct folding of a protein.^{1,2} As a consequence of the thermodynamic hypothesis, the energy landscape underlying the conformational space is best approximated by a funnel model, and the biological active native structure corresponds to the conformation with the lowest free energy content.^{1,3} The bottom of the energy well can be wide and flat, allowing for some conformational diversity that is accessible through thermal motion and that enables protein flexibility as a key element in protein function.

Conformational flexibility is however not considered to include major rearrangements in the secondary structure topology of a protein. More extended differences in the fold of a protein are expected to be separated by sizable energy gaps, and interconversion of these folds requires overcoming considerable energy barriers. An outstanding example for this are SERPINS, where proteolytic cleavage is a prerequisite for the conversion of the central 5-stranded β -sheet A into a 6-stranded β -sheet in a process referred to as a spring-loaded mechanism and which is associated with considerable changes in the thermodynamic stability of the protein.^{4,5} Very recently the assumption that only a narrow conformational distribution exists around a given average native conformation has been questioned in a single molecule study on an RNA enzyme.⁶ Solomatin *et al.* observed that the studied ribozyme continuously switches between severely distinct conformational states within timescales that exceed the enzyme turnover time. The authors raise the question whether similar extended structural interconversions also exists in proteins.

For structural and functional studies, many proteins are heterologously produced in *E. coli* with the caveat that they often accumulate in inclusion bodies and require renaturation before they can be characterized. Current statistics from the "REFOLD" database list about 759 proteins purified by various refolding protocols.⁷ According to the thermodynamic hypothesis, a successful refolding protocol will yield a homogenous protein sample with all molecules displaying a highly similar low energy conformation. The observation that such a sample can be crystallized is often taken as an additional indicator of its homogeneity. Conversely, unsuccessful refolding protocols yield heterogeneous samples, which are prone to aggregation and precipitation and have low chances for being successfully crystallized.

Here, we report the crystal structure of mouse apoM (m-apoM) that was purified upon refolding from inclusion bodies and that displays a highly unusual and yet unobserved 7-stranded lipocalin fold. ApoM is a 25 kDa plasma ligand transport protein that is associated with high density lipoprotein (HDL) particles.^{8,9} Differing from m-apoM, the recently solved crystal structure of human apoM (h-apoM) revealed a common lipocalin fold.¹⁰ Central to the lipocalin fold is an 8-stranded β -barrel that encloses an internal ligand-binding pocket.¹¹ The amino acid sequence of h-apoM and m-apoM are 79 % identical.^{12,13} Both h-apoM and m-apoM are expressed in hepatocytes and in kidney proximal tubule cells and their biological functions are expected to be highly similar.¹³ The high sequence identity and the expected functional similarities strongly suggest that both proteins adopt similar native folds. Yet, the crystal structure of m-apoM reveals a topologically quite distinct 7-stranded β -barrel whereas a common 8-stranded lipocalin fold is observed in h-apoM.

Crystal structure reveals an unusual lipocalin fold

M-apoM (amino acids 22-190) missing the N-terminal hydrophobic anchor helix¹⁰ was purified from inclusion bodies and solubilised with reducing and denaturing solutions prior to the refolding. After successful crystallisation, its three-dimensional structure was determined by X-ray crystallography to a resolution of 2.5 Å (Table 1, Fig. 1, and Supplementary Fig. S1 and Fig. S2). M-apoM has been predicted to display a lipocalin-type fold, but after solving the structure of m-apoM with molecular replacement using h-apoM as a search model,¹⁰ the initial map immediately indicated that there were several large rearrangements within the calyx like β -barrel structure (Supplementary Fig. S1). The final model of m-apoM does not include residues 22 to 23, 107 to 123 and residue 190 since these could not be located in the electron density maps (Fig. 1). Furthermore, none of the residues from the attached N-terminal S-tag are visible in the crystal structure. Superposition of m-apoM with the h-apoM search model yields an RMSD as large as 1.7 Å despite a sequence identity of 79 % (Fig. 2, Supplementary Fig. S3).

Although the overall fold of m-apoM resembles that of the human homologue, a closer inspection, however, reveals a major difference. Whereas h-apoM and all lipocalins solved to date display an 8-stranded β -barrel, m-apoM forms a 7-stranded β -barrel, instead (Fig. 1). To our knowledge, the only other known example of an unusual lipocalin is triabin.¹⁴ Triabin contains also an 8-stranded β -barrel but the connectivity between the strands differs

from the all-antiparallel and fully sequential up-down connectivity characteristic for lipocalins.

Strand reduction in m-apoM is not a mere consequence of the removal of a single β -strand. As described earlier, h-apoM forms a regular 8-stranded β -barrel (labelled A-H) with seven intervening loops (L1-L7) and the barrel is flanked on its exterior by an α -helix, helix-A1 (Fig. 2).¹⁰ In m-apoM strands E and F are displaced from the β -barrel. Instead the N-terminal region of m-apoM contributes a novel β -strand A' to the barrel (residues 28-34) (Fig. 1). This significantly alters the topology and strand connectivity of the barrel. The secondary structures of the swapped segments also differ considerably in both proteins. Former strands E and F from h-apoM are partially disordered in m-apoM (residues 107-123), and only the C-terminal residues of former β -strand F participate in a 3_{10} helix in m-apoM. At the same time the novel strand A' from m-apoM adopts an irregular extended conformation in h-apoM. With the exception of these segments, all additional secondary structure elements are well conserved between m-apoM and h-apoM (Fig. 2).

Lipocalins display very low levels of sequence conservation (<20%)^{15,16} but are characterized by three structurally conserved regions (SCRs), each of which is located within a contiguous peptide segment (Fig. 2). SCR1 includes a 3_{10} helix and the loop segment that connects the 3_{10} helix to strand A of the barrel. SCR1 also contains a conserved tryptophan. SCR2 is formed by residues from strands F and G, intervening loop L6 and encompasses a conserved pair of threonine and aspartate residues. Finally, SCR3 extends from strand-H to the loop L8 and displays a conserved tyrosine and arginine residue. The latter participates in a cation- π -stacking interaction with the conserved tryptophan from SCR1 at the beginning of strand A.¹¹ Already previously, we noticed that in h-apoM only SCR1 and SCR3 are conserved since the threonine and aspartate residues characteristic for SCR2 are missing.¹⁰ In m-apoM the alternative strand topology has numerous consequences for the SCRs. SCR1 no more forms a 3_{10} helix preceding strand A. However Trp47 contained within SCR1, does maintain its characteristic orientation. Withdrawal of strands E and F leads to a complete disruption of the main chain conformation of SCR2. Only SCR3 appears to be folded identically in m-apoM and h-apoM. In m-apoM, Arg151 is able to participate in the same cation- π -stacking interaction with Trp47 as observed in other lipocalins.¹⁷ Thus, despite m-apoM adopting a different topology, a number of the side chain interactions characteristic for SCRs are preserved in m-apoM.

