Structural study on ligand specificity of human vitamin B12 transporters

Jochen Wuerges, Silvano Geremia, Lucio Randaccio

Biochemical Journal Immediate Publication. Published on 5 Feb 2007 as manuscript BJ20061394

Structural study on ligand specificity of human vitamin B12 transporters

Jochen Wuerges, Silvano Geremia[1] and Lucio Randaccio

Centre of Excellence in Biocrystallography, Department of Chemical Sciences, University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy

Short Title: HUMAN VITAMIN B12 TRANSPORTERS

[1] To whom correspondence should be addressed (E-Mail: sgeremia@units.it)

Abbreviations used: DMB, 5,6-dimethylbenzimidazole; Cba, cobamide; Cbi, cobinamide; Cbl, cobalamin; IF, intrinsic factor; HC, haptocorrin; TC, transcobalamin; rms, root mean square.

Studies comparing the binding of genuine cobalamin (vitamin B12) to that of its natural or synthetic analogues have long established increasing ligand specificity in the order haptocorrin, transcobalamin and intrinsic factor, the high-affinity binding proteins involved in cobalamin transport in mammals. Here, ligand specificity is investigated from a structural point of view, for which comparative models of intrinsic factor and haptocorrin are produced based on the crystal structure of the homologous transcobalamin and validated by results of published binding assays. Many interactions between cobalamin and its binding site in the interface of the two domains are conserved among the transporters. A structural comparison suggests that the determinant of specificity regarding cobalamin ligands with modified nucleotide moiety resides in the β-hairpin motif β3–turn–β4 of the smaller C-terminal domain. In haptocorrin, it provides hydrophobic contacts to the benzimidazole moiety through the apolar part of Arg^{357}, Trp^{359} and Tyr^{362}. Together, these large side chains may compensate for the missing nucleotide upon cobinamide binding. Intrinsic factor possesses only the Trp and transcobalamin only the Tyr, consistent with their low affinity for cobinamide. Relative affinity constants for other analogues are rationalized similarly by analysis of steric and electrostatic interactions with the three transporters. The structures also indicate that the C-terminal domain is the first site of cobalamin-binding since part of the β-hairpin motif is trapped between the nucleotide moiety and the N-terminal domain in the final holo-proteins.

Key words: cobalamin, transport proteins, transcobalamin, haptocorrin, intrinsic factor, comparative modeling, structure comparison

INTRODUCTION

The transport of the essential micronutrient cobalamin (Cbl, vitamin B_{12}) from food to cells relies on three successive proteins in mammals, haptocorrin (HC), gastric intrinsic factor (IF) and transcobalamin (TC) [1-3]. Cbl is required by cells as the basis for two enzyme cofactors, methyl-Cbl for methionine synthase and 5'-deoxyadenosyl-Cbl (Ado-Cbl) for methyl-malonyl-CoA mutase [4]. Upon initial uptake of Cbl from food, Cbl becomes bound in the stomach to salivary HC. After proteolysis of HC in the duodenum, Cbl is passed on to IF. Mucosal cells in the ileum absorb the IF-Cbl complex via endocytosis mediated by a specific receptor. In the enterocyte, the IF-Cbl complex is degraded and Cbl is transferred to TC which delivers Cbl to cells via the blood [5]. Only the fraction of Cbl bound to TC is taken up via endocytosis by a specific receptor on most cell types. HC that is also present in plasma cannot facilitate cellular uptake of Cbl with the exception of hepatocytes. TC-Cbl is degraded in lysozomes to release Cbl for further conversion into the important cofactors. Cbl-deficiency as a result of impaired intestinal absorption is more common among the elderly population and may lead to disorders including hematologic abnormalities and defects in the nervous system and metabolism [6].

Table 1 summarizes some properties of the human Cbl-transporters. Investigation of their gene structure suggested that the three proteins have a common evolutionary origin and that TC has diverged first from the ancestral gene [7]. The homologous Cbl-transporters show about 25% overall sequence identity but several shorter stretches possess considerably higher similarity. Glycosylation is observed for IF and to an even higher extent for HC, but not for TC [2]. The proteins have extraordinary affinity for the vitamin, with similar K_d values < 1 pM [8]. However, they show differences in their specificity for Cbl analogues, which are as well produced by microorganisms [9].

Very recently, the structures of human and bovine TC in complex with Cbl have been solved by X-ray crystallography [10]. These allowed the detailed description of the Cbl-binding mode and represent the prototype for the family of mammalian Cbl-transporters since no structures have been reported yet for IF or HC. Their structure determination by X-ray crystallography is likely to remain challenging due to difficulties with crystallization of these highly glycosylated proteins. Detailed structural information for IF and HC is of importance for understanding their way of function in greater detail, especially the interactions with the ligand Cbl, the role of glycosylation and the recognition by specific cell surface receptors.

In particular, knowledge of the Cbl-binding mode for all three proteins is essential to rationalize the observed diversity in ligand specificity. IF shows the highest specificity for Cbl and is suggested to select in the first place genuine Cbl among Cbl analogues in the diet, while HC as the least specific binder may remove other Cbl analogues from the circulation that can harmfully compete for the binding site of Cbl in methionine synthase and methyl-malonyl-CoA mutase [9]. The precise physiological role of HC within the complex process of Cbl internalization is not yet well understood. Furthermore, the structural basis for ligand specificity is of interest in view of the design of Cbl-based bioconjugates. Such bioconjugates show potential as cytotoxic drugs or imaging agents in cancer treatment [11] or may function as Cbl-

competitors to selectively block cell growth in the treatment of AIDS-related lymphoma [12] or cancer [13]. These applications are based on the observation that proliferating cells express more surface receptors for holo-TC, and thus take up proportionally more Cbl, than cells in the stationary phase [11,14].

In this study we analyze all three Cbl-transporters regarding structural determinants of ligand specificity. For this purpose, we produced comparative models of Cbl-complexed human IF and HC guided by our recently determined X-ray structures of human and bovine TC. The models are validated with the help of available experimental data for a variety of structural features and are compared to the human TC structure.

METHODS

Modelling

The comparative modelling of IF and HC is based on the structures of human and bovine TC determined by X-ray crystallography (PDB accession codes 2BB5 and 2BB6, respectively [10]) and on a multiple sequence alignment of IF and HC with these two TC forms shown in Figure 1. This alignment was extracted from a larger multiple sequence alignment to increase its accuracy, employing the following 16 mammalian Cbl-transporting proteins: TC from man (SwissProt entry P20062), chimpanzee (XP_525562), orangutan (Q5REL7), dog (XP_543481), rat (Q9R0D6), mouse (O88968) and cow (Q9XSC9); IF from man (P27352), dog (Q5XWD5), rat (P17267), mouse (P52787) and cow (XP_873322); HC from man (P20061), dog (XP_855361), cow (XP_873306) and hog (P17630). The alignment was performed with CLUSTALW v.1.82 [15] with the Gonnet 250 scoring matrix. 10 models were calculated both for IF and HC by MODELLER 8v1 [16] using default parameters. The Cbl ligand of the TC template structures were included as rigid body ("block residue" definition in MODELLER) whereas water molecules were excluded from the templates. Conserved disulfide bridges were automatically used as geometric restraints for IF and HC. The best model was chosen according to MODELLER's lowest objective function criterion. The respective best model was subsequently refined iteratively by geometry idealization with REFMAC5 [17] and model rebuilding with the graphics program O [18]. The energy profile (MODELLER's DOPE score) of the models was compared to those of the experimentally determined TC structures. The model geometry was analysed with MOLPROBITY [19] regarding Ramachandran plot, van der Waals contacts, hydrogen bonds and steric clashes and with PROCHECK [20] regarding amino acid stereochemistry. Cavities in the models accessible to a probe of 1.4 Å radius (corresponding to a water molecule) were found with the program CAVENV [21] and visualized with O. Root mean square (rms) deviations of Cα-atoms in a pair of models were calculated by DALI [22]. The electrostatic potential on the protein surface was calculated with APBS [23] assuming neutral pH. Protein structures were visualized using PYMOL (DeLano Scientific, South San Francisco, CA, USA). Atomic accessibility areas were calculated with NACCESS (S.L. Hubbard and J.M. Thornton, University College London, London, UK).

