Computational and Structural Biotechnology Journal 19 (2021) 2891–2904

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

A combined evolutionary and structural approach to disclose the primary structural determinants essential for proneurotrophins biological functions

S. Covaceuszach [a,*], L.Y. Peche [a], P.V. Konarev [b], D. Lamba [a,c,*]

[a] Istituto di Cristallografia, Consiglio Nazionale delle Ricerche, Trieste, Italy
[b] A.V. Shubnikov Institute of Crystallography of Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Moscow, Russia
[c] Interuniversity Consortium "Biostructures and Biosystems National Institute", Roma, Italy

ARTICLE INFO

Article history:
Received 5 March 2021
Received in revised form 3 May 2021
Accepted 3 May 2021
Available online 13 May 2021

Keywords:
Proneurotrophin
Structural characterization
Evolutionary analysis

ABSTRACT

The neurotrophins, i.e., Nerve Growth Factor (NGF), Brain-Derived Neurotrophic Factor (BDNF), Neurotrophin 3 (NT3) and Neurotrophin 4 (NT4), are known to play a range of crucial functions in the developing and adult peripheral and central nervous systems. Initially synthesized as precursors, i.e., proneurotrophins (proNTs), that are cleaved to release C-terminal mature forms, they act through two types of receptors, the specific Trk receptors (Tropomyosin-related kinases) and the pan-neurotrophin receptor p75NTR, to initiate survival and differentiative responses.

Recently, all the proNTs but proNT4 have been demonstrated to be not just inactive precursors, but signaling ligands that mediate opposing actions in fundamental aspects of the nervous system with respect to the mature counterparts through dual-receptor complexes formation with a member of the VPS10 family and p75NTR. Despite the functional relevance, the molecular determinants underpinning the interactions between the pro-domains and their receptors are still elusive probably due to their intrinsically disordered nature.

Here we present an evolutionary approach coupled to an experimental study aiming to uncover the structural and dynamical basis of the biological function displayed by proNGF, proBDNF and proNT3 but missing in proNT4. A bioinformatic analysis allowed to elucidate the functional adaptability of the proNTs family in vertebrates, identifying conserved key structural features. The combined biochemical and SAXS experiments shed lights on the structure and dynamic behavior of the human proNTs in solution, giving insights on the evolutionary conserved structural motifs, essential for the multifaceted roles of proNTs in physiological as well as in pathological contexts.

1. Introduction

Neurotrophins (NTs) are known to play differential roles in the developing and adult nervous system and in non-neuronal organs by promoting differentiation, neuronal survival, synaptogenesis and modulating synaptic plasticity [1]. Only four NT genes are known in mammals i.e., Nerve Growth Factor (NGF), Brain-Derived Neurotrophic Factor (BDNF), Neurotrophin 3 (NT3) and Neurotrophin 4 (NT4). They exert their biological functions through two types of receptors, the specific Trk tyrosine kinases (Tropomyosin-related kinases) and the low-affinity pan-neurotrophin receptor p75NTR, which can bind all NTs with similar affinity [2,3].

NTs are initially synthesized as larger precursor proteins, proneurotrophins (proNTs), consisting of a N-terminal pro-domain and a C-terminal mature domain [4]. These precursor proteins, rapidly associate as noncovalent homodimers and can enzymatically be cleaved by means of intracellular proteases (furin and proconvertases, plasmin, tryptase and specific matrix metallo proteases [5]), in the Golgi and endoplasmatic reticulum to produce C-terminal mature NTs [6].

ProNTs were historically considered to be just inactive precursors, with the pro-domains only acting as an intramolecular chaperone and facilitating the folding of the mature domain and regulating the respective secretory pathways. Recent findings

* Corresponding authors at: Istituto di Cristallografia, Consiglio Nazionale delle Ricerche, Trieste, Italy.
E-mail addresses: sonia.covaceuszach@ic.cnr.it (S. Covaceuszach), doriano.lamba@ic.cnr.it (D. Lamba).

https://doi.org/10.1016/j.csbj.2021.05.007

demonstrated that all proNTs but proNT4 can be secreted to the extracellular space and are biologically active ligands [7], by exerting opposing actions in fundamental aspects of the nervous system (such as neuronal survival and synaptic plasticity) with respect to the mature counterparts [8]. ProNTs induce activation, at subnanomolar concentrations, of the apoptotic machinery [9–11] with subsequent cell death of specific neuronal populations; regulate axon guidance in neuronal development [12] and mediate synaptic plasticity, namely Long-Term Depression (LTD), in hippocampal neurons [13]. Besides their physiological apoptotic function, proNTs also play an important role in neuropathology. Indeed, apoptotic pathways activated by proNTs have been proven to be induced after injury and in several neurodegenerative disorders by promoting proNTs secretion to the extracellular space [14]. Likewise NTs, proNTs utilize dual-receptor complexes to mediate their functions. They form a ternary assembly composed of either a member of the VPS10 family, i.e., Sortilin, Sortilin related VPS10 Domain Containing Receptor 2 (SorCS2) or Receptor 3 (SorCS3) and p75NTR [9,15,16], to regulate synaptic plasticity and trigger apoptotic signaling. Therefore, the outcome of NT action depends on the kind of ligand that is secreted by the cells (proNT or mature NT), as well as on the interaction with distinct interplaying receptor complexes. Indeed, both proNTs and NTs are able to interact with VPS10p members, p75NTR and Trk, but in general proNTs preferentially bind to VPS10p members while mature NTs have higher affinity for p75NTR and Trk [9,12]. The only exception is NGF that interacts with SorCS3 tighter than proNGF [17]. Co-expression of SorCS2 (or Sortilin) and p75NTR allows simultaneous binding to proNGF, substantially enhancing the binding affinity and leading to the ternary complex formation required for signaling [9,10,12,15].

Crystallographic studies disclosed NGF and proNGF to be engaged with p75NTR via the same binding site on the mature domain of NGF with the only repositioning of a single loop [18–20]. Similarly, the interactions in the proNGF-SorC2 complex [21] occur via the mature domain. The binding site overlaps those present in the complexes with p75NTR and Trk, respectively. It is worth noting that the other member of VPS10p family, Sortilin, has been shown to specifically recognize the pro-domains of the three proNTs (proNGF, proBDNF, and proNT3) in the formation of the co-receptor complex with p75NTR. In this context, the structural and molecular determinants underpinning the interactions between the pro-domains and the proNTs receptors are still elusive. Indeed, in the previously discussed crystal structure of the complex between p75NTR and proNGF [18] the electron density corresponding to the entire pro-domain was poorly traceable, precluding to predict where it would interact on the receptor. This is likely ascribable to the intrinsically disordered nature of the pro-domain of proNTs thoroughly assessed for hproNGF and in part for hproBDNF by CD, FT-IR, fluorescence, chemical and thermal denaturation, H/D exchange, NMR and limited proteolysis studies [22–31].

Pre-computed disorder predictions of the four hproNTs as well as their disorder-related functions are available by searching the curated IDP database D^2P^2 [32].

The main feature of intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) is the partial or complete lack of a cooperative, folded structure under native conditions, due to the typical composition of their primary sequence enriched in charged, polar and structure-breaking amino acids and lacking bulky hydrophobic amino acids [33].

Consequently, under physiological conditions, every IDP molecule samples a range of conformations over time. Some conformers of the same molecule might resemble a random coil structure, while others display structured elements, even if the dynamic nature of IDPs and the inherent ensemble averaging of in-bulk investigations can make sparsely populated conformations difficult to be detected by traditional biophysical techniques. These so-called transient secondary structures can act as seeds that promote the formation of local specific tertiary conformations upon binding to their partners in a coupled folding and binding process [33–36]. Binding through unfolded or partially unfolded intermediates, though controversial, can provide a kinetic advantage through the "fly-casting" mechanism [37,38]. According to this mechanism, a dimensionality reduction occurs when the folding of a disordered protein is coupled with binding, thereby speeding up the search for specific targets. These molecular interactions can be not only transient and dynamic, but also permanent, enabling IDPs to modify the activity of the partner or to act as protein hubs in intracellular networks. Furthermore, the folding of IDPs upon contact with partners can also be incomplete or even absent in some cases (according to the so-called "fuzziness" phenomenon) [39]. In this framework, disorder in protein can be static, if the IDP adopts a few or a multitude of distinct stable conformations ("polymorphic" model), or dynamic, in the case of the regions that are still unstructured in the complex either neighboring, i.e., "flanking" model, or linking, i.e., "clamp" model, ordered binding regions. If the complex formation does not induce folding of the IDPs, then the bound state is left entirely disordered ("random" model). These specific features determine the functional promiscuity of IDPs under different conditions.

Furthermore, the analysis of structured states of apparently disordered proteins from genomic sequences highlighted the co-occurrences of evolutionary constraints of functional importance [40].