The differences in the main chain topology of m-apoM and h-apoM alter the disulfide bridge pattern that is formed between the 6 strictly conserved cysteines in the human and mouse protein. In h-apoM the N-terminal Cys23 is linked to Cys167 from the C-terminus of helix-A1. This disulfide bond is missing in the structure of m-apoM and there are two solvent exposed free cysteines (Cys23 and Cys169, Fig. 1 and 2). Cys23 could not be modelled in the electron density map. Because of the altered β -barrel topology this residue will be located distant to Cys169 (Fig. 1). Two potential N-glycosylation sites are present in h-apoM (residues Asn135 and Asn148). Of these two residues, only Asn135 is accessible from the surface and is utilised for glycosylation in h-apoM.¹⁸ Since only the asparagine residue from the non-solvent accessible potential glycosylation site is conserved in m-apoM (Asn150), we expect that no N-glycosylation occurs in m-apoM. This corresponds well to the previously published observation that apoM in mouse plasma is not glycosylated.¹³

The switch from an 8-stranded to a 7-stranded β -barrel reduces the size of the ligand-binding pocket enclosed within the barrel. The inner diameter of the 7-stranded barrel is about 29 Å at its broadest opening in m-apoM (measured between the C α positions of Gly93 and Tyr143 in strands D and H, respectively), tapering to a minimum diameter of 19.7 Å (Phe146 and Lys99) in the middle and 16.6 Å (between Tyr149 and Leu104) towards the bottom of the barrel. The corresponding values in h-apoM are 29 Å (between Gly93 and Tyr141), 19 Å (Phe144 and Lys99) and 20 Å (between Tyr147 and His103), respectively. We reckon that as a result of the narrowing of the lower part of the barrel, the 7-stranded β -barrel in m-apoM will not allow for binding of long-chain fatty acid ligands as previously observed for h-apoM.¹⁰ In m-apoM, access to the bottom of the ligand-binding pocket is blocked by residues Met73, Tyr149, Phe71, Leu84 and Leu147. In the solved crystal structure a glycerol molecule is bound on the top of the binding pocket and interacts with residues Glu138, Tyr102, Arg145, Trp100 and a water molecule. An additionally bound 2-propanol molecule interacts with residues Tyr102, Tyr149, Phe49, Phe63, Leu84, Leu147 and Trp100. It is noteworthy that the same residues are also involved in ligand binding in h-apoM. Taken together, differences in the β -barrel size are expected to severely alter the ligand-binding specificity of m-apoM.

Molecular dynamics simulations suggest comparable fold stabilities

In order to investigate the 7-stranded m-apoM in more detail and to relate its stability to the 8-stranded h-apoM fold, we performed molecular dynamics (MD) simulations for 50 ns at 300 K for three systems, namely the 8-stranded h-apoM, the 7-stranded m-apoM and a hypothetical 8-stranded m-apoM (Fig. 3, Supplementary Fig. S4). In the simulations, the 8-stranded and the 7-stranded β -barrels remained stable during all calculations. This is in agreement with thermal unfolding experiments reported for h-apoM and mutants thereof, which revealed melting temperatures of around 72 °C.¹⁰ The higher overall RMSD value detected for 7-stranded m-apoM can be fully attributed to fluctuations in the segment joining β -strands D and G. Omission of this flexible loop from the RMSD calculation results in values similar to those observed in the 8-stranded barrels (Supplementary Fig. S4, cyan curve).

Analysis of stabilizing interactions in m-apoM revealed two salt-bridges, which cannot be formed in h-apoM and in the 8-stranded m-apoM model due to the alternative geometry of strand A' and of the D to G segment (Fig. 3c). The first interaction occurs between Glu25 at the N-terminus of strand A', and Arg103. The second salt-bridge is formed between Asp36 and either Lys122 or Arg118, which are both located in the D to G segment. Since Asp36 and Arg103 are not conserved between m-apoM and h-apoM these salt-bridges could not be formed in a hypothetical 7-stranded h-apoM structure (Fig. 2). One could speculate that this might explain why during the purification and refolding an unusual 7-stranded β -barrel is obtained for m-apoM but not for h-apoM, which was produced using a similar refolding protocol.¹⁰

High-temperature unfolding simulations were performed at 500 K using a previously described protocol to gain a more detailed insight into the relative stability of the various secondary structure elements in the different fold topologies.¹⁹ As a measure for unfolding, distances between adjacent strands in the β -barrel were monitored (Supplementary Fig. S5). The calculations show that both the 7-stranded and the 8-stranded β -barrel exhibit a remarkable thermal stability. The contacts between strands B and C, as well as G and H remain stable over the entire simulation time. Interestingly, in all three simulations unfolding starts at a very similar position and involves a loss of the contacts formed between A' in 7-stranded m-apoM or F in h-apoM and the 8-stranded m-apoM model (Supplementary Fig. S5). When viewing unfolding as the reversal of the folding process, the present simulations suggest that inclusion of strand A' in case of the 7-stranded barrel and likewise strand F in case of the 8-stranded barrel occurs during a late step of the folding process and culminates in

the closure of the barrel. Overall, these simulations suggest that, at least in case of m-apoM, both topologies display similar stabilities and that only a low energy barrier might separate the 7-stranded topology from the 8-stranded topology in m-apoM.

Near native conformation or folding decoy?

Although the crystal structure of m-apoM reveals a 7-stranded β -barrel, the MD simulations suggest that the sequence of m-apoM is at the same time fully compatible with an 8-stranded barrel conformation. The concurrent occurrence of an 8-stranded conformation for m-apoM is also supported by the observed high sequence identity between m-apoM and h-apoM and the similar biological roles attributed to both proteins. We therefore conclude that, very likely, m-apoM is able to adopt both a 7-stranded and an 8-stranded β -barrel conformation. This immediately raises the question: What, if any at all, could be the biological relevance of the 7-stranded conformation? One could envision that switching between an 8-stranded and 7-stranded conformation could facilitate ligand release. The hydrophobic residues lining the calyx become shielded from the solvent after release of the ligand and upon conversion of the fold into a narrower 7-stranded barrel. Toggling between both conformations might be energetically not very expensive as suggested by the MD simulations. Possibly also, the observed strand removal and exchange hints to a mechanism in which strand exchange occurs in *trans*, namely through association with yet unknown proteins, which could serve as regulators for ligand release. Similar scenarios where protein-protein interactions occur through a zip-in-zip-out β -strand addition mechanism have increasingly gained interest in structural biology.^{20,21}

An alternative and possibly more obvious explanation for the occurrence of a 7-stranded β -barrel m-apoM structure is that the observed structure does not represent a native and biological functional state but rather a folding intermediate or even a misfolded state. As an on-pathway intermediate,²² the 7-stranded β -barrel possible provides a snapshot of a folding state that immediately precedes the closure of the barrel in a final folding step. In case of integral membrane proteins with β -barrel topology, like for example the outer membrane proteins in bacteria, it has been proposed that barrel formation occurs through successive β -strand augmentation in a process that is assisted by chaperone proteins.^{23,24} If one considers the 7-stranded β -barrel m-apoM structure to represent an off-pathway folding decoy,²² it becomes apparent that the topological differences that we observe are quite

remarkable and to our knowledge unprecedented. Recognized examples of crystallisable folding artefacts are so far limited to domain swapping, where two identical protein chains exchange a part of their secondary structural elements to form an intertwined dimer or higher-order oligomer.²⁵ Although domain swapping fulfils in many proteins a biological function, for some crystal structures its occurrence has been acknowledged as an artefact.²⁶ Since in a domain swapped structure all interactions between secondary structural elements are preserved and only slight alterations occur in the connectivity between secondary structure elements, it appears comprehensible that *in-vitro* refolding or even crystallisation could give rise to artificially domain swapped structures. In m-apoM however, the observed differences are significantly more extensive than those seen so far in domain swapped proteins.

However, even if the structure we determined represents an off-pathway folding artefact, the structure might still provide an interesting model for furthering our understanding of evolutionary processes. Strand invasion/withdrawal has been postulated to have played an important role during fold evolution.²⁷ Thus the fold difference that can be seen between 8-stranded lipocalins and retinol-binding protein²⁸ on one hand and the 10-stranded β -barrel that is present in retinoic acid-binding protein²⁹ on the other hand can be explained by an invasion of a β -hairpin between strands D and E in retinol-binding protein.²⁷ Evolutionary plasticity might also have lead to the unusual strand connectivity in triabin. For triabin it has been suggested that an evolutionary shift in function might have stabilized the observed strand exchange.^{14,30} M-apoM might represent an interesting case in which functional promiscuity that is expected to occur at intermediate stages during protein evolution is overruled purely by structural promiscuity. Fully functional m-apoM can inherently adopt both the 7-stranded and the 8-stranded β -barrel conformation, a fortuitous mutation could easily topple the equilibrium between both conformations and then lock a mutated m-apoM into either of these folds.