Deposition of atomic coordinates

The atomic coordinates have been deposited in the Protein Data Bank (http://www.rcsb.org) with accession number 2CKT for IF and 2CKV for HC.

RESULTS AND DISCUSSION

Description of the IF and HC models and comparison to TC

Comparative models were produced for the human forms of IF and HC on the basis of the sequence alignment with human and bovine TC (Figure 1) and guided by the recently determined crystal structures of human and bovine TC [10]. The overall correctness of the modelling is indicated by the absence of bad overlaps and large holes. The models show compact globular domains, well defined hydrogen-bonds for buried polar residues and secondary structure elements, as well as extended hydrophobic domain cores. Small cavities are found which in size and location are similar to the water-occupied cavities observed in the crystal structures of TC. The energy profiles of the models are similar to those of the TC template structures. IF and HC show reasonable stereochemistry with regard to rms deviations from ideal geometry and Ramachandran plot statistics (Supplementary Table 1).

As the template protein TC, the comparative models of IF and HC possess a two-domain architecture in which a larger N-terminal domain ("α-domain"), composed of 12 α-helices in an α_6-α_6 barrel and a short 3/10 helix, is connected by a flexible linker region to a smaller C-terminal domain ("β-domain") consisting mainly of two β-sheets. The Cbl ligand, with its 5,6-dimethylbenzimidazole (DMB) moiety coordinated to the Co ion ("base-on" form), is tightly enclosed in the domain interface (Figure 2). The sequence identity of human IF and HC as well as the rms deviations of superimposed Cα-atoms of their models are given in Table 2, including also the two TC forms. The regions of elevated similarity between the Cbl-transporters in the sequence alignment of Figure 1 were mapped on the experimentally determined human TC structure. Regarding the α-domain, these regions coincide with the inner six (even-numbered) helices, i.e. the core of the domain, while similarity is spread over the entire β-domain.

An experimental proof of the presence of a two-domain architecture in human IF has been given previously [24,25]. Two fragments, about 30 kDa and 20 kDa in size, were obtained from natural proteolysis of recombinant IF (50 kDa) and closely corresponded to the α- and β-domain of TC according to their N-terminal sequencing. Each fragment alone was able to bind Cbl with high affinity and a mixture of both fragments, provided the presence of Cbl, could be assembled into a ternary complex which behaved in a very similar way as uncleaved IF-Cbl regarding Cbl-binding and receptor recognition.

The finding [26] that truncation of 50 amino acids from the C-terminal end of human IF abolished its Cbl-binding ability is consistent with our IF model since this truncation leads to the loss of all H-bond interactions and most of the hydrophobic contacts between Cbl and the β-domain of IF.

All secondary structure elements of TC are present also in the models of IF and HC (Figure 1). Helices α3, α5 and α13 are shorter in IF than in the other proteins, while particularly extended loops are

found between helices α4 and α5 in TC, between α8 and α9 in HC and between strands β1 and β2 in IF. The flexible linker is three and four residues longer in IF and HC, respectively. Swapping of amino acids is observed between IF and HC in the hydrophobic core of the β-domain, involving the hydrophobic contact Leu^{325} – Phe^{370} in IF and Phe^{335} – Leu^{381} in HC. This concerted mutation is located at the same position as that between human and bovine TC which involves a leucine and a valine.

Disulfide bridges

Human and bovine TC show three disulfide bridges in the N-terminal α-domain. In human TC, these are Cys^{3} – Cys^{249}, Cys^{147} – Cys^{187} and Cys^{98} – Cys^{291}. Results from initial multiple sequence alignments indicated that in IF all three disulfide bridges are conserved while for HC, the conservation of only the first two bridges is obtained. The presence of the third bridge, however, appears possible since one of the involved Cys residues (Cys^{82}) is already conserved and the second (Cys^{285}) is shifted away from the conserved position by only one amino acid towards the C-terminus. An alignment featuring the conservation of all three disulfide bridges also in HC can thus be obtained by moving, both for TC and IF, one insertion gap from the nearby linker region to a position N-terminally to the involved Cys residue. A model of HC calculated with the third bridge shows excellent stereochemistry in the loops hosting the bridging Cys residues, i.e. the introduction of a restraint for this third bridge does not lead to distorted backbone geometry.

Apart from the three conserved disulfide bridges, an additional fourth bridge is found in the model of HC, where Cys^{365} and Cys^{370} connect the β-strands β4 and β5 in the C-terminal domain. No restraint for this bridge was present during modelling and initial refinement. The possibility of the presence of such a bridge was raised earlier based on a primary sequence analysis [27]. A multiple alignment of four known HC sequences (from man, dog, cow and hog) shows that the fourth disulfide bridge is a feature unique to the human form. In conclusion, the following disulfide bridges are present in our models: Cys^{8} – Cys^{228}, Cys^{125} – Cys^{164}, Cys^{85} – Cys^{270} in IF and Cys^{3} – Cys^{242}, Cys^{132} – Cys^{174}, Cys^{82} – Cys^{285}, Cys^{365} – Cys^{370} in HC.

Glycosylation sites

All glycosylation sites that were predicted on the basis of the Asn-X-Thr/Ser consensus sequence are found to be solvent-exposed in our comparative models (Figure 2). The IF model shows glycosylation sites exclusively on the β-domain, as was confirmed experimentally [24]. Asn^{293} was suggested to be glycosylated due to the inability to identify this particular residue during sequencing. Instead, HC also possesses two potential glycosylation sites on the α-domain which fall on the surface region recently proposed as receptor-recognition site of TC [10]. The presence of carbohydrates on the α-domain of HC, but not on TC, may therefore explain how cell surface receptors for TC discriminate between the two plasma Cbl-transporters TC and HC. Moreover, assuming that the receptor-recognition site is located on the α-domain alone [26], the lack of carbohydrates on the α-domain of IF may account for the observation that glycosylation does not effect receptor-binding of IF [28]. Recently, however, the presence of both domains was found to be required for binding of IF to its specific receptor cubilin [25].