Interestingly, IDPs generally appear to evolve more rapidly, to be more permissive to different point mutations and to have higher amino acids deletion and insertion rates compared to ordered proteins [41]. In this regard, proteins can display different modes of how sequence conservation can relate to disorder as a functional feature [42]. While in some cases the IDPs' disorder itself is conserved, but not the sequence facilitating it (i.e., flexible disorder), other IDPs retain disorder together with a highly conserved sequence (i.e., constrained disorder). Thus, it is of primary interest to integrate structural and evolutionary analysis in order to reveal common mechanisms used by members of the proNTs family to deal with these fast rates of evolution and to gain insights into the structural and molecular determinants underpinning their functional adaptability.

Structural characterization of IDPs by conventional methods as X-ray crystallography and cryo-electron microscopy is rather challenging due to their extremely high flexibility in solution. Indeed, X-ray crystallography is not applicable because not only the intrinsic conformational heterogeneity of this class of proteins usually hampers crystallization, but also because, in the crystal structures of proteins containing IDRs, the corresponding electron density maps are poorly resolved. Therefore, high-resolution structural characterization of IDPs mainly relies on Nuclear Magnetic Resonance (NMR) spectroscopy that is especially well suited to probe structurally flexible systems and dynamics [43,44]. Unfortunately, the size of the functional proNTs homodimers hampers to exploit this technique.

Small Angle X-ray Scattering (SAXS) represents instead the experimental technique of choice to deal with the structural flexibility of IDRs [45]. It has already been proven successful to structurally characterize a member of the proNTs family, i.e., mouse proNGF [26,46]. This approach allows probing the size and shape of unstructured proteins and their complexes with receptors. Although its intrinsic low-resolution nature, SAXS analysis allows to quantitatively describe systems characterized by large-scale protein fluctuations and the presence of multiple species and/or conformations in solution [47–49], employing ensembles of conformers to describe the experimental data [50].

Here we present a computational biology approach coupled to an experimental study by SAXS aiming to uncover the structural and dynamical basis of the biological function displayed by proNGF, proBDNF and proNT3 but missing in proNT4. An *in silico* bioinformatic analysis, following an evolutionary approach recently implemented [51], allowed to elucidate the functional adaptability and to identify key structural features of the members of the proNTs family, by predicting their structural disorder and secondary structure propensities and computing the rates of evolution per site and the rates of structural properties transitions.

A combined experimental approach was then exploited to shed lights on the structure and the dynamic behavior of the human members of proNTs family in solution. We report on complementary biochemical and SAXS experiments aiming to uncover the molecular mechanisms underpinning the biological functions conserved in proNGF, proBDNF and proNT3, but lost in proNT4 during evolution.

2. Materials and methods

2.1. Bioinformatics

The intrinsic disorder property within the amino acid sequence of the pro-domains of the four hproNTs, was characterized by using a set of commonly per-residue disorder predictors, such as MFDp2 [52], the IUPred suite for the prediction of short and long intrinsically disordered regions, respectively [53], PONDR VLXT [54], PONDR VSL2 [55], PONDR VL3 [56] and PONDR FIT [57]. Similarly, the results of different secondary structure predictors (i.e., Psipred [58], Jpred4 [59], PredictProtein [60] and RaptorX SS3 [61]) were compared by analyzing the amino acid sequences of the pro-domains of the four hproNTs.

Vertebrate sequences were collected by searching the NCBI database, using NCBI BLAST [62] by the Blastp (protein–protein BLAST) algorithm. The protein sequences of the four human proNTs, i.e., hproNGF: P01138; hproBDNF: P23560; hproNT3: P20783; hproNT4: P34130 [63] were used as query. A filter for high sequence coverage was applied and for each of the species only the sequences with the highest BLAST Max score were selected to minimize redundancy. In the case of high sequences abundance of a single taxon, i.e., Mammalia and Aves, sequences have been further filtered in order to equally cover all the orders of each taxa. For Multiple Sequence Alignment (MSA), datasets of the whole vertebrate sequences and of the representative organism were constructed for each of the proNTs. The species included in the representative datasets were: *Homo sapiens*, *Pongo abelii*, *Ornithorhynchus anatinus*, *Monodelphis domestica*, *Mus musculus*, *Columba livia* and *Gavia stellata* (for proNGF, proBDNF and proNT3); *Gekko japonicus, Podarcis muralis, Nanorana parkeri, Rhinatrema bivittatum, Amblyraja radiata* and *Rhincodon typus* (for proBDNF and proNT3 and proNT4); *Latimeria chalumnae* and *Paramormyrops kingsleyae* (for proBDNF and proNT3 and proNT4); and *Anarrhichthys ocellatus* (for proNGF, proNT3 and proNT4). All datasets were aligned using Clustal Omega [64] and the tree building tool was employed to construct the corresponding phylogenetic trees. Structural disorder propensity of all unaligned vertebrate datasets was predicted using MFDp2 (Multilayered Fusion-based Disorder predictor v. 2.00), which allows to process simultaneously up to 100 protein sequences [52]. It combines per-residue disorder probabilities predicted by MFDp with per-sequence disorder content predicted by DisCon [65]. Novel post-processing filters are applied to provide disorder predictions with improved predictive quality. The order–disorder propensity scores were mapped onto the MSA and taxon position in the tree to be visualized as a heat map and were further converted into a binary matrix to analyze the evolutionary dynamics of structural order–disorder (i.e., 1 for order and 0 for disorder). Secondary structure propensity was predicted for all sequences in each dataset using Jpred 4 [59], the latest version of the popular JPred protein secondary structure prediction server, which provides predictions by the JNet algorithm, one of the most accurate methods for secondary structure prediction, being its secondary structure prediction accuracy reported to be 82%. As for the analysis of the order–disorder propensity, the results of the secondary structure propensity were mapped to be visualized in a heat map and then converted into a binary matrix (i.e., 1 for helix and strand, and 0 for loop), to examine the evolutionary dynamics of secondary structure-loop transitions. The substitution rates of amino acids for each vertebrate dataset alignment were analyzed using ConSurf [66], a bioinformatics tool that applies statistical inference methods, machine learning and multiple sequence alignment to attribute a conservation score to each residue based on the phylogenetic relationship between the protein and its homologous sequences.

The evolutionary dynamics of structural order–disorder and predicted secondary structures were analyzed by calculating the evolution rates based on the phylogenetic trees and the binary matrices. The Gain Loss Mapping Engine (GLOOME) [67] was used in order to study the gain/loss transition events of the structural property by adopting the phyletic patterns. The gaps were treated as missing data and the outputs of evolution rates were standardized as Z-score.

2.2. DNA constructs

The cDNA encoding for human BDNF (hBDNF), human NT3 (hNT3) and human NT4 (hNT4) were purchased from Harvard Medical School PlasmID Database (clone_Ids: HsCD00005889, HsCD00001730 and HsCD00000988) corresponding to *Homo sapiens brain-derived neurotrophic factor (*NCBI Gene ID: 627*), Homo sapiens neurotrophin 3* (NCBI Gene ID: 4908) *and Homo sapiens neurotrophin 4* (NCBI Gene ID: 4909), respectively.

hBDNF, hNT3 and hNT4 in fusion with the corresponding N-terminal pro-domains (hproBDNF, hproNT3 and hproNT4) without the secretion leader sequence, corresponding to codons 57–741 of hpreproBDNF cDNA, 57–771 of hpreproNT3 cDNA and 75–630 of hpreproNT4 cDNA were amplified by PCR using the following forward and reverse primers, respectively:

5′- CTTTAAGAAGGAGATATACATATGGCCCCCATGAAAGAAGCAAACATC

3′-GGCTTTGTTAGCAGCCGGATCTCATCTTCCCCTTTTAATGGTCAATGTAC

for hproBDNF

5′- CTTTAAGAAGGAGATATACATATGAACAACATGGATCAAAGGAGTTTG

3′- GGCTTTGTTAGCAGCCGGATCTCATGTTCTTCCGATTTTTCTCGAC

for hproNT3

5′- CTTTAAGAAGGAGATATACATATGCAACCCCCACCCTCAACATTGC

3′- GGCTTTGTTAGCAGCCGGATCTCAGGCCCGGCCAGTCCG

for hproNT4.

They were cloned in pET22b vector by Restriction free Cloning [68], removing the bacterial secretion leader, the polylinker and the C-terminal His Tag. After screening by colony PCR, the positive clones were verified by automated DNA sequencing.

All the sequenced clones for hproBDNF carried the V66M polymorphism that was indeed present in the purchased plasmid. The sequence encoding the wild type protein was generated from the hproBDNF V66M construct by retro mutation (M66V) according the Stratagene site-directed mutagenesis protocol.

The following forward and reverse primers were employed respectively:

5′-GCTGACACTTTCGAACACGTGATAGAAGAGCTGTTGG

3′- CCAACAGCTCTTCTATCACGTGTTCGAAAGTGTCAGC.

The construct was verified by automated DNA sequencing.