The question whether the seven-stranded lipocalin fold describes an important biological conformation or, alternatively, a mere refolding artifact cannot be answered at this time. However, our data, namely molecular dynamic calculations and the fact that the protein can be crystallized and its structure determined, strongly suggest that the energetic stability of this 7-stranded structure is close to that of the common 8-stranded lipocalin fold. If this holds true, than the crystal structure of m-apoM provides atomic insight into a novel near-native-like conformation that is accessible to lipocalins and possibly also to other β -barrel proteins. A clearer understanding of the biological relevance of this observation, however, will have to await further experimental data.

Protein Data Bank accession number

Structural data (coordinates and structure factor amplitudes) of the final m-apoM model have been deposited with the Protein Data Bank under Accession Code 2XKL.

Acknowledgements

We are thankful to Uwe Müller from BESSY synchrotron Berlin for help during data collection. This work was supported in part by the *Sonderfonds* of the University Erlangen-Nuremberg, a grant from the Swedish Research council (grant #07143) and the Swedish Heart-Lund Foundation.

References

1. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223–230.
2. Dobson, C. M. (2003). Protein folding and misfolding. *Nature* **426**, 884-890.
3. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10-19.
4. Gettins, P. G. (2002). Serpin structure, mechanism, and function. *Chem. Rev.* **102**, 4751-4804.
5. Klieber, M. A., Underhill, C., Hammond, G. L. & Muller, Y. A. (2007). Corticosteroid-binding globulin, a structural basis for steroid transport and proteinase-triggered release. *J. Biol. Chem.* **282**, 29594-29603.
6. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. (2010). Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463**, 681-684.
7. Chow, M. K., Amin, A. A., Fulton, K. F., Whisstock, J. C., Buckle, A. M. & Bottomley, S. P. (2006). REFOLD: an analytical database of protein refolding methods. *Protein Expr. Purif.* **46**, 166-171.
8. Xu, N. & Dahlback, B. (1999). A novel human apolipoprotein (apoM). *J. Biol. Chem.* **274**, 31286-31290.
9. Zhang, X. Y., Jiao, G. Q., Hurtig, M., Dong, X., Zheng, L., Luo, G. H., Nilsson-Ehle, P., Ye, Q. & Xu, N. (2004). Expression pattern of apolipoprotein M during mouse and human embryogenesis. *Acta Histochem.* **106**, 123-128.

10. Sevvana, M., Ahnstrom, J., Egerer-Sieber, C., Lange, H. A., Dahlback, B. & Muller, Y. A. (2009). Serendipitous fatty acid binding reveals the structural determinants for ligand recognition in apolipoprotein M. *J. Mol. Biol.* **393**, 920-936.
11. Flower, D. R., North, A. C. & Attwood, T. K. (1993). Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci.* **2**, 753-761.
12. Deakin, J. E., Papenfuss, A. T., Belov, K., Cross, J. G., Coggill, P., Palmer, S., Sims, S., Speed, T. P., Beck, S. & Graves, J. A. (2006). Evolution and comparative analysis of the MHC Class III inflammatory region. *BMC Genomics* **7**, 281.
13. Faber, K., Axler, O., Dahlback, B. & Nielsen, L. B. (2004). Characterization of apoM in normal and genetically modified mice. *J. Lipid Res.* **45**, 1272-1278.
14. Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W. D., Huber, R. & Bode, W. (1997). Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatomine bug. *Proc. Natl. Acad. Sci. USA* **94**, 11845-11850.
15. Flower, D. R. (1996). The lipocalin protein family: structure and function. *Biochem. J.* **318**, 1-14.
16. Flower, D. R., North, A. C. & Sansom, C. E. (2000). The lipocalin protein family: structural and sequence overview. *Biochim. Biophys. Acta* **1482**, 9-24.
17. Greene, L. H., Chrysina, E. D., Irons, L. I., Papageorgiou, A. C., Acharya, K. R. & Brew, K. (2001). Role of conserved residues in structure and stability: tryptophans of human serum retinol-binding protein, a model for the lipocalin superfamily. *Protein Sci* **10**, 2301-2316.
18. Duan, J., Dahlback, B. & Villoutreix, B. O. (2001). Proposed lipocalin fold for apolipoprotein M based on bioinformatics and site-directed mutagenesis. *FEBS Lett.* **499**, 127-132.
19. Schulenburg, C., Weininger, U., Neumann, P., Meiselbach, H., Stubbs, M. T., Sticht, H., Balbach, J., Ulbrich-Hofmann, R. & Arnold, U. (2010). Impact of the C-terminal disulfide bond on the folding and stability of onconase. *Chembiochem.* **11**, 978-986.
20. Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E. & Waksman, G. (2006). Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol. Cell* **22**, 831-842.
21. Remaut, H. & Waksman, G. (2006). Protein-protein interaction through beta-strand addition. *Trends Biochem. Sci.* **31**, 436-444.

22. Baldwin, R. L. (1996). On-pathway versus off-pathway folding intermediates. *Fold. Des.* **1**, R1-8.
23. Robert, V., Volokhina, E. B., Senf, F., Bos, M. P., Van Gelder, P. & Tommassen, J. (2006). Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol.* **4**, e377.
24. Knowles, T. J., Jeeves, M., Bobat, S., Dancea, F., McClelland, D., Palmer, T., Overduin, M. & Henderson, I. R. (2008). Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. *Mol. Microbiol.* **68**, 1216-1227.
25. Rousseau, F., Schymkowitz, J. W. & Itzhaki, L. S. (2003). The unfolding story of three-dimensional domain swapping. *Structure* **11**, 243-251.
26. Schoch, G. A., D'Arcy, B., Stihle, M., Burger, D., Bar, D., Benz, J., Thoma, R. & Ruf, A. (2010). Molecular switch in the glucocorticoid receptor: active and passive antagonist conformations. *J. Mol. Biol.* **395**, 568-577.
27. Grishin, N. V. (2001). Fold change in evolution of protein structures. *J. Struct. Biol.* **134**, 167-185.
28. Zanotti, G., Berni, R. & Monaco, H. L. (1993). Crystal structure of liganded and unliganded forms of bovine plasma retinol-binding protein. *J. Biol. Chem.* **268**, 10728-10738.
29. Kleywegt, G. J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K. & Jones, T. A. (1994). Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-trans-retinoic acid and a synthetic retinoid. *Structure* **2**, 1241-1258.
30. Murzin, A. G. (1998). How far divergent evolution goes in proteins. *Curr. Opin. Struct. Biol.* **8**, 380-387.
31. Homeyer, N., Essigke, T., Meiselbach, H., Ullmann, G. M. & Sticht, H. (2007). Effect of HPr phosphorylation on structure, dynamics, and interactions in the course of transcriptional control. *J. Mol. Model.* **13**, 431-444.
32. Wartha, F., Horn, A. H. C., Meiselbach, H. & Sticht, H. (2005). Molecular Dynamics Simulations of HIV-1 Protease Suggest Different Mechanisms Contributing to Drug Resistance. *J. Chem. Theory Comput.* **1**, 315-324
33. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38, 27-38.
34. Fiser, A., Do, R. K. & Sali, A. (2000). Modeling of loops in protein structures. *Protein Sci.* **9**, 1753-1773.