Cbl environment in IF and HC

The lower side of Cbl (termed α-side) is occupied by the nucleotide moiety and is embedded in the protein (see discussion below). The upper axial side (β-side) contains in our models the ligand H_2O. This ligand can interact via H-bonds with the main chain oxygen of Pro^{149} in IF or Gln^{159} in HC. It possesses solvent accessibility which is, however, restricted by the C-terminal part of the loop between helices α7 and α8. The β-side's accessibility to solvent is in agreement with experimental data on kinetics for binding of Cbl with different upper ligands. The fact that very similar binding constants are obtained for Cbl with small ligands, such as CN or H_2O, and larger ligands such as 5'-deoxyadenosyl, is long considered as evidence that the β-side is at least partly solvent-exposed and thus does not play a role in determining ligand specificity [29]. As concluded in [1], the transporters show comparable affinities for the various Cbl forms and do generally not discriminate between Cbl with different upper axial ligands (Supplementary Table 2). A special feature in this respect (with unknown physiologic function) may be attributed to TC which treats H_2O-Cbl differently to CNCbl and N_3Cbl in that it replaces the H_2O ligand with a histidine side chain. Spectroscopic studies on the kinetics of ligand exchange at the upper axial position of H_2O-Cbl upon binding to any of the three transporters [8], revealed reactivity of Co with externally supplied histidine for IF- and HC-bound Cbl, but not for TC-bound Cbl. In fact, the recent X-ray structure revealed that upon binding of H_2O-Cbl to TC, the side chain of a histidine residue forms a coordination bond to the Co ion [10]. The Co-coordinating histidine in TC is not conserved in IF or HC as is visible from the sequence alignment in Figure 1 and consequently, H_2O remains coordinated.

Most of the polar and hydrophobic interactions between Cbl and TC are observed also in our comparative models of IF and HC (Table 3). Water-mediated polar interactions are not included in this comparison. IF misses the H-bonds to the oxygen atom of the corrin side chains *a* and *f* which are present in both TC and HC. Regarding hydrophobic contacts, IF and HC show identical interactions with an exception being the smaller Val^{351} in IF at the position of Tyr^{362} in HC and TC close to the 5,6-dimethylbenzimidazole (DMB) and an additional contact of a threonine C^{γ} atom in IF. TC differs from those hydrophobic interactions in three regions. First, the tryptophane close to DMB (Trp^{348} in IF, Trp^{359} in HC) is occupied by a Ser^{359} and three water molecules. Second, an additional contact exists at the pyrrole ring A of the corrin macrocycle and involves Met^{270} in human TC. Third, the phenylalanine side chain is absent at the upper side of the corrin macrocycle between ring C and D. All three transporters have in common the hydrophobic nature of the protein environment in van der Waals contact with the DMB moiety, as has early been concluded from the observation that CN^- is unable to displace the DMB from the Co ion in IF [30].

Compared to the solvent accessible surface of Cbl bound to TC (≈80 $Å^2$ or 6.6%), the Cbl ligand shows an accessibility twice as high (≈163 $Å^2$ or 13.3%) when bound to IF and half the accessibility when bound to HC (≈39 $Å^2$ or 3.2%). In IF, the 5'-hydroxyl group of the nucleotide moiety appears easily accessible to solvent water molecules. The corrin side chains *d, e* and *f* as well as an oxygen of the phosphate group possess restricted accessibility while side chain *a – c* are inaccessible. The same holds for HC where, in addition, side chain *f* is inaccessible.

Protein Surface Properties

As already found for human TC, no strong electrostatic potential is calculated from the models at the interface of Cbl and either domain of IF or HC, as it can be expected for binding of a neutral ligand (Figure 3). Regarding the domain contact area, a moderate degree of complementarity in the potential appears to exist only for IF which shows the unique feature of a slightly positively charged interface on the α-domain. Instead, the interface surfaces of HC and TC are neutral to slightly negative.

Figure 3 illustrates also the shape complementarity between the ligand and the two domains of the transporters. A depression in the otherwise flat interface of the α-domain to the β-domain accommodates about half of the ligand and a small portion of the β-domain which we identified as the turn between strands β3 and β4. In the holo-form of the transporters, this turn is in part wrapped around the nucleotide moiety of Cbl and appears trapped between the α-domain and Cbl. This spatial arrangement can hardly be achieved if Cbl was already bound to the α-domain. As a consequence, it seems necessary for a correct formation of the holo-form that the protein binds Cbl first with its β-domain. Subsequently, the α-domain can attach to the Cbl–β-domain complex and induces (if not already accomplished) the wrapping of the mentioned turn in the β-domain around Cbl's nucleotide moiety. In this latter step, a large amino acid is involved in IF (Trp^{348}) and HC (Trp^{359}), but only a small residue in TC (Ser^{359}). An additional indication for the sequence of Cbl-binding "β-domain before α-domain" is the observation for IF that Cbl binds to the isolated β-domain with higher affinity than to the isolated α-domain [25].

Ligand specificity of the three Cbl-transporters

An early study [9] demonstrated that the affinity of both IF and TC for 14 different Cbl analogues decreases upon modification of the corrin side chains or the nucleotide moiety. In contrast, HC tolerates major structural differences and even the total lack of the nucleotide moiety. A later study focused on the influence of modifications at the nucleotide moiety on ligand recognition [31]. These assays allowed to file Cbl-binding specificities in the order HC << TC < IF. The likely physiological role thereof implies that the most specific protein IF filters out various Cbl analogues before they pass to the plasma, whereas HC in the blood plasma acts as a scavenger of potentially toxic Cbl-analogues, thus exploiting its lowest level of discrimination among the three transporters.

Attempts to explain the observed differences in specificity have so far been unsuccessful. Earlier, a correlation between His-Co coordination on the upper side of the corrin ring (as is observed for TC) and high ligand selectivity was ruled out by comparative kinetic analysis [8]. The lack of correlation is now corroborated by our finding that the His-Co coordination is present neither in IF nor in HC. Analysis of conservations among the transporters' primary structure, even if supplemented by the detailed knowledge of Cbl interactions with TC [10], did not suffice to identify crucial residues for ligand specificity. This prompted us to produce and examine comparative models for IF and HC in order to address this issue on the level of tertiary structures.

In the following, our structural analysis of ligand specificity is based on observations from affinity measurements that employ a variety of Cbl analogues [9,31]. The affinity of a transport protein for a

given Cbl analogue is expressed as the ratio of the apparent affinity constant to that for CN-Cbl and is determined by an analogue's ability to competitively inhibit the binding of [^{57}Co]CN-Cbl [9]. The Supplementary Table 2 presents a compilation of investigated Cbl analogues and their affinity relative to CN-Cbl. Modifications can be grouped into those concerning the corrin side chains (*a* – *g*) and those concerning the nucleotide arm, in particular the DMB moiety.

The first three analogues in the Supplementary Table 2 involve the replacement of a neutral amide group with a negatively charged carboxyl group at one of the corrin side chains *b*, *d* or *e* (see Figure 4 and the on-line Multimedia adjuncts). Each of these replacements leads in all three protein models to the loss of H-bond interactions to main chain oxygen atoms (Table 3). However, the replacement shows little effect only concerning chain *e*. A difference between the modification at chain *b* or *d* on one hand and chain *e* on the other hand is the spatial freedom for the latter chain to adapt its conformation according to the introduced electrostatic repulsion while the *b*- and *d*-chain are tightly packed against the protein. Regarding the modification at chain *b*, the negative charge of the introduced O^- suffers from electrostatic repulsion to the backbone oxygens and additionally to two residues in IF (Glu^{379}, Asp^{383}), and to only one residue in TC (Asp^{393}) but no other residue in HC. Likewise, the electrostatic surface potential in the contact region of chain *d* is more negative for IF than for the other two transporters (Figure 3). This can explain the trend of increasing specificity in the row HC < TC < IF regarding these modifications.