2.3. Protein expression, refolding and purification

Human proNTs expression and purification were performed based on the method previously described for hproNGF [69]. pET22b constructs encoding for the proNTs were transformed into Rosetta (DE3) strain (Stratagene) and protein expression was induced at OD(600 nm) = 1.0 by 1 mM IPTG overnight at 30 °C. Pulsed refolding of hproNT3 was performed as previously reported for hproNGF, while in the case of hproBDNF and hproNT4 the refolding buffer was further optimized and 750 mM Arginine, 100 mM Tris pH 8.5, 5 mM EDTA, 5 mM reduced glutathione, 2 mM oxidized glutathione, 0.05% (v/v) PMSF was employed. After extensive dialysis against 50 mM Sodium Phosphate pH 7.0, 1 mM EDTA and 0.05% (v/v) PMSF at 4 °C, the proteins were purified by cationic exchange chromatography on HiTrap SP column (GE Healthcare) and concentrated by stirred ultrafiltration cells (Millipore).

2.4. Differential Scanning Fluorimetry (DSF)

DSF experiments were performed using a CFX96 Touch Biorad Real-Time PCR system (Bio-Rad) with λ_{ex} = 470–505 nm and λ_{em} = 540–700 nm. The proNTs stocks in 50 mM Sodium Phosphate buffer pH 7.0 were mixed at final concentration of 0.5 mg/mL with SYPRO Orange (Sigma) at final concentration of 90×. Fluorescence was measured in a 20°–90 °C temperature range with increments of 0.2 °C/min. The melting temperatures (T_m), represented by the inflection points of the transition curves, were calculated using Boltzmann Sigmoid fit. Experiments were performed in triplicate.

2.5. Small-angle X-ray scattering data collection and data processing

SAXS measurements on 5 different concentrations (the ranges are reported in Table S1) of the four hproNTs in 50 mM Sodium Phosphate pH 7.0 were performed on the P12 beamline EMBL SAXS-WAXS (Hamburg, Germany at PETRAIII/DESY) [70]. Data were collected with 20x 0.05 sec exposures using a Pilatus 2 M pixel X-ray detector, sample-detector distance 3.00 m, wavelength 1.24 Å. No radiation damage was detected, comparing scattering profiles for the collected frames.

Data were processed employing the ATSAS 2.6.0 program package [71]. In detail, normalization to the intensity of the transmitted beam, averaging of the frames for each sample and buffer subtraction were performed by PRIMUS [72]. Since for the hproBDNF sample some inter-particle repulsion effect was observed at increased protein concentration, the low s-data of diluted samples (0.41 mg/ml), where inter-particle interactions did not occur, were merged with the high s-data of the concentrated samples.

Guinier approximation [73] was applied to calculate the radius of gyration Rg and the forward scattering I(0), assuming that at very small angles ($0 < s < 1.3/R_g$), the intensity $I(s) = I(0) \exp(-1/3(R_g s)^2)$. GNOM [74] was used to compute the maximum sizes D_{max} and the pair distance distribution functions p(r). Molecular masses (MM) were estimated by comparison of the calculated forward scattering I(0) of the samples with that of the standard solution of bovine serum albumin (MM 66 kDa). The Porod approximation [75]:

$$V_p = \frac{2\pi^2 I(0)}{\int I_{exp}(s) s^2 ds}$$

was used to calculate the excluded volume of the hydrated protein molecule (V_p).

The web-based server tool SWISS-MODEL [76] was exploited to build a 3D homology model of mature hBDNF dimer using the high-resolution X-ray crystal structure of the hBDNF/hNT3 heterodimer (PDB_ID 1BND) [77] as a template. The homodimeric assembly was subjected to 4000 steps of steepest-descent *in vacuo* energy minimization, followed by 10,000 steps of conjugate gradient *in vacuo* energy minimization using GROMACS 4.5 [78].

Rigid-body modeling was performed by using CORAL [79] in order to generate the approximate conformations of the missing regions of the propeptides. The crystallographic structures of hNT3 (PDB_ID 1NT3) [80] and hNT4 (PDB_ID 1HCF) [81] and of the obtained high resolution 3D models of hNGF [82] and hBDNF, have been kept fixed and either P1 or P2 symmetry have been imposed. Ten independent CORAL runs were performed for each data set. The resulting models were superimposed by using the program SUPCOMB [83] and averaged by using DAMAVER [84] to identify the most typical models representing the human proNTs conformation in solution. The similarity of the resulting models was estimated by DAMAVER calculating the normalized spatial discrepancy parameter (NSD) [83]. NSD values $\leq$ 1.0 are expected for similar models.

The Ensemble Optimization Method (EOM) [85] allowed to quantitatively assess propeptide flexibility and size distribution of possible conformers. Randomized 10,000 conformations were initially generated for the hproNTs based on amino acid sequences and the crystallographic structures of hNT3 and hNT4 and the 3D-structural models of hNGF and hBDNF, used as rigid bodies, imposing the native conformation option and applying P2 symmetry. The scattering profiles of these pools of randomly generated conformations, computed by CRYSOL [86], were compared and representative structures, whose scattering curve fits to the experimental scattering curve, were selected by a genetic algorithm.

3. Results

3.1. The bioinformatic analysis of the evolution of structural order-disorder and secondary structures propensities highlights a divergent behavior of proNT4 with respect to the other members of the proNTs family

The comparison of the pro-domains among the human members of the proNTs family highlights a high variability both in their length and in their sequences (Fig. 1a), being the pro-domain of hproNT4 the most divergent.

Indeed, its length is almost half of the other members and it shares a lower sequence similarity. It displays a lower mean percentage of pair sequence identity (i.e., 17.7 ± 2.0% instead of 23.7 ± 3.6%) and almost halved mean percentage of pair sequence homology (i.e., 24.2 ± 1.4% instead of 43.0 ± 3.1%) with respect to the others hproNTs. The mature domains of hproNTs, instead, show a lower degree of variability. hNT4 displays very similar and much higher mean percentage of pair sequence identity (i.e., 52.2 ± 2.6%) and homology (i.e., 73.2 ± 3.7%) if compared to the others hNTs (55.0 ± 2.5% and 71.8 ± 4.4%, respectively). In analyzing the sequence similarity within this family, it deserves attention a main structural feature, at present reported only for proNGF and proBDNF [26,29], i.e., the intrinsic disorder of the pro-domains. It is known that IDRs evolve more rapidly than globular folded domains. The significant differences in sequence conservation between the pro-domains and the mature regions within hproNT family prompt to the intrinsic disorder to be a common feature among all the pro-domains. Thus, by taking into account that disorder is a fundamental property of the proteome, robustly