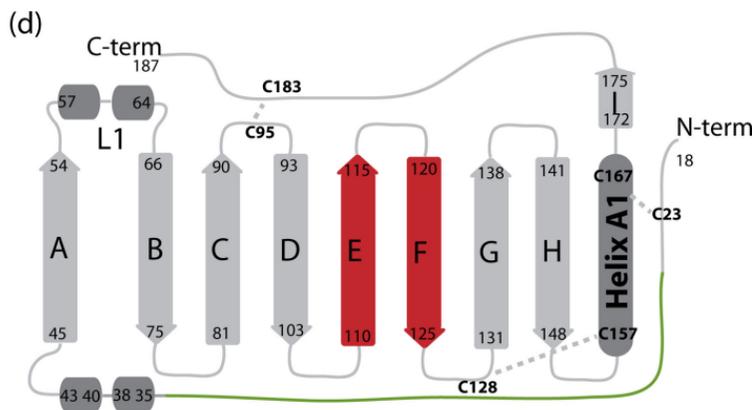
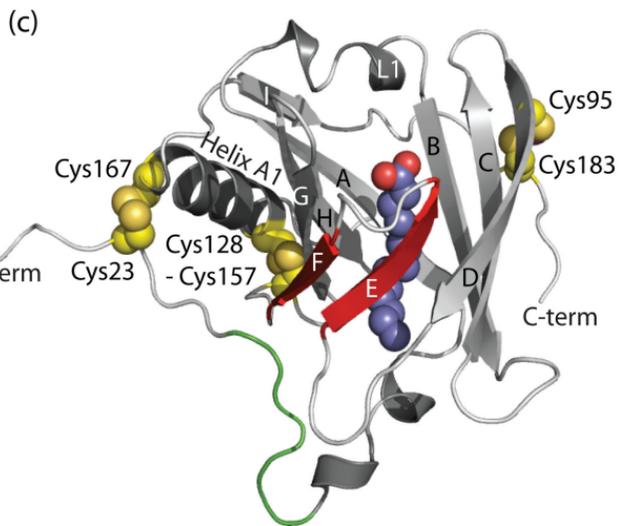
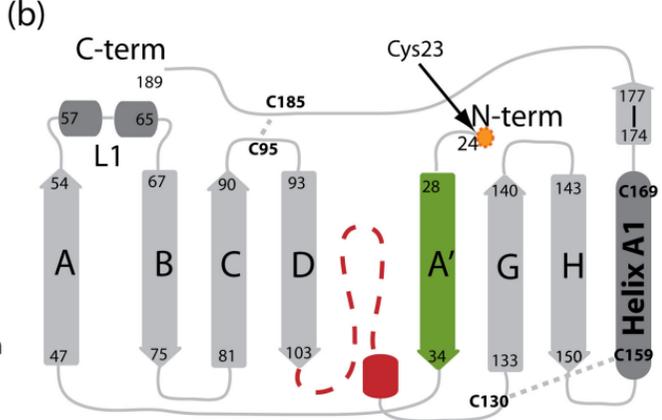
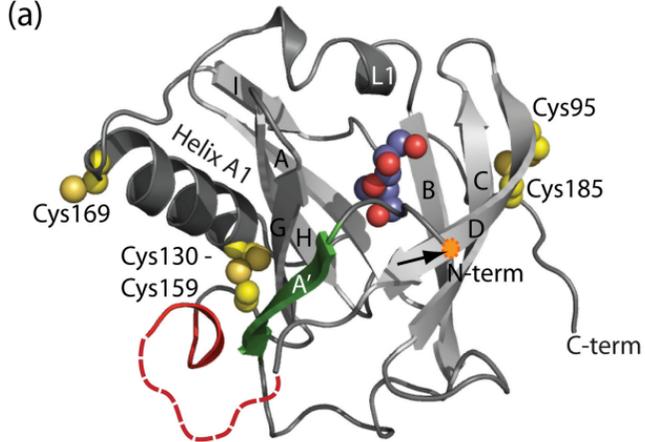
35. Fiser, A. & Sali, A. (2003). ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **19**, 2500-2501.
36. Ahnstrom, J., Faber, K., Axler, O. & Dahlback, B. (2007). Hydrophobic ligand binding properties of the human lipocalin apolipoprotein M. *J. Lipid Res.* **48**, 1754-1762.
37. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800.
38. McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **63**, 32-41.
39. Terwilliger, T. C. (2004). Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **60**, 2144-2149.
40. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240-255.
41. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126-2132.
42. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458-463.
43. DeLano, W. (2003). The PyMOL Molecular Graphics System. *DeLano Scientific LLC, San Carlos, CA.*

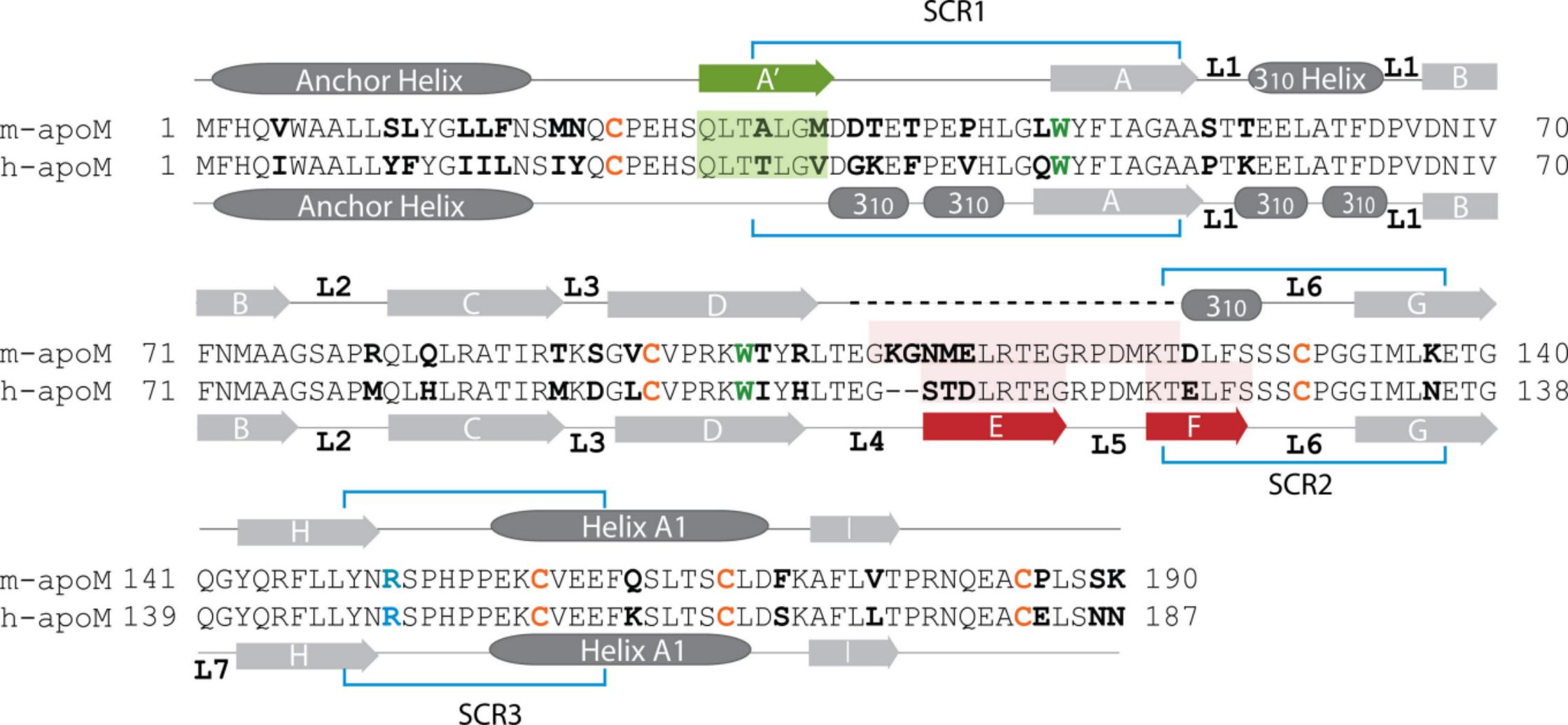
Figure legends

Fig. 1. Crystal structures and topology plots of m-apoM (a, b) and h-apoM (c, d). Residues forming strands E and F in h-apoM are colored in red and shown as broken lines followed by a 3_{10} helix in m-apoM (a and b). The newly formed N-terminal strand A' is colored green in m-apoM (a, b) and the corresponding region also highlighted in green in h-apoM (c, d). The bound ligands are shown as sphere models and colored in blue and red. The cystines and the free cysteine Cys169 of m-apoM are shown in yellow. The position assumed for missing residue Cys23 in m-apoM is highlighted by an orange circle (a and b).