Chain *d* (as also chain *c*) is fully surrounded by the β-domain alone, i.e. the domain proposed to first accommodate the ligand. Electrostatic repulsion of a modified Cbl will thus disturb the binding process at the earliest step. In contrast, chains *b* and *e* are located in the domain interface, and their repulsion will be fully effective only at the subsequent step of formation of the domain interface. This difference is expected to account for the lower affinity of the *d*-monocarboxylic Cbl-analogue with respect to the *b*- and *e*-forms. Analogue 4 combines the three modifications and confirms the mentioned trend in specificity. The swapping of chain *e* from below the corrin plane to above (analogue 5) is neither connected to significant loss of affinity nor differences in affinity among the three proteins and, as judged from our models, does not result in steric clash.

The second group of analogues in the Supplementary Table 2 involves modifications at the DMB moiety on the lower side ("α-side") of Cbl. This group can be divided into analogues which preserve the coordination bond between the Co ion and the nitrogen atom N3B of DMB (analogues 6-14 in "base-on" form) and those which do not (analogues 15-19 in "base-off" form). Some analogues were examined in both studies (analogues 7, 10, 12, 16). In the study by Kolhouse and Allen [9], affinity ratios are systematically lower than those obtained by Stupperich and Nexø [31]. These lower affinities might be due to contamination of protein samples with trace amounts of CN-Cbl or other CN-Cbl analogues, as stated by the authors [9]. Stupperich and Nexø concluded that modifications on the DMB moiety are tolerated by all binding proteins as long as the Co-DMB coordination is left intact [31]. Upon disruption of this coordination bond, both IF and, to a slightly smaller extent, TC show strongly decreased affinity, in contrast to HC which maintains most of its affinity. Generally, the size of the nucleotide's base had only minor effects on the affinity constants. These conclusions are in full agreement with structural

features of our protein models. Smaller modifications at the DMB can be accommodated by the proteins' environment around the nucleotide base (Figure 4 and the on-line Multimedia adjuncts). For base-off analogues, the models suggest that in the final holo-form of the transporter, the nucleotide moiety should approach the Co ion in a way similar to a base-on form in order to avoid major steric clashes.

A striking difference among the Cbl-transporters is the ability of HC to bind with high affinity even those analogues that lack the nucleotide moiety, such as cobinamide (Cbi). In all transporter models, the lower side of Cbl which hosts the nucleotide moiety is surrounded by a region built of the β–hairpin motif β3–turn–β4. It resembles an "arm" at one side of the β-domain as "body" (Figure 4A). HC shows the unique feature of three large residues in van der Waals contact with the nucleotide. Two aromatic side chains, Trp^{359} and Tyr^{362}, flank the DMB group and the apolar part of the Arg^{357} side chain packs against the ribose group (Figure 4B and the on-line Multimedia adjuncts). Trp^{359} is present in IF (Trp^{348}) but absent in TC (Ser^{359} is at this position). On the contrary, Tyr^{362} is present also in TC but absent in IF (Val^{351} is found instead). At the position of Arg^{357} there is a small polar residue both in IF (Thr^{346}) and in TC (Ser^{357}). These structural differences suggest that the space of the nucleotide moiety of Cbl can be filled upon binding of the nucleotide-lacking Cbi to HC by the side chains of Arg^{357}, Trp^{359} and Tyr^{362} without major main chain rearrangements (Figure 4C and the on-line Multimedia adjuncts). Appropriate side chain conformations of these three residues may enable HC to provide hydrophobic interactions towards the apolar lower side of Cbi and polar contacts towards the solvent. In contrast, IF and TC provide only one aromatic side chain and thus leave a large part of Cbi's apolar lower side accessible to solvent molecules. Major conformational rearrangements in the turn and adjacent amino acids of β-strands 3 and 4 might improve the interaction between Cbi and the hairpin, but remain in any case less optimal than in HC, as is evident from the binding assays. In addition, the positively charged arginine can balance the negative charge at the α-side of cobinamides, stabilizing this Cbl analogue in HC.

We predict that mutations in HC at one or several positions of the three mentioned residues significantly change the binding of Cbl-analogues. Replacing these large residues with smaller residues, such as serine or threonine as present in TC or IF, is expected to deteriorate the packing of the apolar lower side of Cbi to the hairpin motif. Likewise, mutations to render TC and IF equal to HC at the three residues should produce less selective forms of these two Cbl-transporters. In addition, the results of such a study may answer the question whether or not binding of Cbl-analogues without the nucleotide moiety involves structural rearrangements also outside the interface between the analogue and the protein, for instance, in regions determining the flexibility of the hairpin motif.

Binding of Cbl-bioconjugates to TC

In addition to the Cbl analogues compiled in the Supplementary Table 2, binding of Cbl bioconjugates has been studied in great detail for TC [12,32-34]. The effect that the attachment of a molecule to Cbl had on the interaction with TC was systematically investigated in competitive binding assays using [^{57}Co]-CN-Cbl, varying the type and location of the attached conjugate. Attachments at the corrin side chain *e* showed the smallest decrease in binding (about a factor of 2). Increasingly strong impairment of binding was observed for attachment on side chains in the row *b*, *d* and *c*. Biotin conjugates attached either to the

Co ion or the 5'-hydroxyl of the ribose moiety had little effect on binding [12]. These results are rationalized by the three-dimensional structure of human holo-TC [10], showing that sites of conjugation with minor impact on binding are indeed those which suffer less from steric problems with the surrounding protein, i.e. the β-axial position of Co, the 5'-hydroxyl and the side chain *e*. These locations can accommodate conjugates without restriction on size if an appropriate spacer moiety is inserted between Cbl and the conjugate [12,32,34].

For the use of Cbl derivatives as therapeutic or diagnostic agents, the correct binding of the derivative to TC is not the only requisite. It is of importance that a complex of TC with a Cbl-derivative shows cellular uptake and intracellular trafficking similar to that of the complex of TC with natural Cbl. McLean et al. investigated the ability of monomeric and dimeric Cbl derivatives to block the growth of leukemia cells [13]. They found that all those bioconjugates with modifications at the side chain *e* were good inhibitors of cell proliferation in that they act as competitors with natural Cbl for binding to TC and, at the same time, do not allow coenzyme activity. The constraint that Cbl conjugation should not alter too much the binding affinity to the transport protein holds additionally for IF and HC if the resulting derivative is intended for oral administration. Our comparative models suggest that the functional groups of Cbl can accommodate conjugates in a similar way when bound to IF or TC. In the case of HC, however, available space at the corrin side chains appears more restricted.