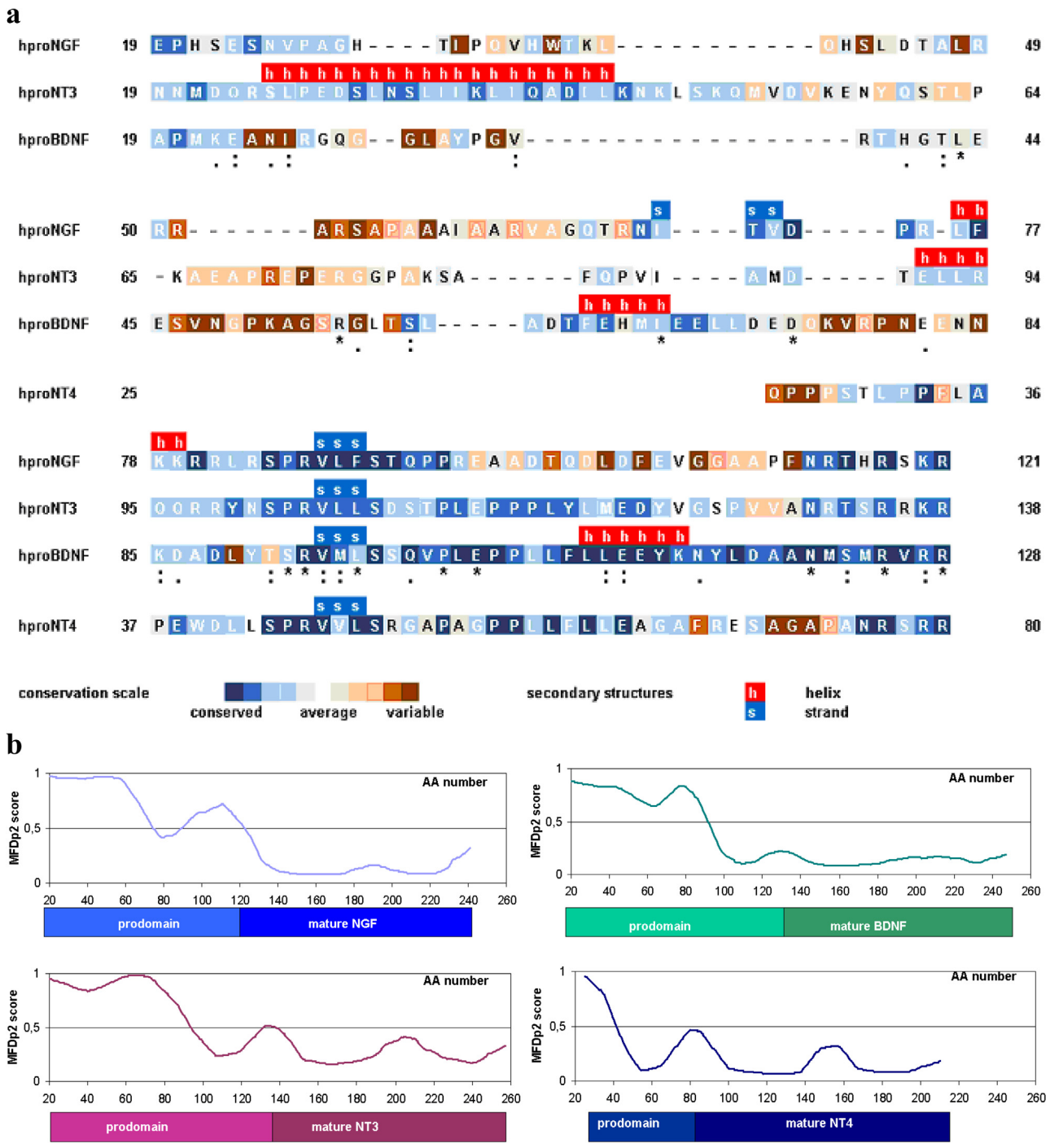

Fig. 1. Bioinformatic comparison of the primary sequences of the hproNTs: (a) multiple alignment of the primary sequences of the corresponding pro-domains. Each residue is colored according the conservation scores based on the phylogenetic relationship between the pro-domains of hNTs and their homologous sequences calculated by ConSurf [66]. Above each sequence, the secondary structure propensity based on the Jpred4 (Drozdetskiy et al. 2015) prediction score (i.e., red for helix and blue for strand) is reported. (b) MFDp2 predictions of intrinsically disordered regions [59]; below each graph the schematic representation of the corresponding amino-acid sequence is shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predictable from primary sequence relying on characteristic patterns of amino acids distribution and overall amino acids content [87], we extended the analysis of the disorder content to the other two members of the proNTs family (Fig. 1b). We assessed a set of commonly used per-residue disorder predictors, selected on their specific sensitivities to different features associated with intrinsic disorder (Supp. Fig. S1). Almost all the predictors concur with the reported intrinsically disordered nature of the pro-domain of hproBDNF and point to a very low disorder content in the pro-domain of hproNT4, mainly confined at the first 10 residues. On the contrary, a wide variability in the degree of disorder has been observed for the pro-domains of hproNGF and hproNT3 (Supp. Fig. S1a and c). Considering that the intrisic disorder of the pro-domain of hproNGF and hproBDNF has been experimentally characterized [26,29,31], the most accurate predictor resulted to be MFDp2 [52], (Fig. 1B), which allows to quantitatively predict the disorder content by means of DisCon [65], a predictor that takes advantage of a set of descriptors that aggregate and hybridize information originated from sequence, evolutionary profiles, and secondary structure prediction, solvent accessibility, flexibility,

and annotation of globular domains. Indeed, while the pro-domain of hproNT3 is estimated to be mainly disordered having a higher disorder content than the other two biologically active hproNTs (being the DisCon index of 0.379, 0.297 and 0.414 for hproNGF, hproBDNF and hproNT3, respectively), hproNT4 is characterized by a significantly lower disorder content (being the DisCon index of 0.104). Moreover, the comparison of the evolutionary conservation of the amino acid sequences (Fig. 1a) suggests that the pro-domains of hproNGF, hproBDNF and hproNT3 are characterized by constrained disorder, especially at the level of the regions that are predicted to be involved in secondary structure formation, according to the consensus obtained employing three different secondary structure predictors (Supp. Fig. S2).

To this regard, despite their designation as "disordered", IDPs do not generally behave as structure-less random polymer chains and frequently display residual secondary structures. Indeed, although the high degree of intrinsic disorder of the pro-domains, hproNGF, hproBDNF and hproNT3 are characterized by both the presence of helical content, that is absent in hproNT4, and by a conserved short beta strand that, in turn, iscommon to the pro-domains of all the members of the family.

In order to further investigate the relationship between rates of amino acid substitution in the pro-domains and their structural and functional properties across evolution, we calculated the conservation Z-scores for each position, being positives the scores for those sites evolving faster than average and negative the scores for those sites evolving slower than average (upper graphs, Fig. 2).

Moreover, to elucidate the changes in protein dynamics/structure of the pro-domains that are likely to be highly related to their biological function, we focused on the evolutionary patterns of structural properties of the proNTs members in vertebrates. The heat maps of secondary structure propensity and order–disorder structure propensity (Supp. Figs. S3 and S4) have been mapped to the MSA, ordered by the taxa position in the phylogenetic trees. The dynamic evolution of those structural properties has been analyzed by calculating the corresponding transition rates (lower graphs, Fig. 2). Despite each of the pro-domains is characterized by a specific evolutionary pattern of the structural properties, common features as well as diverging behaviors can be clearly identified.

Although the four pro-domains are very divergent both in length and in sequence, the secondary structure prediction results showed that all the pro-domains share a highly conserved beta strand (Supp. Fig. S3). The fact that this short secondary structure element, ranging between 2 and 4 residues, is extremely conserved across evolution with very similar transition patterns and rates (Fig. 2) in the proNTs superfamily, might suggest that it could act as an essential seed, mediating pro-domain biological functions associated with NT folding and maturation. An additional conserved secondary structure helix element is found in all the pro-domains but not in proNT4. Besides showing a highly conserved predicted long helical structural element (ranging 12 to 25 residues), located close to the amino-terminal, proNT3 shares with proNGF an homolog region (ranging 3 to 10 amino acids) having helix propensity, which is highly conserved across evolution. A small, but distinct, difference in the transition patterns and rates (Supp. Fig. S3a and c and Fig. 2a and c, respectively) has been detected. In addition, proBDNF shows conserved helix secondary structure propensity mapped in two regions (Supp. Fig. S3b), whose transition patterns and rates (Fig. 2b) are very similar to the helical structure shared by proNGF and proNT3. Since Sortilin recognizes a binding site encompassing amino acids 44–103 of proBDNF [88], the amino-terminal alpha helix element might then be regarded as the structural seed involved in this interaction, despite differences in the amino acid composition if compared to the corresponding alpha helical segment shared by proNGF and proNT3. Nevertheless, a typical hallmark of IDPs is their tolerance to mutations across evolution of paralogs allowing the biological function to be preserved. It is known that proBDNF derives from a different intermediate neurotrophin ancestral gene with respect to the one that originated proNGF and proNT3 [89], so explaining the basis for the divergence of the primary sequence with respect to the other biologically active proNTs.

Interestingly, the pro-domain of proNT4 is characterized by an apparent distinct behavior with respect to all the other members. Indeed, no conserved pattern of helical secondary structure could be identified in proNT4. A helical structure propensity is only present in a number of mammalian orders, but is almost completely absent in primates (Supp. Fig. S3d). This divergent behavior of proNT4 from the other members of the family might suggest that the very conserved helical structure could be involved in proNT specific biological functions associated with cell death and LTD, which on the contrary are not mediated by proNT4.

The comparison of the order–disorder propensity in the proNT family seems to corroborate this divergent behavior of proNT4. Indeed, while the proNGF pro-domain shows a conserved propensity of structural disorder, excepting very few order propensities being observed in some mammalians (Supp. Fig. S4a), opposed to proNT4, proBDNF and proNT3, it shares a specific order–disorder propensity pattern through evolution in this domain. Indeed, both proBDNF and proNT3 pro-domains are characterized by a major transition to disorder from amphibia and reptiles to birds, non-placental mammals, mammals and primates (Supp. Fig. S4b and c). On the contrary, proNT4 shows an opposed behavior with a major transition to order increasing from amphibia and reptiles to mammals and primates (Supp. Fig. S4d).

This finding is further supported by comparing the distributions of the rates of evolution in each of the pro-domains (upper graphs, Fig. 2). It is worth noting that conserved shared features that are related to the biological function are considered more reliable for slowing amino acid substitution rates in disordered regions, which instead generally have a faster rate of evolution. To this regard, all the proNTs pro-domains except proNT4 are characterized by a disorder propensity. Moreover, besides proNT3 that shows the lower rate of amino acid substitution (Fig. 2c) along almost the whole pro-domain sequence, the regions of the lowest rate of evolution in the pro-domains of proNGF and proBDNF are coincident with amino acid sequences with secondary structure propensity (Fig. 2a and b). The fact that, most of the sequence of the pro-domain of proNT4 (but the amino and the carboxy terminals) is characterized by a low rate of evolution and by the absence of helical structure propensity prompts to the alpha helical region present in proNGF, proNT3 and proBDNF as a key structural determinant in mediating pro-domains biological functions.

3.2. hproNT4 shows a higher conformational stability compared to the other members of hproNTs family

The preliminary part of the biophysical characterization dealt with the expression and purification of the hproNTs. The obtained yields for hproBDNF were similar to those for hproNGF (10 mg/L culture), but despite any further optimization the yields for hproNT3 and hproNT4 were ten times lower (1–1.5 mg/L culture). Nevertheless, the purity of the resulting samples was > 95% (Supp. Fig. S5). The absence of aggregation was checked by dynamic light scattering experiments (data not shown), being the % of polydispersity 15, 13 and 16 for hproBDNF, hproNT3 and hproNT4, respectively. All these data confirmed the suitability of the samples for the subsequent biochemical and SAXS experiments.