Fig. 2. Sequence alignment of m-apoM and h-apoM together with the secondary structure elements observed in the crystal structures. Sequence differences are highlighted in bold. Residues that form the new strand A' in m-apoM are highlighted in green. Residues corresponding to strands E and F in h-apoM are highlighted in red. Conserved cysteines and tryptophans are shown in orange and green, respectively. Structural conserved regions (SCRs) characteristic for lipocalins are marked with blue brackets. Conserved residue Arg151 (m-apoM) that participates in a cation- π interaction with Trp47 in both m-apoM and h-apoM is highlighted in blue. A loop, which is not visible in the electron density maps and which interconnects strand D and a 3_{10} helix in m-apoM, is indicated by a broken line (For a more detailed comparison of h-apoM with other lipocalins please see Sevvana et al,¹⁰).

Fig. 3. Molecular dynamic (MD) simulation of m-apoM at 300 K over 50 ns following an established protocol.^{31,32} Stabilization of m-apoM occurs via two salt bridges: (a) between Glu25 and Arg103 and (b) between Asp36 and Lys122 or, alternatively, Asp36 and Arg118. Salt-bridges were analyzed with the molecular visualization program VMD1.8.5.³³ Residues Asp36 and Arg103 are marked with an asterix because they are not conserved between m-apoM and h-apoM. (c) Schematic representation of the 7-stranded m-apoM structure used for the MD calculation. The loop connecting β -strand D to β -strand G in m-apoM was not resolved in the X-ray structure. It was added using SwissModel and Modloop and is colored in red.^{34,35} Strand A' of m-apoM is highlighted in green. The salt bridge forming residues are shown in a stick representation.





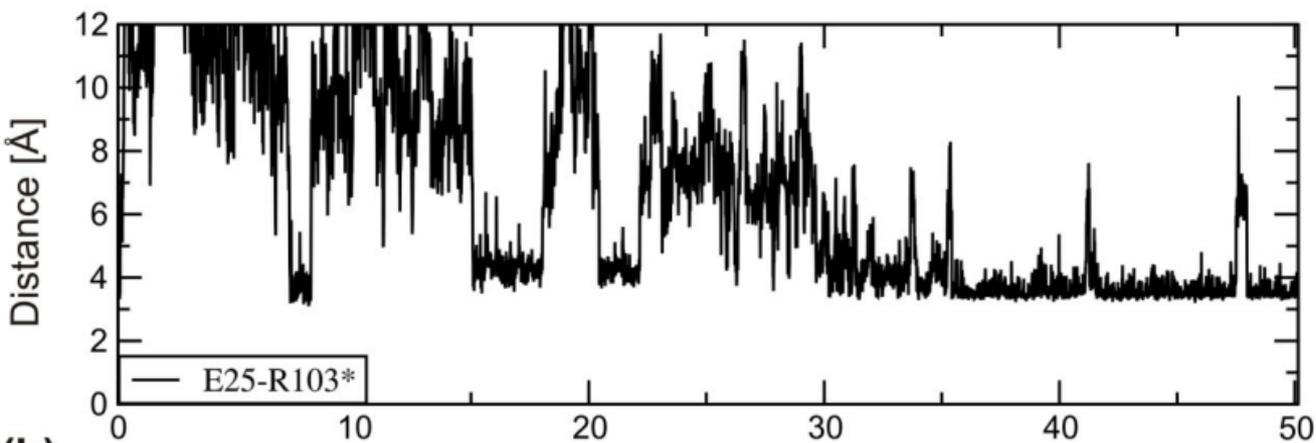
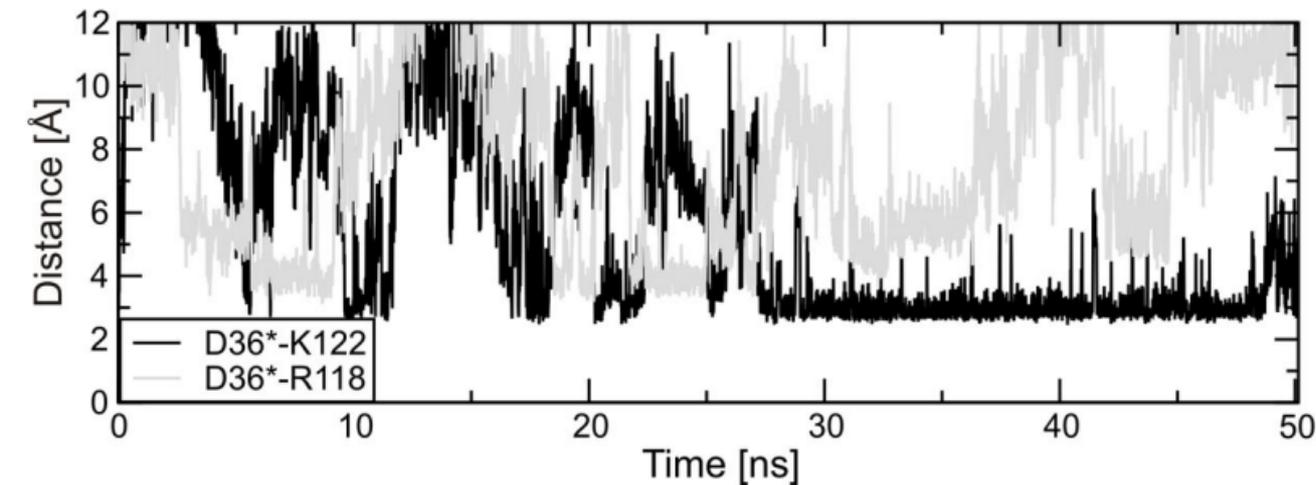
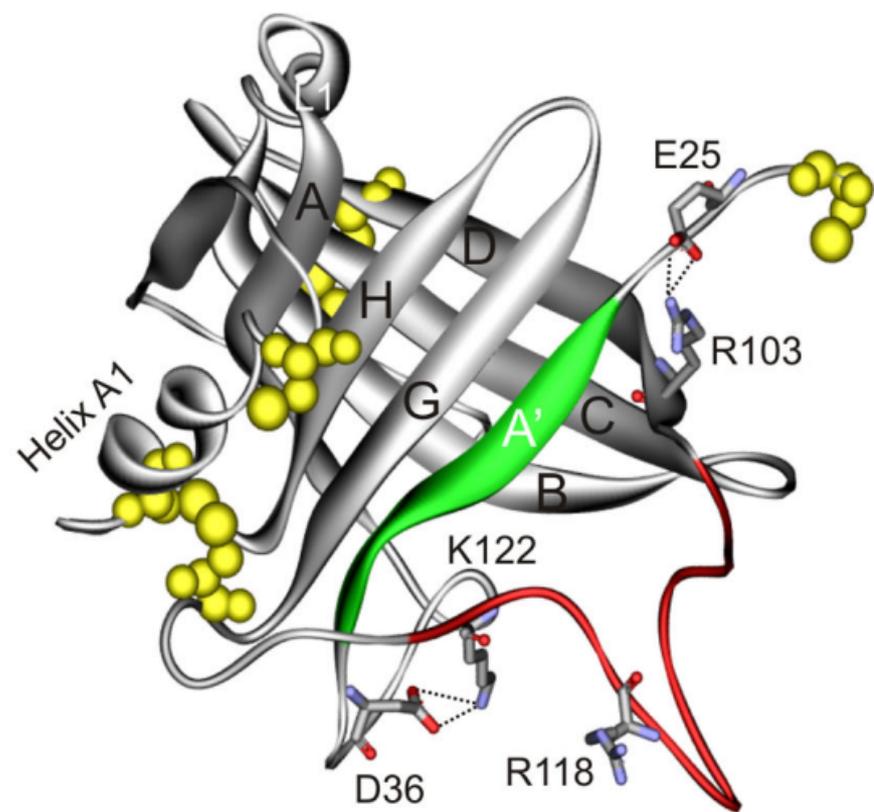
(a)**(b)****(c)**