In conclusion, this study presented the first structural comparison of the three human Cbl-transport proteins, in particular of their interaction with Cbl and Cbl-analogues. This has permitted a rationalization of biochemical data in the literature about the difference in ligand specificity and the binding of bioconjugates. The protein models indicate the importance of the β3-turn-β4 hairpin motif in the smaller β-domain regarding high affinity of HC for Cbi. Three large amino acid side chains are proposed to supply hydrophobic contacts to the apolar side of Cbi, allowing HC, in contrast to IF and TC, to compensate for the lack of the nucleotide. The position of the hairpin's turn, trapped partially between the α-domain and the nucleotide moiety of Cbl, is suggested to require Cbl-binding first to the β-domain before the compact sandwich complex α-domain–Cbl–β-domain can be achieved. Polar interactions at the corrin side chains of Cbl are very similar in the three proteins; a finding that is consistent with their similar equilibrium dissociation constants. Many interactions of corrin side chains involve the protein backbone and their modification to carboxylic acids causes in general electrostatic repulsion from the tightly packed protein environment. Subtle differences in the proteins' structure may manifest themselves in the kinetics of binding to the different analogues but resolution of such features requires more detailed assays in addition to those assays relative to Cbl available to date. New experimental approaches have recently been presented in the literature [35].

ACKNOWLEDGMENTS

This work was supported by the Ministero dell'Istruzione dell'Università e della Ricerca, Rome (FIRB2003-RBNE03PX83).

REFERENCES

1. Nexø, E. (1998) Cobalamin binding proteins. In: Vitamin B_{12} and B_{12}-proteins (Kräutler, B., Arigoni, D. and Golding, T., eds), pp. 461-475, Wiley-VCH, Weinheim
2. Alpers, D. H. and Russell-Jones, G. J. (1999) Intrinsic Factor, Haptocorrin, and their Receptors. In: Chemistry and Biochemistry of Vitamin B_{12} (Banerjee, R., ed.), pp. 411-439, John Wiley & Sons, New York
3. Rothenberg, S. P., Quadros, E. V. and Regec, A. (1999) Transcobalamin II. In: Chemistry and Biochemistry of Vitamin B_{12} (Banerjee, R., ed.), pp. 441-473, John Wiley & Sons, New York
4. Banerjee, R. and Ragsdale, S. W. (2003) The many faces of vitamin B12: Catalysis by cobalamin-dependent enzymes. Annu. Rev. Biochem. **72**, 209-247
5. Quadros, E. V., Regec, A. L., Kahn, K. M., Quadros, E. and Rothenberg, S. P. (1999) Transcobalamin II synthesized in the intestinal villi facilitates transfer of cobalamin to the portal blood. Am. J. Physiol. **277**, G161-G166
6. Wolters, M., Strohle, A. and Hahn, A. (2004) Cobalamin: a critical vitamin in the elderly. Prev. Med. **39**, 1256-1266
7. Li, N., Seetharam, S. and Seetharam, B. (1995) Genomic structure of human transcobalamin II: Comparison to human intrinsic factor and transcobalamin I. Biochem. Biophys. Res. Commun. **208,** 756-764
8. Fedosov, S. N., Berglund, L., Fedosova, N. U., Nexø, E. and Petersen, T. E. (2002) Comparative analysis of cobalamin binding kinetics and ligand protection for intrinsic factor, transcobalamin, and haptocorrin. J. Biol. Chem. **277**, 9989-9996
9. Kolhouse, J. F. and Allen, R. H. (1977) Absorption, plasma transport, and cellular retention of cobalamin analogues in the rabbit. Evidence for the existence of multiple mechanisms that prevent the absorption and tissue dissemination of naturally occurring cobalamin analogues. J. Clin. Invest. **60**, 1381-1392
10. Wuerges, J., Garau, G., Geremia, S., Fedosov, S. N., Petersen, T. E. and Randaccio, L. (2006) Structural basis for mammalian vitamin B_{12} transport by transcobalamin. Proc. Natl. Acad. Sci. U.S.A. **103**, 4386-4391
11. Russell-Jones, G., McTavish, K., McEwan, J., Rice, J. and Nowotnik, D. (2004) Vitamin-mediated targeting as a potential mechanism to increase drug uptake by tumours. J. Inorg. Biochem. **98**, 1625-1633
12. Pathare, P. M., Wilbur, D. S., Heusser, S., Quadros, E. V., McLoughlin, P. and Morgan, A. C. (1996) Synthesis of cobalamin-biotin conjugates that vary in the position of cobalamin coupling. Evaluation of cobalamin derivative binding to transcobalamin II. Bioconjugate Chem. **7**, 217-232
13. McLean, G. R., Pathare, P. M., Wilbur, D. S., Morgan, A. C., Woodhouse, S. V., Schrader, J. W. and Ziltener, H. J. (1997) Cobalamin analogues modulate the growth of leukemia cells *in vitro*. Cancer Res. **57**, 4015-4022
14. Amagasaki, T., Green, R. and Jacobsen, D. W. (1990) Expression of Transcobalamin II Receptors by

Human Leukemia K562 and HL-60 Cells. Blood **76**, 1380-1386

15. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. **31**, 3497-3500
16. Sali, A. and Blundell, T. L. (1993) Comparative modelling by satisfaction of spatial restraints. J. Mol. Biol. **234**, 779-815
17. Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D **53**, 240-255
18. Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard M. (1991) Improved methods for building protein models to electron density maps and the location of errors in these models. Acta Crystallogr. A **47**, 110-119
19. Davis, I. W., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. **32**, W615-W619
20. Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283-291
21. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D **50**, 760-763
22. Holm, L. and Sander, C. (1996) Mapping the protein universe. Science **273**, 595-603
23. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. **98**, 10037-10041
24. Fedosov, S. N., Fedosova, N. U., Berglund, L., Moestrup, S. K., Nexø, E. and Petersen, T. E. (2004) Assembly of the intrinsic factor domains and oligomerization of the protein in the presence of cobalamin. Biochemistry **43**, 15095-15102
25. Fedosov, S. N., Fedosova, N. U., Berglund, L., Moestrup, S. K., Nexø, E. and Petersen, T. E. (2005) Composite organization of the cobalamin binding and cubilin recognition sites of intrinsic factor. Biochemistry **44**, 3604-3614
26. Tang, L. H., Chokshi, H., Hu, C. B., Gordon, M. M. and Alpers, D. H. (1992) The intrinsic factor (IF)-cobalamin receptor binding site is located in the amino-terminal portion of IF. J. Biol. Chem. **267**, 22982-22986
27. Fedosov, S. N., Berglund, L., Nexø, E. and Petersen, T. E. (1999) Sequence, S-S Bridges, and Spectra of Bovine Transcobalamin expressed in Pichia pastoris. J. Biol. Chem. **37**, 26015-26020
28. Gordon, M. M., Hu, C., Chokshi, H., Hewitt, J. E. and Alpers, D. H. (1991) Glycosylation is not required for ligand or receptor binding by expressed rat intrinsic factor. Am. J. Physiol. **260**, G736-G742
29. Gräsbeck, R. (1967) Intrinsic factor and the transcobalamins with reflections on the general function and evolution of soluble transport proteins. Scand. J. Clin. Lab. Invest. Suppl **95**, 7-18