In details, Differential Scanning Fluorimetry (DSF) spectroscopy was exploited in order to compare the thermal stability of the hproNTs. Indeed, monitoring protein thermal denaturation by

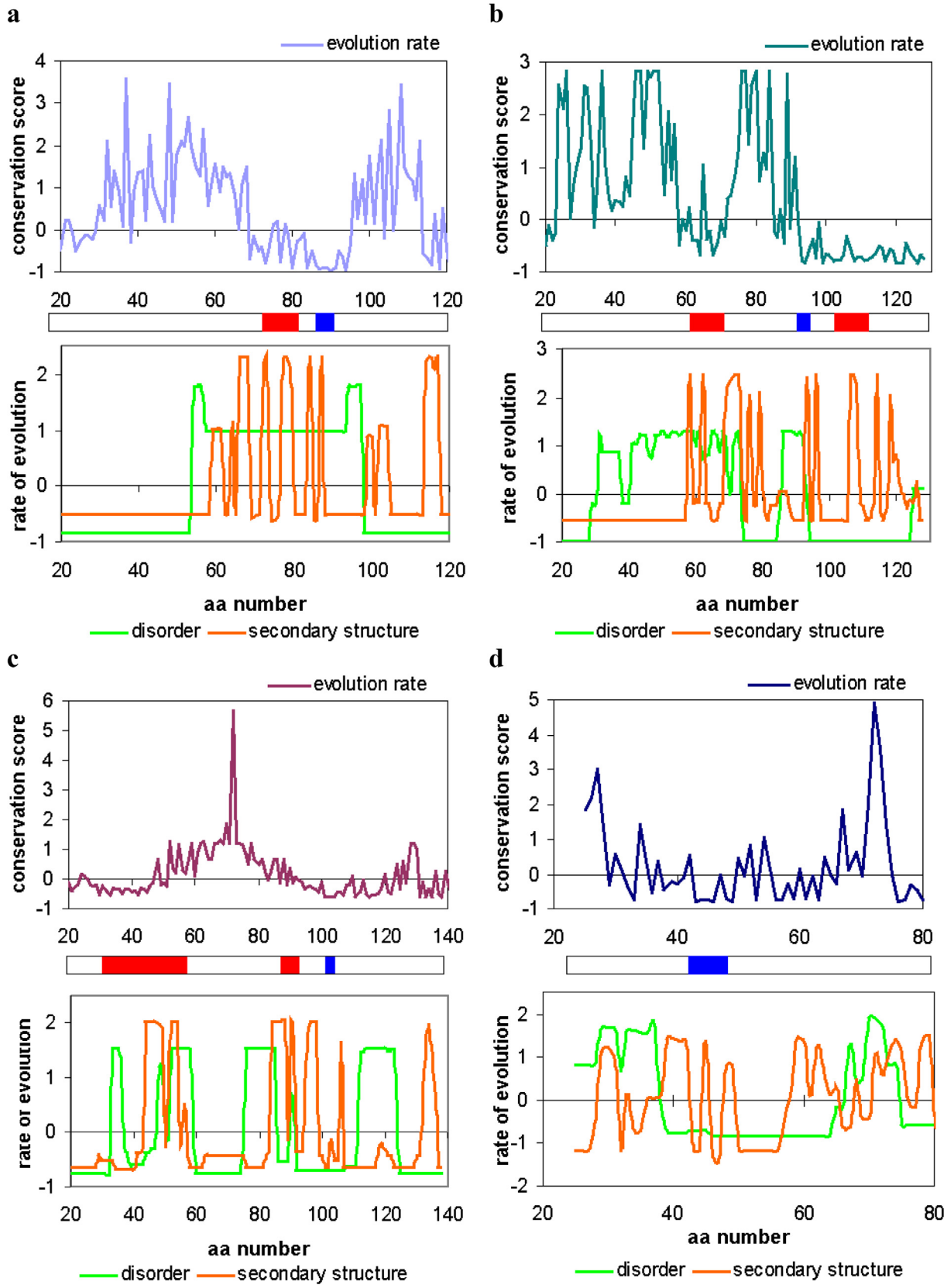

Fig. 2. Rates of evolution per site of hproNTs and evolutionary dynamics of structural properties: a) hproNGF, b) hproBDNF, c) hproNT3 and d) hproNT4. Two graphs are shown in each panel: the rate of evolution per site (top) and the evolutionary dynamics of predicted order–disorder and secondary structures (bottom). The secondary structure propensity bar (middle) indicates conserved helix and strand structures as red and blue box, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using a high quantum yield hydrophobic fluorescent dye (SYPRO Orange) allowed to measure the melting curves, whose first derivative can be used to calculate T_m, a parameter that well describes thermal and conformational stability of proteins with low levels of surface hydrophobicity.

The designed experimental protocol resulted to be not suitable for the hproNT3 sample due to the strong interactions with the hydrophobic fluorophore occurring in the pre-transition region (data not shown). The resulting high initial signal did not allow monitoring protein melting by the fluorescence increase. This behavior is likely ascribable to the higher hydrophobicity profile of hproNT3 compared to the other human members of the family, as estimated by their respective GRAVY indexes.

Indeed, the thermo-melting profiles of the other hproNTs (Fig. 3a) show low and flat background fluorescence in the pre-transition region, and a bimodal profile that can be ascribed to the presence of the two domains. In details, the second sharp transition at high T is due to the unfolding of the globular folded domain of the mature proNT (confirmed by measuring the melting curve of the mature hNGF domain, data not shown). Instead, the first transition that occurs at low T, is likely the result of the denaturation of the pro-domains. In order to better compare pro-domains denaturation of the hproNTs, the respective melting curves have been normalized on the first transition (Fig. 3b).

Interestingly, the melting curves of the pro-domain of hproNGF and hproBDNF are almost superimposable (with T_m of 37.8 ± 0.5 °C and 38.3 ± 0.4 °C for hproNGF and hproBDNF, respectively), whereas the pro-domain of hproNT4 shows a significant positive shift in its melting curve (T_m 44.3 ± 0.3 °C). This reflects the higher conformational stability in solution of the pro-domain of hproNT4 with respect to the more flexible conformations of the pro-domains of hproNGF and hproBDNF. This experimental result corroborates the bioinformatic analysis of the structural order–disorder propensity in the pro-domain of proNTs, which highlighted a lower disorder content in the pro-domain of hproNT4 if compared to the other members of proNTs family.

3.3. hproNT4 shows a lower conformational plasticity in solution with respect to the other members of the hproNTs family

Solution techniques such as SAXS play a central role in the structural studies of IDPs, because they allow to gain insights not only on protein conformations in solution, but also on protein plasticity that is the main hallmark of the physiological functions of IDPs.

Thus, we performed batch SAXS measurements on serial dilutions of the four hproNTs samples (see Supp. Table S1, that summarizes the overall parameters from the SAXS data), without observing any concentration dependent association state changes. The resulting scattering curves of the four hproNTs are shown in Fig. 4 with the corresponding Guinier plots reported in the inserts.

In all cases, the linearity of the Guinier plots at $0 < s < 1.3/R_g$ confirmed the homogeneity of the samples, although a small level of aggregation could be detected in both hproBDNF and hproNT4 samples. The MM values, derived from the calculation of the corresponding excluded Porod volume V_p, were in agreement with values expected for homodimeric species at all the measured concentrations, taking into account that for globular proteins the Vp in $Å^3$

should numerically be about 1.5–1.7 times larger than the MM in Da, according to an empirical finding by Petoukhov et al. 2012 (Table S1). These findings were also confirmed by the estimation of the MM from the relative forward scattering intensity (s = 0, with s the scattering vector). The increased dimensions of the proNTs with respect to those of the corresponding mature NTs, have been highlighted by comparing the experimental scattering patterns to those computed using the high-resolution crystallographic structures and the models of the mature NTs (dot line in Fig. 4). This caused a poor fit to the experimental scattering curves (i.e., the discrepancy χ^2 values are 5.233, 9.853, 4.241 and 2.754 for hproNGF, hproBDNF, hproNT3 and hproNT4, respectively).

The computed distance distribution functions p(r) (Fig. 5) of the hproNTs are in agreement with the typical features of IDPs, being very asymmetric if compared to the highly symmetric p(r) of globular proteins and displaying an extended tail, due to the variety of extended conformations that co-exist in solution.