Table 1 Data collection and refinement statistics

Data Collection Statistics	
Wavelength (Å)	0.9184
X-ray Source	Bessy Synchrotron
Detector	MarCCD (165 mm)
Space group	P6 ₅ 22
Unit Cell Parameters	a = b = 53.93 Å, c = 207.48 Å, $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$
Matthews Coefficient(Å ³ /Da ⁻¹)	2.28
Molecules/asymmetric unit	1
Solvent content	46.02
Resolution (Å) ⁺	2.50 (2.65 – 2.50)
No. of Reflections (Unique)	52683 (6802)
Redundancy ⁺	7.74 (8.08)
Completeness (%) ⁺	99.8 (99.9)
Mean I/(σ I) ⁺	10.84 (3.56)
R _{sym} (%) ⁺	18.7 (58.6)
Wilson B-factor (Å ²)	26.489
Refinement Statistics	
Final R-factor (%)	
Working set	18.73
Working set + test set	19.03
Final free R-factor (%) [#]	24.73
RMSDs	
Bond lengths (Å)	0.012
Bond angles (°)	1.607
Mean B-value (Å²)	
Main chain atoms	19.72
Side chain atoms	21.25
Ligand	30.00
Solvent	22.43
No. of protein atoms	1163
Ligand atoms	94
No. of solvent atoms	15
Ramachandran plot	
(% ; core / additionally / generously /disallowed)	90.4 / 8.0 / 0.8 / 0.8
PDB ID code	2XKL

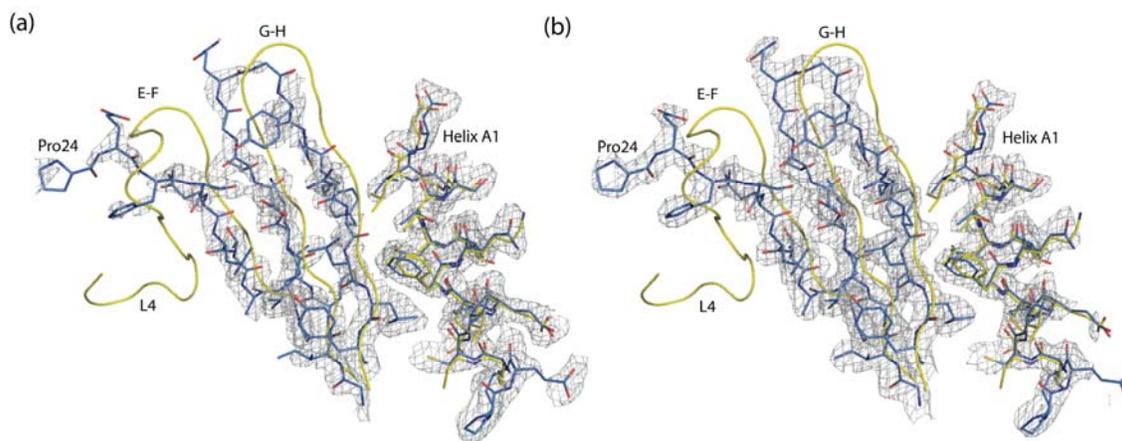
+ Numbers in parenthesis are for the highest-resolution shell. # 5% of reflections have been chosen as R_{free} set. R_{sym} is calculated as $\sum_n \sum_i |I_i - \langle I \rangle| / \sum_n \sum_i \langle I \rangle$ where I_i is the i^{th} observation of the n^{th} reflection and $\langle I \rangle$ the mean of all observations of the n^{th} reflection.

Truncated m-apoM (amino acids 22-190; swiss-prot entry Q9Z1R3) was originally

amplified from a mouse liver 5' stretch plus cDNA library (Clontech) and subsequently cloned into pET30Xa/LIC vector (Novagen) containing an N-terminal His-tag followed by an S-tag. A thrombin cleavage site was present after the His-tag and an additional FXa-cleavage site separated the S-tag from m-apoM. Protein production in *E.coli* resulted in the formation of inclusion bodies. Protein purification was essentially carried out as described earlier.³⁶ In short, the N-terminal hexa-histidine (hexa-His) tag allowed purification of the protein using a nickel chromatography column prior to any refolding. The protein was refolded upon dialysis with a protocol including oxidized and reduced glutathione and further purified with a Q-Sepharose column (GE Healthcare Bio-Sciences, Uppsala, Sweden). The N-terminal hexa-His-tag was then cleaved off with thrombin and removed with an additional Q-Sepharose column. Removal of the S-tag with FXa was not successful. As a consequence, the protein was crystallized with the S-tag attached. M-apoM migrated as a single homogenous band on SDS-PAGE under both reducing and non-reducing conditions with an apparent molecular weight just above 25 kDa. The purity of m-apoM was identical to that of the previously described human protein, i.e. > 95 %.³⁶

M-apoM was crystallized using the hanging drop vapor diffusion method and condition 40 (0.1 M Sodium citrate, pH 5.6, 20% v/v 2-propanol, 20% w/v PEG 4000) of the Hampton Crystal Screen 1 solutions (Hampton Research, Aliso Viejo, CA). The crystal was flash frozen in liquid nitrogen using 20% ethylene glycol as a cryoprotectant for data collection at 100 K. A single native data set was collected at BESSY synchrotron at MX beamlines in Berlin. The diffraction data were processed using the programs *XDS* and *XSCALE*.³⁷ The protein crystallized in space group P6₅22 with one molecule in the asymmetric unit, and crystals diffracted to a maximum resolution of 2.50 Å.

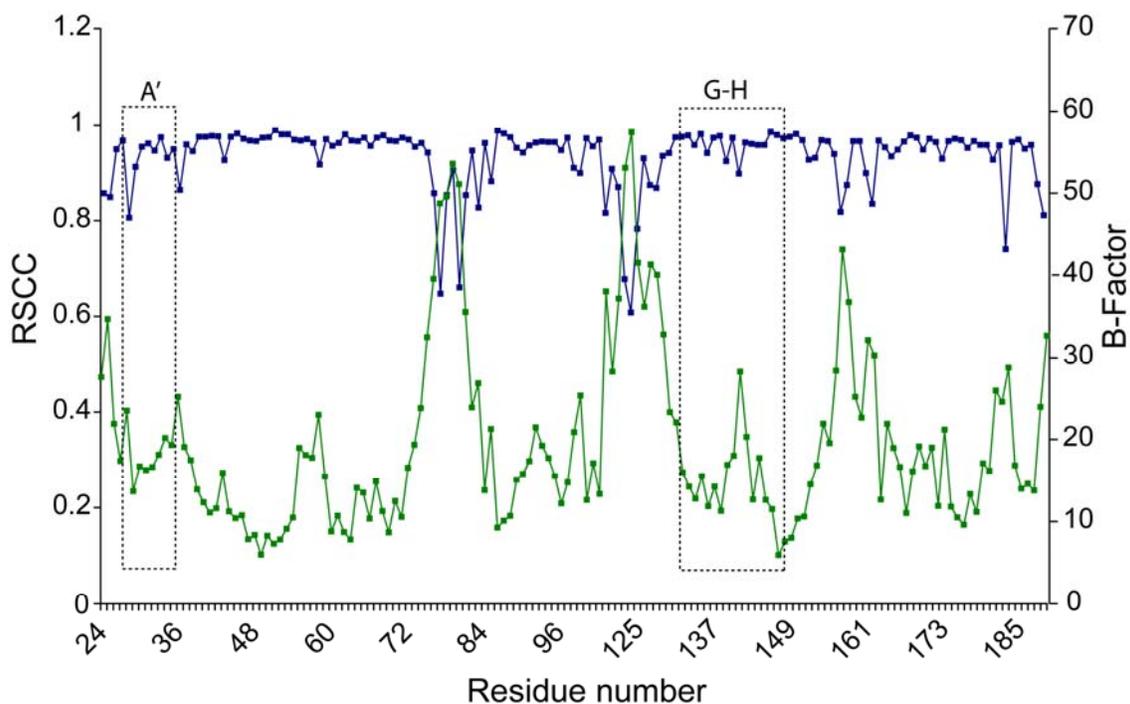
The structure of m-apoM was solved by molecular replacement using the program *PHASER*³⁸ and the published h-apoM structure as a search model (PDB ID: 2WEW). Model bias was minimized using prime and switch phasing in *RESOLVE*.³⁹ The initial maps immediately indicated that there was a rearrangement in the β -barrel topology in comparison to the human homologue and search model. The structure was refined using TLS refinement in *REFMAC5*⁴⁰ alternating with model building in real space using *2mFo-DFc* and *mFo-DFc* maps in *COOT*,⁴¹ followed by cycles of *ARP* solvent building.⁴² The structure was validated as indicated in the supplementary information (Supplementary Fig. S1 and S2) and illustrated using *PYMOL*.⁴³