30. Lien, E. L., Ellenbogen, L., Law, P. Y. and Wood J. M. (1974) Studies on the Mechanism of Cobalamin Binding to Hog Intrinsic Factor. J. Biol. Chem. **249**, 890-894
31. Stupperich, E. and Nexø, E. (1991) Effect of the cobalt-N coordination on the cobamide recognition by the human vitamin B12 binding proteins intrinsic factor, transcobalamin and haptocorrin. Eur. J. Biochem. **199**, 299-303
32. Pathare, P. M., Wilbur, D. S., Hamlin, D. K., Heusser, S., Quadros, E. V., McLoughlin, P. and Morgan, A. C. (1997) Synthesis of cobalamin dimmers using isophthalate cross-linking of corrin ring carboxylates and evaluation of their binding to transcobalamin II. Bioconjugate Chem. **8**, 161-172
33. Wilbur, D. S., Hamlin, D. K., Pathare, P. M., Heusser, S., Vessella, R. L., Buhler, K. R., Stray, J. E., Daniel, J., Quadros, E. V., McLoughlin, P. and Morgan, A. C. (1996) Synthesis and *nca*-radioiodination of arylstannyl-cobalamin conjugates. Evaluation of aryliodo-cobalamin conjugate binding to transcobalamin II and biodistribution in mice. Bioconjugate Chem. **7**, 461-474
34. Wilbur, D. S., Pathare, P. M., Hamlin, D. K., Rothenberg, S. P. and Quadros, E. V. (1999) Radioiodination of cyanocobalamin conjugates containing hydrophilic linkers: preparation of a radioiodinated cyanocobalamin monomer and two dimmers, and assessment of their binding with transcobalamin II. Bioconjugate Chem. **10**, 912-920
35. Fedosov, S. N., Grissom, C. B., Fedosova, N. U., Moestrup, S. K., Nexø, E. and Petersen, T. E. (2006) Application of a fluorescent cobalamin analogue for analysis of the binding kinetics. A study employing recombinant human transcobalamin and intrinsic factor. FEBS J. **273**, 4742-4753
36. Henikoff, S. and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U.S.A. **89**, 10915-10919
37. Anton, D. L., Hogenkamp, H. P. C., Walker, T. E. and Matwiyoff, N. A. (1980) Carbon-13 Nuclear Magnetic Resonance Studies of the Monocarboxylic Acids of Cyanocobalamin. Assignments of the *b*-, *d*-, and *e*-Monocarboxylic Acids. J. Am. Chem. Soc. **102,** 2215-2219
38. Nexø, E., Olesen, H. and Hansen, M. R. (1979) Strength of binding of methyl-, 5'-deoxyadenosyl-, cyano- and hydroxocobalamin to human transcobalamin I and II and intrinsic factor. In: Vitamin B_{12} (Zagalak, B. and Friedrich, W., eds.), pp. 851-854, Walter de Gruyter, Berlin, New York

TABLE 1. Properties of Cbl transporters in humans

Property	IF*	HC*	TC*
Chromosome	11	11	22
Source	mainly gastric mucosa	gastric mucosa, white blood cells, glands, etc.	vascular endothelium
Occurrence	stomach, small intestine	body fluids (mainly saliva, plasma)	body fluids (mainly plasma)
Function	Cbl-transport from proximal intestine to epithelial cells of the ileum filtering out Cbl-analogues (?)	salivary HC: binding of dietary HC in the stomach (protection of Cbl from acid hydrolysis and from uptake by intestinal fauna ?) plasma HC: Cbl recruitment from stores, scavenging of Cbl-analogues (?)	Cbl-transport from epithelial cells to tissue and organ cells Facilitation of Cbl-uptake into cells
Amino acid contents (M_W)	399 (43.4 kDa)	410 (45.6 kDa)	409 (45.5 kDa)
Carbohydrate contents (w/w)	9-15 %	30-40%	0
K_d (H_2O-Cbl)†	1 pM	0.01 pM	0.005 pM

* IF and HC data are taken from [2], TC data from [3,5].

† Equilibrium dissociation constants at pH 7.5 and 20°C taken from [8]

TABLE 2. Comparison of primary and tertiary structures of Cbl-transporters

The upper diagonal gives sequence identities (relative to the longer sequence in each pair) and, in parenthesis, sequence similarity ("positives" in the substitution matrix of Henikoff and Henikoff [36]). The lower diagonal reports rms deviations of Cα-atoms for equivalenced residues in the models superimposed with DALI [22]. The number of these residues is given in parenthesis.

	human IF	**human HC**	**human TC**	**bovine TC**
human IF	–	30% (48%)	23% (38%)	22% (38%)
human HC	1.3 Å (387)	–	27% (45%)	25% (43%)
human TC	1.6 Å (379)	1.5 Å (389)	–	73% (81%)
bovine TC	1.2 Å (381)	1.1 Å (395)	1.2 Å (409)	–

TABLE 3. Interactions between Cbl and its human transport proteins

Cbl atom	IF	HC	TC
Co (coordination bond)	–	–	His^{173} N^{ε}
H-bond Interactions (entries in parenthesis show weak H-bonding, i.e. N-O distance > 3.4 Å)			
side chain *a*: O28	–	Asn^{217} N^{δ}	Asn^{224} N^{δ}
N29	(Asp^{153} $O^{\delta1}$/$O^{\delta2}$)	(Asp^{163} $O^{\delta2}$)	Asp^{176} $O^{\delta2}$
	Asp^{204} $O^{\delta1}$/$O^{\delta2}$	Asn^{217} O^{δ}	Asn^{224} O^{δ}
side chain *b*: O33	–	–	–
N34	Asn^{378} O	Ser^{389} O	(Leu^{388} O)
side chain *c*: O39	Phe^{370} N	Leu^{381} N	Leu^{379} N
N40	Phe^{370} O	Leu^{381} O	Leu^{379} O
	(Leu^{377} O)	Leu^{388} O	Leu^{387} O
side chain *d*: O44	(Val^{352} N)	(Ile^{363} N)	(Leu^{363} N)
N45	Val^{352} O	Ile^{363} O	Leu^{363} O
	(Trp^{368} O)	(Trp^{379} O)	(Trp^{377} O)
side chain *e*: O51	Ser^{112} O^{γ}	Asn^{120} N^{δ}	–
N52	Ser^{112} O	Thr^{119} O	Thr^{134} O
side chain *f*: O58	–	(Gln^{123} N^{ε})	Gln^{138} N^{ε}
N59	–	–	–
side chain *g*: O62	(Gln^{252} N^{ε})	(Gln^{266} N^{ε})	(Gln^{273} N^{ε})
	Tyr^{115} O^{η}	(Tyr^{122} O^{η})	(Tyr^{137} O^{η})
N63	(Gln^{252} O^{ε})	Gln^{266} O^{ε}	Gln^{273} O^{ε}
	Asp^{153} $O^{\delta1}$	Asp^{163} $O^{\delta1}$	Asp^{176} $O^{\delta2}$
nucleotide arm: O4	His^{73} N^{ε}	Arg^{357} $N^{\eta1}$	Gln^{86} N^{ε}
O5	(Ser^{347} N)	Ser^{358} N	Leu^{358} N
Hydrophobic Interactions			
5,6-dimethylbenzimidazole	Thr^{346} C^{γ}	Arg^{357} C^{β}	–
	Trp^{348}	Trp^{359}	Ser^{359}
	Val^{351}	Tyr^{362}	Tyr^{362}
	Gly^{380}	Gly^{391}	Gly^{390}
nucleotide arm	Gly^{72}	Gly^{70}	Gly^{85}
	Leu^{76}	Leu^{74}	Leu^{89}
	Leu^{119}	Leu^{126}	Leu^{141}
methyl groups: ring A	–	–	Met^{270}
	Tyr^{206}	Phe^{219}	Tyr^{226}
ring C	Tyr^{367}	Tyr^{378}	Phe^{376}
	Tyr^{399}	Tyr^{410}	Trp^{409}
ring D	Tyr^{115}	Tyr^{122}	Tyr^{137}
	Phe^{150}	Phe^{160}	–