Interestingly, the structural parameters derived from the scattering curves depict a scenario regarding the proNTs conformations in solution that was hard to predict based just on the relative lengths of their primary structures. Indeed, the radii of gyration (R_g) and consistently the maximum dimensions (D_{max}) derived from the p(r)s (Table S1) do not directly correlate with the MM of the hproNTs. In details, the hproNGF shows the most compact conformation with an R_g of 28.8 ± 0.01 Å and Dmax of 95 ± 3 Å. Consistently hproNT3, whose pro-domain is 17 residues longer, is slightly less compact with an R_g of 31.9 ± 0.05 Å and D_{max} of 110 ± 5 Å. On the contrary, hproBDNF, whose pro-domain is inter-

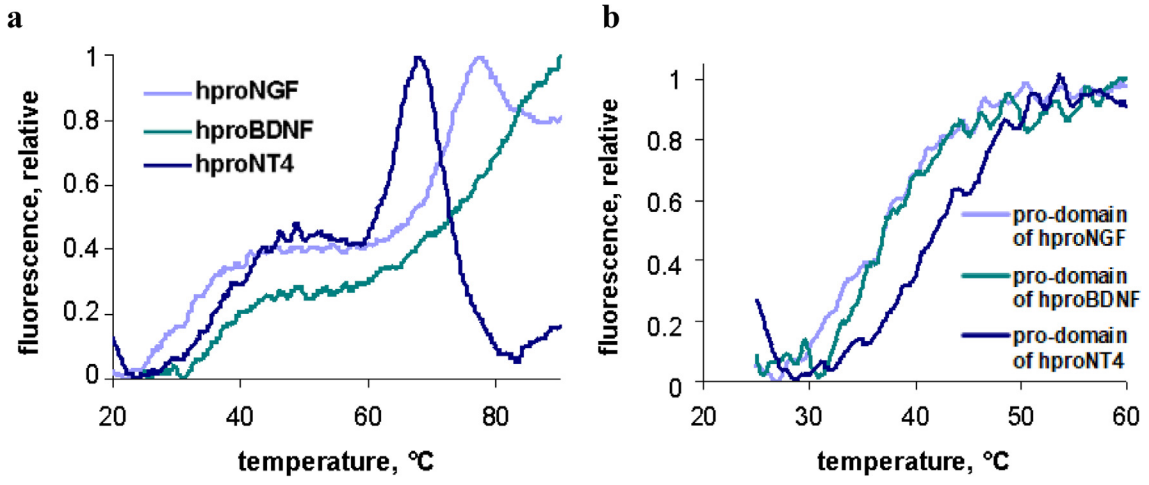

Fig. 3. The pro-domain of hproNT4 shows an increased thermal stability compared to the ones of hproNGF and hproBDNF. Thermal denaturation assay by DSF with SYPRO© dye showing the overall melting curves in the sampled temperature range (a) and the melting profiles relative to the first transition (b), which describes the denaturation of the pro-domain. Experiments were performed in triplicate: the mean curve is shown for clarity.

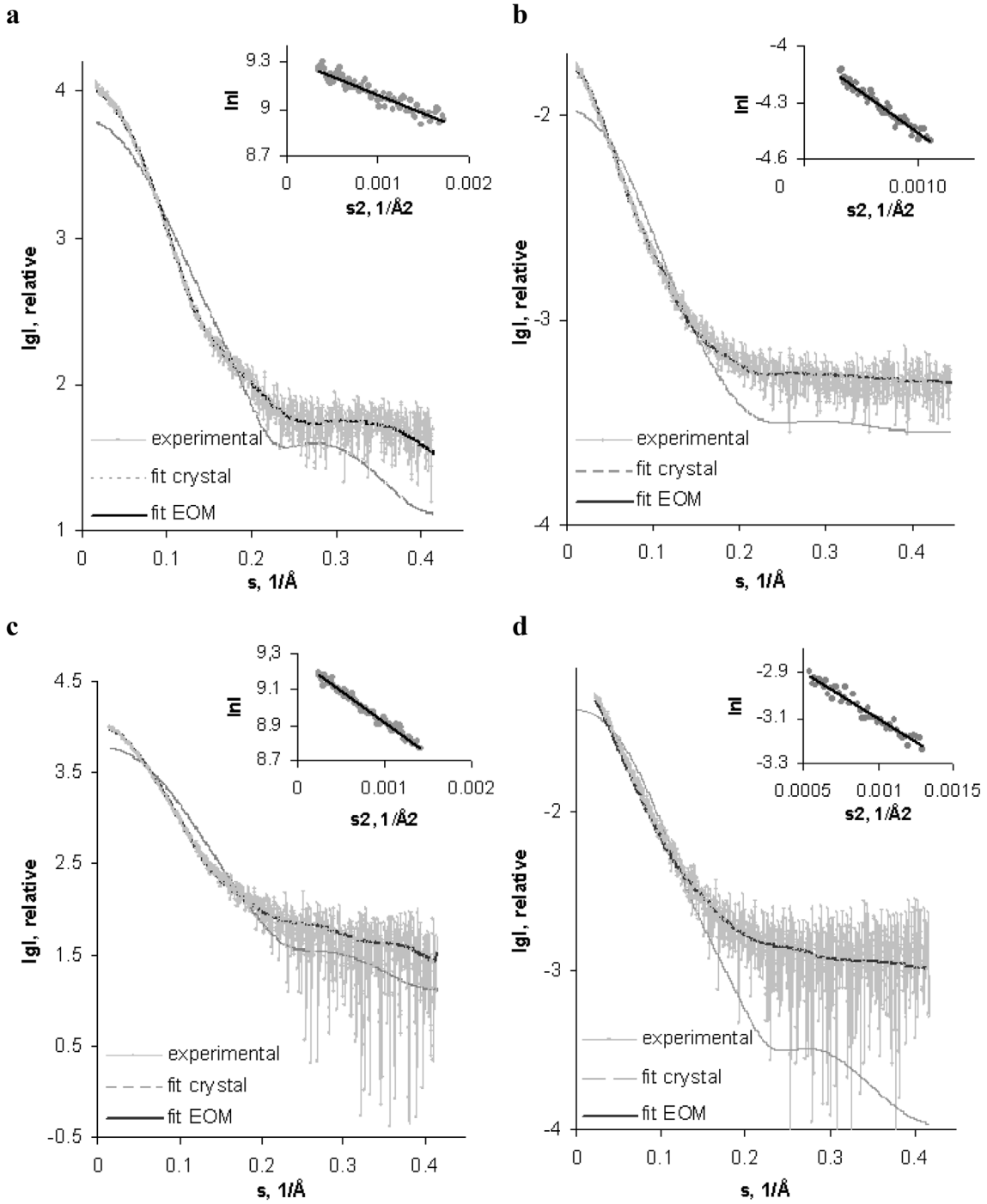

Fig. 4. Experimental SAXS patterns and the obtained fits of the hproNTs: hproNGF (a), hproBDNF (b), hproNT3 (c) and hproNT4 (d). Experimental SAXS patterns, scattering calculated from the crystallographic models of the respective mature NTs by CRYSOL (fit crystal) and scattering calculated from models obtained by EOM (fit EOM) are plotted as indicated. The plots display the logarithm of the scattering intensity as a function of momentum transfer (s). The corresponding Guinier plots are shown in the inserts.

mediate in length among the former hproNTs, is by far the most elongated with an R_g of 37.2 ± 0.06 Å and D_{max} of 125 ± 6 Å. The most striking result concerns hproNT4, whose extended conformation is comparable to that of hproBDNF, with an R_g of 35.0 ± 0.05 Å and Dmax of 115 ± 6 Å, despite its pro-domain is half in length.

Besides deciphering the molecular dimensions of a protein in solution, SAXS data are extremely useful when used in combination with other methods (in particular with high-resolution techniques such as X-ray crystallography), especially in the case of proteins containing IDRs. Indeed, SAXS measurements allow to infer the low-resolution shape of a protein in solution. Thus, to get insights of the mean conformation of the hproNTs, multiple runs of the program CORAL [79] were employed for the 3D reconstruction of the missing pro-peptide regions, by assuming either an asymmetrical conformation or by applying P2 symmetry (Supp. Figs. S6 and S7, respectively). The resulting models, which provide a representation of the average conformations in solution, corroborate the calculated structural parameters that point to larger

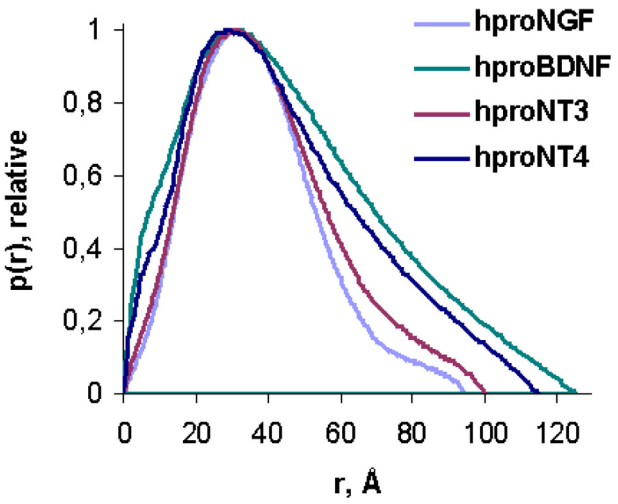

Fig. 5. Comparison of the distance distribution functions of the hproNTs.

dimensions of hproNT4 in comparison to hproNGF and hproNT3, despite the differences in the length of the respective pro-domains (for details, see Supplemetary Material).