Supplementary Fig. S1. (a) OMIT map calculated after deletion of strands-A' and H-G from the final m-apoM model and after refinement of this truncated model with *REFMAC*.¹ (b) Electron density map ($2mF_o-DF_c$) of the final refined m-apoM model. The m-apoM model is shown in blue sticks and as a reference h-apoM in yellow. Both maps are contoured at 1σ .

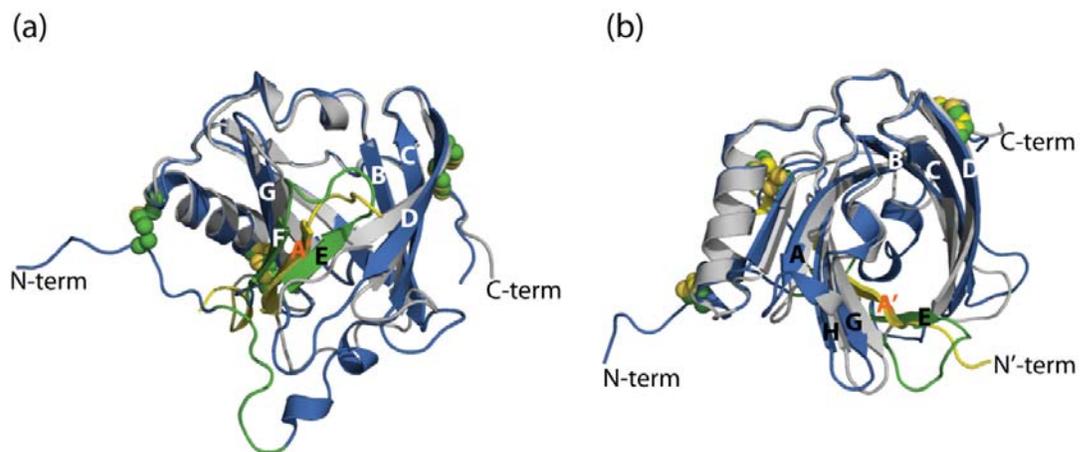
Methods: Owing to the unusual fold of m-apoM we carried out numerous validation checks at all stages of the structure determination process: (I) *Data validation:* The R_{sym} of the data set was very high (18.7%). Since the hexagonal space group is of high symmetry (P6₅22), we tested the data for possible merohedral twinning. The result was negative. The high R-factors result from the weak data that we collected from a very small crystal. (II) *Validation of the molecular replacement solution:* The structure solution with h-apoM as a search model using *PHASER*,² yielded a solution with a rotation function Z-score of 7.0 and translation function Z-score of 18.8. An initial refinement step yielded an R_{work} of 39.4 % and an R_{free} of 45%. The first sigmaA map from *REFMAC*¹ indicated clear difference density in regions with major conformational differences. Model bias was then minimized using the prime and switch phasing method implemented in *RESOLVE*.³ The resulting composite OMIT map showed improved electron densities in the regions exhibiting major conformational differences between h-apoM and m-apoM and was subsequently used to build the m-apoM model. After the first model building the R-factors decreased to 29.28% and 34.43% (R_{work} and R_{free}). A portion of the final refined map is shown in Supplementary Fig. S1b. (III) *Validation of the refined model:* The model was validated using the *MOLPROBITY*⁴ server, which yielded a Molprobity score of 2.88 (47th percentile; 100th percentile is the best among structures of comparable resolution). Model validation with other available software and servers (*PROCHECK*,⁵ *SFCHECK*,⁶ *WHATCHECK*⁷) indicated no major problems in the refined model. We further calculated an omit map after deleting the regions strand-A' and H-G from the final refined m-apoM model and applying one round of refinement with

REFMAC.¹ The omitted region displays clear density in the OMIT map (Supplementary Fig. S1a).

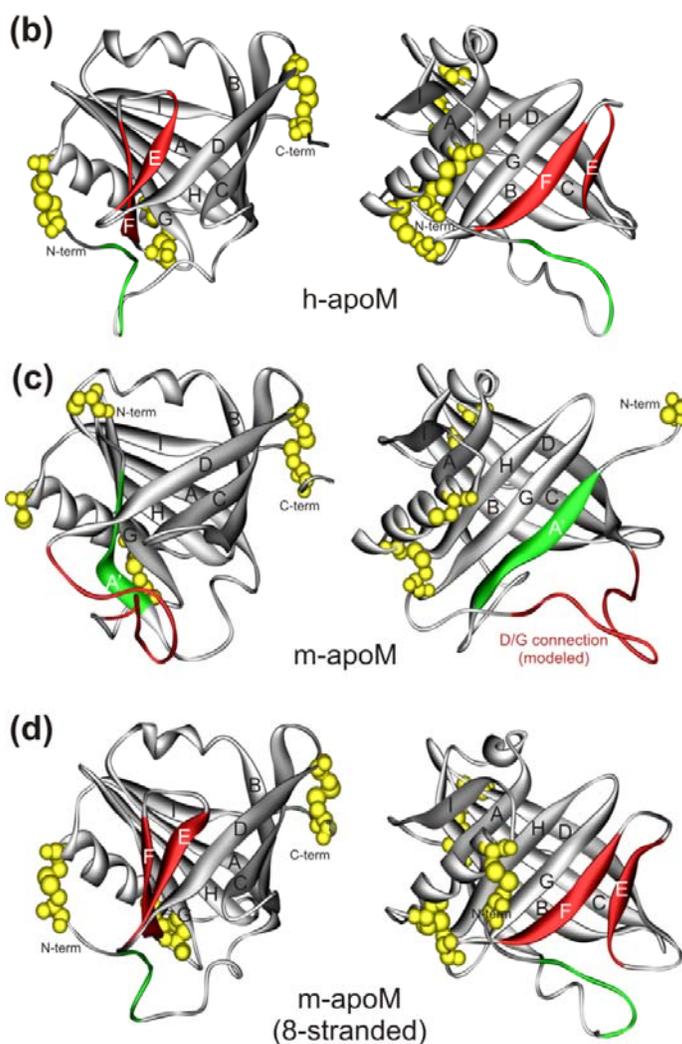
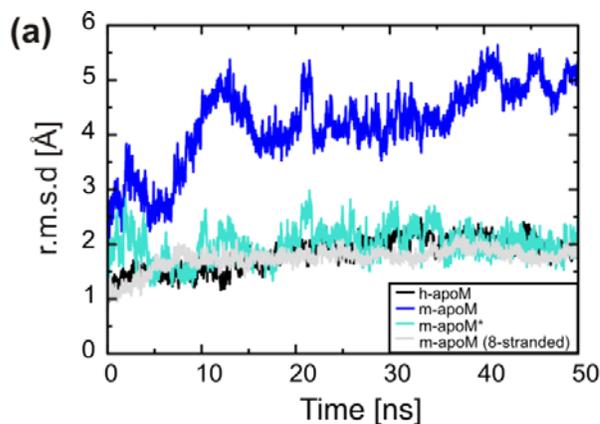


Supplementary Fig. S2. Plot showing real space correlation coefficients (blue line) and B-factors (green line) for each residue. The N-terminal residues modeled as strand A' and the strands G-H that differ in the superposition of m-apoM and h-apoM show high correlation coefficients.