Figure 1. Sequence alignment of human and bovine TC with human IF and HC

This alignment together with the known three-dimensional structures of the two TC forms were used for comparative modelling of IF and HC. One letter amino acid codes in italics represent secondary structure elements in the α-domain and β-domain, respectively. α-helices (α) and β-strands (β) are numbered consecutively from the N- to the C-terminus. Residues participating in hydrophobic interactions with the ligand Cbl are highlighted in bold. Residues participating in H-bonds to Cbl are marked by the letter for the involved corrin side chain above the residue (those residues employing their side chain are additionally highlighted in grey). The His residue which coordinates to the Co ion of Cbl in human and bovine TC (after displacement of H_2O at the upper axial position of H_2O-Cbl) and Cys residues of disulfide bridges are marked by rectangular boxes. Potential Asn-linked glycosylation sites in IF and HC (based on the Asn-X-Thr/Ser consensus sequence) are underlined. See Supplementary Figure 1 for a colour version of this figure.

Figure 2. Comparative models of IF and HC

(**A**) Ribbon representation of the IF model showing secondary structure elements, the orientation of the ligand Cbl (magenta) with respect to the N- and C-terminal domain, the disulfide bridges (yellow sticks) and potential Asn-linked glycosylation sites (red). The part of the β–domain coloured in blue corresponds to the 50 C-terminal residues which were removed from the protein in the Cbl-binding studies of Tang *et al.* [26], see text. (**B**) Ribbon representation of HC using the colour scheme of panel A. (**C**) Stereo view of superimposed backbone traces of human Cbl-transporters. The comparative models of IF and HC are shown in green and blue, respectively, and the human TC structure as determined by X-ray crystallography is coloured orange.

Figure 3. Surface properties of Cbl transport proteins

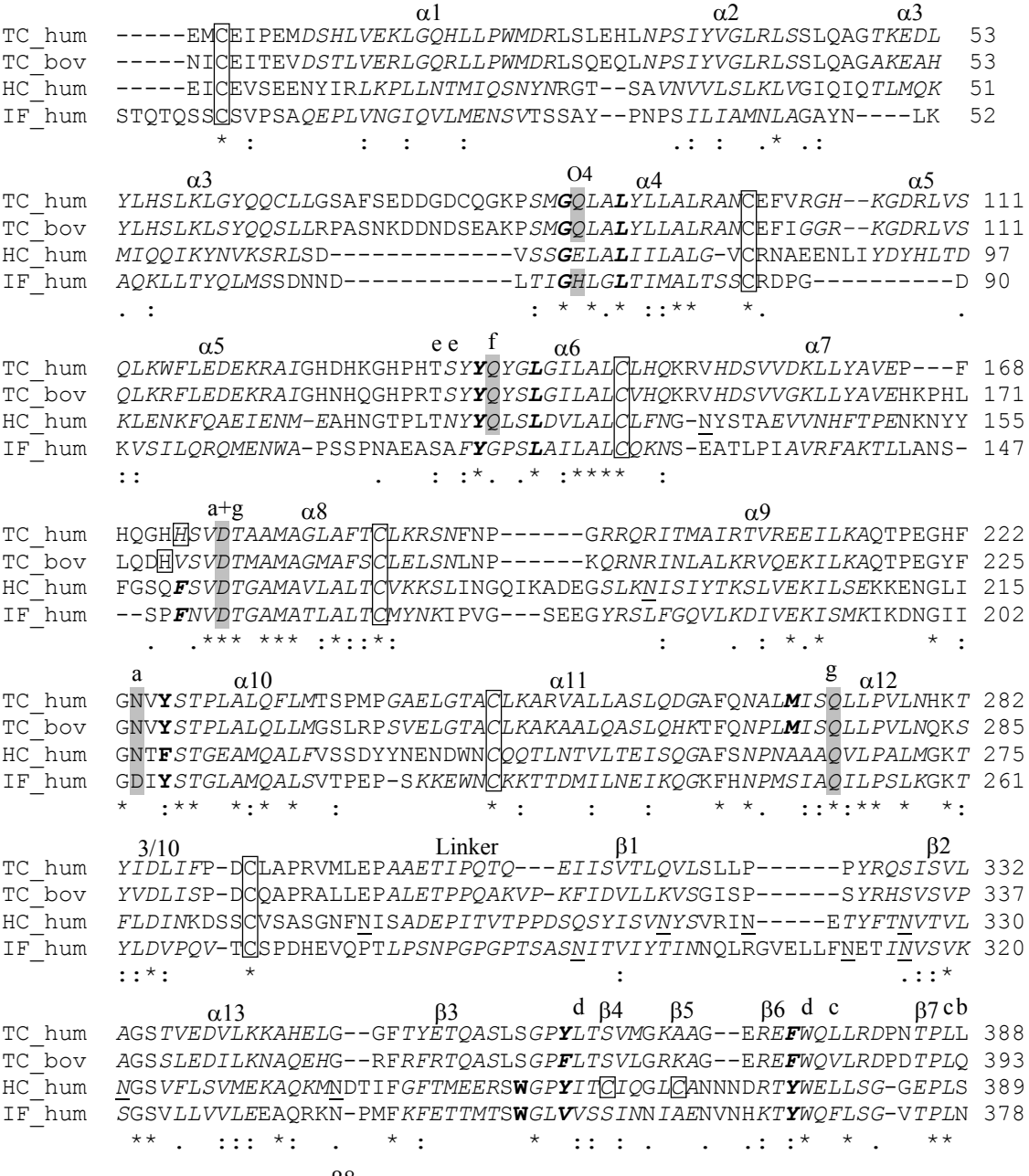
The electrostatic potential (blue, positive potential; red, negative potential) is mapped on the surface of each protein. The proteins are presented here in an artificial ”open” conformation in which the α-domain (left) is separated from the β-domain (right) to allow the view on their Cbl-hosting interface. The green line represents the linker region (Figure 1). A neutral or slightly negatively potential dominates on all interface regions with the exception of the α-domain of IF, thus possessing moderate charge complementarity. Shape complementarity with the ligand Cbl is much more significant; about half of Cbl’s surface is in contact with the α-domain (most of the upper β-side of the corrin ring and side chains *a* and *g*) and with the β-domain (most of the lower α-side of the corrin ring and side chains *c* and *d*). A region of the β-domain (bordered by a yellow line) wraps around part of the nucleotide moiety on Cbl’s α-side and fits into a depression on the α-domain (also bordered yellow). This region corresponds to the turn between β-strands β3 and β4 (see Figure 4A); in the holo-form structures, it is trapped between the α-domain and Cbl (bottom scheme). The scheme illustrates the wrapping of the turn around Cbl’s α-side at the step of complex formation with the α-domain but this conformational change may be induced

already upon binding of Cbl to the β-domain.