As discussed, IDPs and multi-domain proteins that include large IDRs, like the proNTs family, are extremely dynamical structures in solution that are very hard to characterize at an atomic resolution by conventional methods especially in the case of proteins whose MM exceeds the limits of NMR. To this regard SAXS is not only suitable to determine averaged structural parameters and low resolution structures of flexible proteins in solution, but allows to analyze protein dynamic behavior providing the basis for the structural characterization of IDRs. Although SAXS typically yields low-resolution information, the combination of SAXS and atomistic simulations can provide insight into conformational changes in proteins and protein flexibility [45,85,90,91]. In particular, the Ensemble Optimization Method (EOM) is a powerful tool that allows to facilitate the construction of conformational ensembles for which the ensemble-averaged theoretical SAXS profile is in agreement with experimentally determined SAXS profiles [85]. The resulting conformational ensemble provides a rich dataset that can be used to study the role of flexibility in protein function. Therefore, the development of this tool has deeply improved the characterization of flexible systems, and in particular of IDPs. Indeed, instead of only a mean R_g value, obtained by the Guinier plot and the p(r) profile just depicting an averaged behavior in solution, the EOM generates the statistical distributions of Rg and the representative conformational models that reflect the conformational heterogeneity of the macromolecule in solution.

Therefore, besides qualitatively confirming the intrinsic disorder porperties of human proNTs by Kratky plot analysis (data not shown), EOM was employed to obtain a quantitative characterization of plasticity and size distribution of possible multiple configurations of the four proNTs in solution. EOM searches for optimal conformational ensemble models of flexible proteins to match the experimental scattering profiles (i.e χ^2 of 1.007, 0.595, 0.564 and 1.066 for hproNGF, hproBDNF, hproNT3 and hproNT4, respectively; see fit EOM in Fig. 4). As shown in Fig. 6, the behavior of the four proNTs in solution is significantly different.

Indeed, the profile of the R_g distribution of hproNGF closely resembles that of hproNT3 (Fig. 6a and c, respectively). Overall, both curves show a bimodal distribution with a sharp and symmetric peak in the range of 20–30 Å and an extended tail, consistent with their disordered nature that is more pronounced in the case of hproNT3, where the distribution of the selected ensemble is as wide as the distribution of randomly generated models. Although the peak accounts for the great majority of the hproNGF conformers population (75%), the corresponding peak of hproNT3 is less populated, being only 46.2% of the total. Accordingly, the tail of the hproNGF profile spans a shorter range (30–50 Å), which is less populated (25%) compared to the one of hproNT3 that covers a range of 30–62 Å (accounting for the 53.8% of the population with a significant residual of 5% at high range, i.e., 50–62 Å). Thus, these two hproNTs in solution are characterized by a high degree of conformational plasticity that is the typical hallmark of IDPs. This feature is even more dramatic in the case of hproBDNF (Fig. 6b). In fact, an almost continuous distribution of the selected ensembles is observed that covers a wider range with respect to the distribution of randomly generated models. The profile is still bimodal and it is characterized by two peaks. The main one (62% of the total population) is sharp and symmetric with R_g ranging 22–42 Å whereas the second peak is broader and asymmetric (38% of the total population), with a tail accounting for very extended conformations (0.5% of the total population), with R_g spanning between 42 and 70 Å and having a maximum around 49 Å.

Intriguingly, the profile obtained for hproNT4 is very divergent and points to a significantly lower conformational heterogeneity in solution. The R_g values of the individual members of the conformational ensemble (Fig. 6d) spanned a range shifted to higher dimensions (ranging from 28 to 46 Å), characterized by a roughly unimodal distribution with the main population (91.8%) spanning between 32 and 42 Å. Although the shift of the peak position towards higher Rg values accounts for the overall more extended conformation with respect to hproNGF and hproNT3 (in agreement with the parameters calculated for the mean conformation in solution), the distribution of hproNT4 is significantly narrower if compared to the other members of the hproNT family, confirming the predicted lower degree of disorder. Its reduced plasticity might be regarded as a key feature that prevents hproNT4 to fulfill proNTs biological functions, related to receptor binding.

4. Discussion

The present study aimed to understand the molecular determinants essential to mediate physiological functions of the intrinsically disordered proNTs. To this end, widely adopted biophysical and bioinformatics approaches have been exploited for their structural characterization. The results allowed us to shed lights on the dynamics and conformational landscapes of biologically active proNTs and identify specific and conserved structural motifs that are likely to be involved in the interactions with physiological binding partners.

Concerning the evolution of the proNTs, their molecular phylogeny suggests that this family is the result of two gene duplications [89]. The two different intermediate neurotrophin ancestral genes, resulting from the first duplication event, gave separately rise to two couple of paralogues (i.e., proNGF/proNT3 and proBDNF/proNT4, respectively). These duplications events took place early in vertebrate history, before the appearance of cartilaginous fish, but after the split of lampreys and hagfish from the common vertebrate lineage. The consequence of this early evolution is a high variability at the level of the pro-domains both in sequence length and in sequence composition, being the pro-domain of hproNT4 the most divergent. To this regard, the analysis of the molecular evolution of three avian NTs genes revealed that the high conservation exhibited by the BDNF pro-domain results from intense functional constraints that are relaxed in NGF and somewhat relaxed in the NT3 pro-domains respectively [93]. It is worth noting that the NT4 gene has apparently been lost during evolution

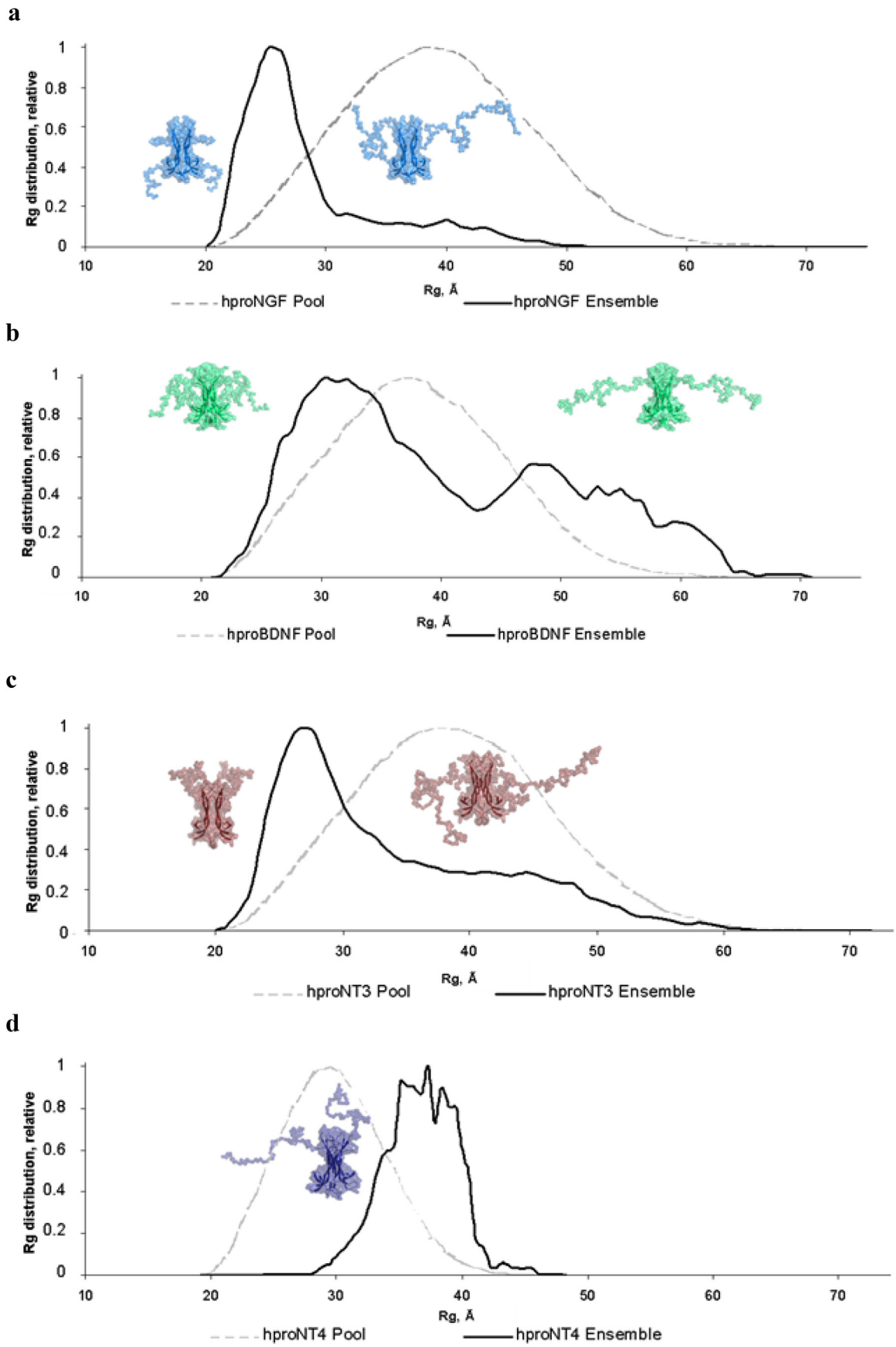

Fig. 6. hproNT4 displays a reduced conformational heterogeneity in comparison with the other members of neurotrophin family. R_g distributions obtained by EOM for hproNGF (a), hproBDNF (b), hproNT3 (c) and hproNT4 (d). The distributions for the initial random pools of models and for the selected ensembles are shown by grey dot lines and black solid lines, respectively. The representative conformations (semitransparent surfaces) are shown near the distributions: compact on the left and extended conformation on the right. Folded homodimers of mature NTs are depicted as cartoons. Figure produced by Pymol [92].

in avian taxon, prompting BDNF to likely play a compensatory role in the development of the avian nervous system.