Methods: In order to further validate the refined model against its electron density, we calculated the real space correlation coefficient and average B-value of each residue in the final electron density map using the *EDS* server.⁸

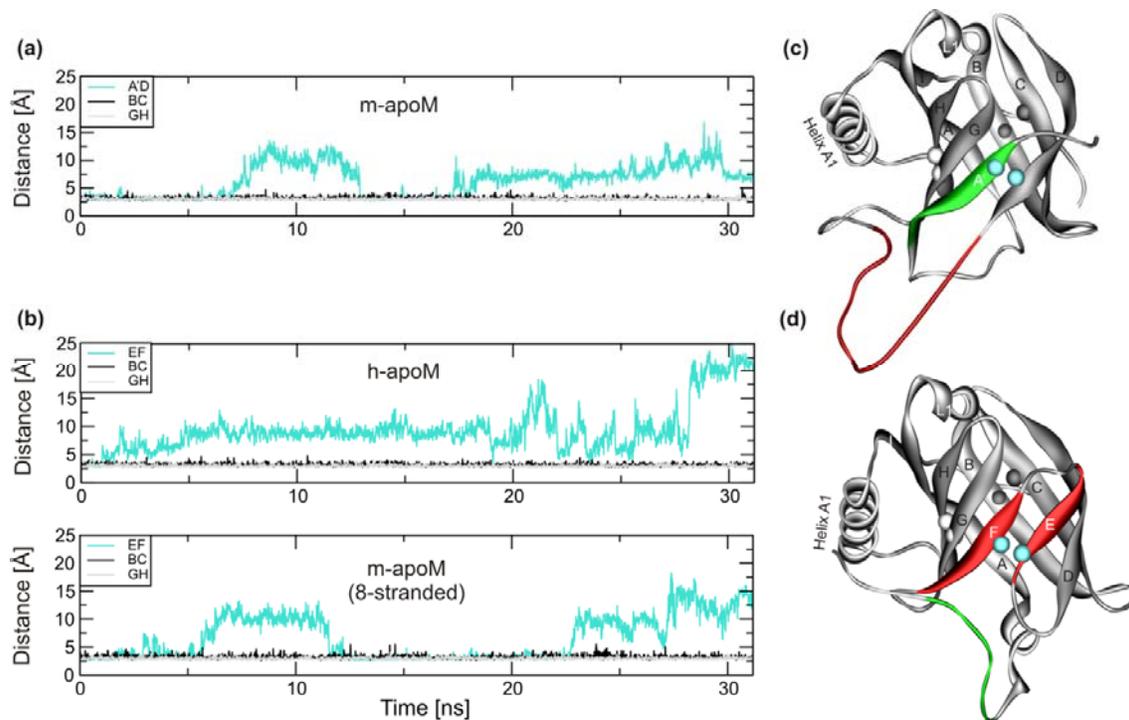


Supplementary Fig. S3. Superposition of h-apoM (blue) and m-apoM (gray). (a) Side view, (b) top view. The disulfide bonds, strands E and F and the N-terminal region corresponding to inserted strand A' in mApO_M are highlighted in green in the h-apoM structure. The identical regions are shown in yellow in m-apoM. Selected secondary structure elements are marked.



simulation: (b) h-apoM, (c) m-apoM, and (d) modeled 8-stranded m-apoM. The inserted N-terminal strand A' in m-apoM and strands E and F in h-apoM and 8-stranded m-apoM model are highlighted green and red, respectively. The D to G connection modeled in m-apoM (c) is shown in red.

Supplementary Fig. S4. Stability of m-apoM, h-apoM and of a hypothetical 8-stranded m-apoM model as monitored by molecular dynamics calculations at 300 K. The 8-stranded m-apoM model was obtained by modelling the mouse sequence onto the h-apoM crystal structure using SwissModel.⁹ The loop connecting β -strands D and G in m-apoM was not resolved in the X-ray structure and was therefore supplemented using SwissModel and Modloop.^{10,11} Any bound ligands were removed to allow comparability of the systems. The three systems were minimized and subjected to a 50 ns molecular dynamics (MD) simulation at 300 K according to an established protocol.^{12,13} (a) Backbone r.m.s.d. values calculated for the entire models. Also shown for m-apoM: r.m.s.d. values with the flexible segment joining strands D and G omitted (m-apoM*). Two views of the structures at the end of the 50 ns molecular dynamics



Supplementary Fig. S5. Unfolding MD simulations at 500 K. To detect the starting points for thermal unfolding, hydrogen bonding distances were analyzed between three different pairs of β -strands in each of the following three structures: (a) m-apoM and (b) h-apoM and 8-stranded m-apoM model. The location of the residues, for which main chain hydrogen bonding distances were analyzed, are depicted as colored balls in m-apoM (c) and h-apoM (d). The following distances were analyzed:

m-apoM:

A'D: Leu29-N – Arg103-O

BC: Val70-N – Thr87-O

GH: Leu136-N - Leu147-O

h-apoM and (modelled 8-stranded m-apoM):

EF: Leu111-N (Leu113-N) - Thr121-O (Thr123-O)

BC: Val70-N – Thr87-O

GH: Leu134-N (Leu136-N) – Leu145-O (Leu147-O)

Supplementary References

1. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D* **53**, 240-255.
2. McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. Sect. D* **63**, 32-41.
3. Terwilliger, T. C. (2004). Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr. Sect. D* **60**, 2144-2149.
4. Chen, Y. W., Stott, K. & Perutz, M. F. (1999). Crystal structure of a dimeric chymotrypsin inhibitor 2 mutant containing an inserted glutamine repeat. *Proc. Natl. Acad. Sci. USA* **96**, 1257-1261.
5. Laskowski, R. A., Moss, D. S. & Thornton, J. M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049-1067.
6. Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. Sect. D* **55**, 191-205.
7. Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature* **381**, 272.
8. Kleywegt, G. J., Harris, M. R., Zou, J. Y., Taylor, T. C., Wahlby, A. & Jones, T. A. (2004). The Uppsala Electron-Density Server. *Acta Crystallogr. Sect. D* **60**, 2240-2249.
9. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201.
10. Fiser, A., Do, R. K. & Sali, A. (2000). Modeling of loops in protein structures. *Protein Sci.* **9**, 1753-1773.
11. Fiser, A. & Sali, A. (2003). ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **19**, 2500-2501.
12. Homeyer, N., Essigke, T., Meiselbach, H., Ullmann, G. M. & Sticht, H. (2007). Effect of HPr phosphorylation on structure, dynamics, and interactions in the course of transcriptional control. *J. Mol. Model.* **13**, 431-444.
13. Wartha, F., Horn, A. H. C., Meiselbach, H. & Sticht, H. (2005). Molecular Dynamics Simulations of HIV-1 Protease Suggest Different Mechanisms Contributing to Drug Resistance. *J. Chem. Theory Comput.* **1**, 315-324