Figure 4. Structural analysis of protein-ligand interactions

(**A**) Overview of the ligand position in the transporter with an emphasis on the β–hairpin motif in the β–domain. The ligand (here H_2O-Cbl) is shown as magenta sticks, the β-strands 3 and 4 as blue ribbons and the rest of the protein with its surface in brown (α-domain) or yellow (β-domain). (**B**-**D**) Stereo views of superimposed structures of the transporters in a 5 Å neighbourhood of the ligand (carbon atoms of TC in green, IF in blue, HC in magenta). All labels refer to HC. (**B**) Interactions of H_2O-Cbl with the proteins. (**C**) Interactions of the nucleotide-lacking dicyanocobinamide (Cbl-analogue no. 16 in the Supplementary Table 2). Only HC can fill the space occupied by the nucleotide moiety of Cbl after appropriate changes in the side chain conformations of Arg^{357}, Trp^{359} and Tyr^{362}. (**D**) Environment of a cobamide (here NZA-Cba, Cbl-analogue no. 7) as an example of analogues with small modifications at the DMB moiety. For each panel of this figure, a Multimedia adjunct is provided in the Supplementary Material on-line to facilitate the visualization of the structural features.

Figure 1



Biochemical Journal Immediate Publication. Published on 5 Feb 2007 as manuscript BJ20061394

Figure 2

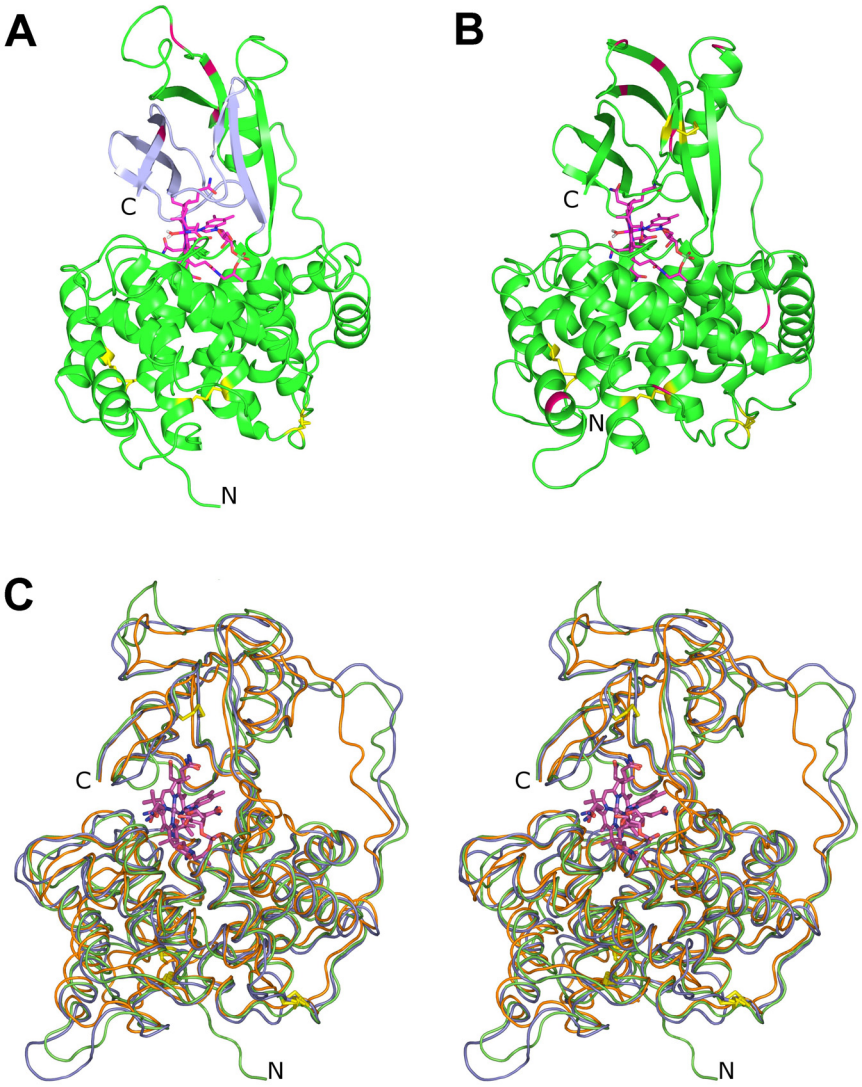


Figure 3

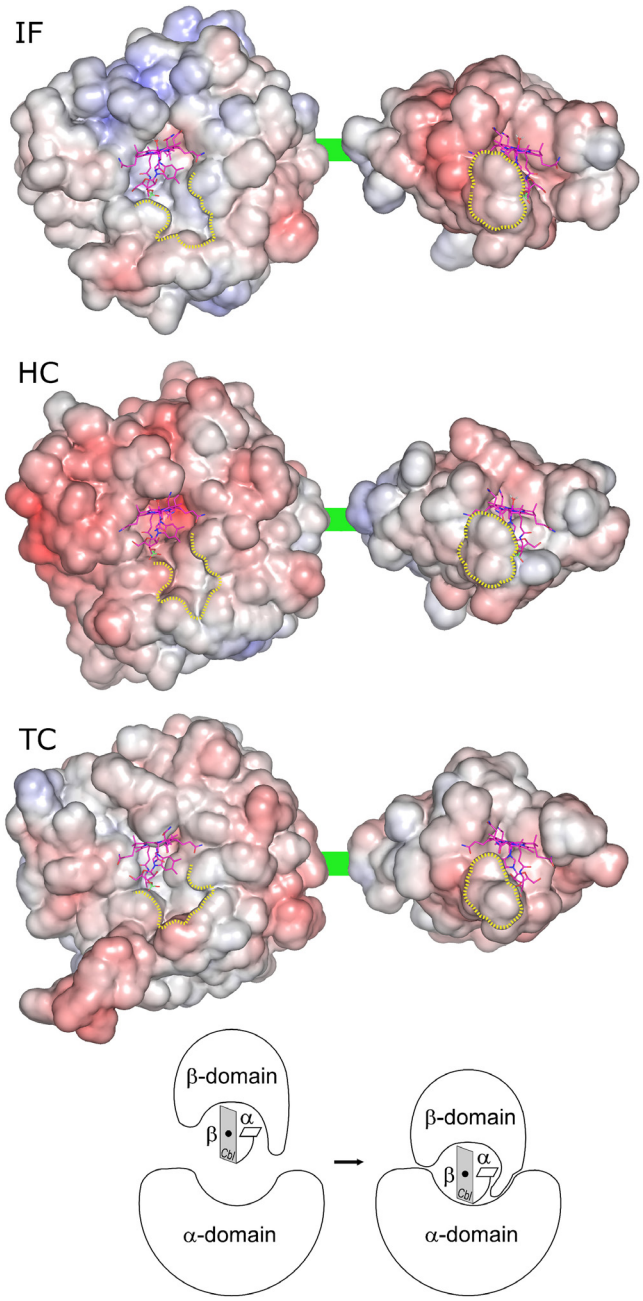


Figure 4