Previous structural studies on members of proNTs family revealed that the pro-domains of proNGF and proBDNF lack of folded structure [26,29]. In agreement with these findings, the prediction of order–disorder propensity of the pro-domain of proNTs presented in this study, experimentally validated by DSF and SAXS measurements, also showed that the three members of this family but proNT4 are highly disordered in structure. To this regard, the present analysis on the evolution of the structural order–disorder propensity reveals an opposite transition of the order–disorder propensity for proNT4 with respect to other proNTs during the evolution of vertebrates that is most likely related to the loss of its biological functions (i.e., apoptosis and LTD). The EOM analysis of the SAXS patterns allowed to characterize the wide variety of conformational ensembles as result of the hproNTs high flexibility. hproNT4 shows a very low degree of disorder comparing to the others hproNTs. Indeed, it explores a rather restricted conformational space with a global low plasticity that in turn prevents from adapting to the physiological binding partners or to the changing environments of the cell, which on the contrary are known conditions to rapidly modulate IDPs flexibility and conformation [94]. Moreover, comparing the EOM profiles, it is tempting to speculate that the active conformers, which are able to fulfill the proNTs biological functions associated with receptors binding, are clustered in the main peak of the three bimodal distributions (ranging from 20 to 30 Å). These conformational ensembles can be considered in dynamic equilibrium with all the other population of conformers due to the high degree of disorder observed for hproNGF, hproBDNF and hproNT3. This crucial behavior might be considered as the structural determinant that provides flexible and rapid adaptation networks triggering proNT biological activities. On the contrary, hproNT4 shows a very limited plasticity in solution, sampling a very limited range of conformations that are more extended respect to the putatively active conformation ensemble of the other members of the family.

Even if IDPs are highly dynamic and do not form tertiary structures, they are known to contain variable amounts of transient secondary structures. The presence of these transiently formed structural elements has been reported to facilitate partner recognition and to tune the thermodynamics and kinetics of the binding interactions [95–97]. Interestingly, the present analysis on the evolution of secondary structure propensity highlighted conserved secondary structures in all vertebrates.

All proNTs share a conserved small beta strand secondary structure that might act as a pre-formed structural element involved in the chaperonin function of the pro-domains that are known to assist proper folding of the mature part.

Most importantly, the three biologically active proNTs are characterized by the predicted presence of an alpha helix secondary structure, which exhibits lower rates of evolution than average. This structural determinant is absent in the pro-domain of proNT4 and therefore might be envisaged as the structural basis mediating receptor recognition in a highly specific manner. Regarding this feature, previous studies both on hproBDNF by CD spectroscopy [98] and on hproNGF by hydrogen/deuterium (H/D) exchange and by combined NMR-SAXS measurements [46,99] confirmed the presence of a localized higher-order structure motif in their respective pro-domains. The observation that the reported conserved alpha-helix structure is predicted to be proximal to the observed H/D exchange protection, can be ascribed to the fact that the first four backbone amide hydrogens at the N terminus of an alpha-helix are unprotected because they are not engaged in hydrogen bonding.

The findings that the disorder and the alpha-helix secondary structure propensities appear to be highly conserved throughout evolution of the three biologically active proNTs, prompt these features to likely play in concert to mediate receptor binding. Indeed, these structural determinants could act as functional modules within the frames of disordered regions in furtherance of the interaction with the globular domains of the binding partners, assisted by flexible structures around them [100], and thus be constrained to maintain the binding affinity to the specific receptors of the VPS10p family. Moreover, being predicted to be conserved in disordered structures across evolution, these elements that have lower rates of evolution are generally considered to be relevant to mediate physiological protein–protein interactions. Accordingly, the three proNTs, which play similar critical function in their interaction with specific receptors of the VPS10p family, are characterized by constrained disordered structures in vertebrates. These observations are in line with the well-known hallmarks of IDPs, by which essential intermolecular interactions and conserved binding partners could constrain disordered regions through evolution.

5. Conclusion

In conclusion, the disordered and dynamic nature of the pro-domain of biologically active proNTs have been addressed, for the first time, in the present study, by a combined evolutionary and biophysical approach. The present study represents so far, an important advance in knowledge with respect to the only previous reported phylogenetic analyses limited to the molecular evolution of three avian NTs genes, i.e., proNGF, proBDNF and proNT3 [93] and to BDNF in 36 vertebrates [101].

The gained deeper insights on the evolutionary conserved structural motifs appears to be the main determinant essential for the multifaceted roles of proNTs in physiological as well as in pathological contexts.

CRediT authorship contribution statement

S. Covaceuszach: Conceptualization, Formal analysis, Investigation, Visualization, Writing. **L.Y. Peche:** Formal analysis, Investigation. **P.V. Konarev:** Formal analysis, Investigation. **D. Lamba:** Conceptualization, Formal analysis, Investigation, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the support of the EMBL EU funding for recording at EMBL Hamburg SAXS beamline and the SAXS group of Professor Svergun and Dr Alberto Cassetta for assistence during SAXS data collection. PVK acknowledges the support from Ministry of Science and Higher Education of the Russian Federation within the State assignment FSRC 'Crystallography and Photonics' RAS (SAXS analysis).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.05.007.

References

[1] Chao MV. Neurotrophins and their receptor: a convergence point for many signaling pathways. Nat Rev Neurosci 2003;4:299–309.

[2] Dechant G, Barde Y-A. The neurotrophin receptor p75NTR: novel functions and implications for diseases of the nervous system. Nat Neurosci 2002;5 (11):1131–6.
[3] Huang EJ, Reichardt LF. Trk receptors: roles in neuronal signal transduction. Annu Rev Biochem 2001;72:609–42.
[4] Lu B, Pang PT, Woo NH. The yin and yang of neurotrophin action. Nat Rev Neurosci 2005;6(8):603–14.
[5] Friedman WJ. Proneurotrophins, seizures, and neuronal apoptosis. Neuroscientist 2010;16(3):244–52.
[6] Seidah NG, Benjannet S, Pareek S, Chrétien M, Murphy RA. Cellular processing of the neurotrophin precursors of NT3 and BDNF by the mammalian proprotein convertases. FEBS Lett 1996;379:247–50.
[7] Lee R, Kermani P, Teng KK, Hempstead BL. Regulation of cell survival by secreted proneurotrophins. Science 2001;294:1945–8.
[8] Schweigreiter R. The dual nature of neurotrophins. BioEssays 2006;28 (6):583–94.
[9] Nykjaer A, Lee R, Teng KK, Jansen P, Madsen P, Nielsen MS, et al. Sortilin is essential for proNGF induced neuronal cell death. Nature 2004;427 (6977):843–8.
[10] Teng HK, Heng KK, Lee R, Wright S, Tevar S, Almeida RD, et al. ProBDNF induces neuronal apoptosis via activation of a receptor complex of p57NTR and sortilin. J Neurosci 2005;25:5455–63.
[11] Yano H, Torkin R, Martin LA, Chao MV, Teng KK. Proneurotrophin-3 is a neuronal apoptotic ligand: evidence for retrograde-directed cell killing. J Neurosci 2009;29(47):14790–802.
[12] Glerup S, Olsen D, Vaegter C, Gustafsen C, Sjoegaard S, Hermey G, et al. SorCS2 regulates dopaminergic wiring and is processed into an apoptotic two-chain receptor in peripheral glia. Neuron 2014;82(5):1074–87.
[13] Mizui T, Ishikawa Y, Kumanogoh H, Lume M, Matsumoto T, Hara T, et al. BDNF propeptide actions facilitate hippocampal LTD and are altered by the common BDNF polymorphism Val66Met. Proc Natl Acad Sci U S A 2015;112(23):E3067–74.
[14] Jansen P, Giehl K, Nyengaard JR, Teng K, Lioubinski O, Sjoegaard SS, et al. Roles for the pro-neurotrophin receptor sortilin in neuronal development, aging and brain injury. Nat Neurosci 2007;10 (11):1449–57.
[15] Deinhardt K, Kim T, Spellman DS, Mains RE, Eipper BA, Neubert TA, et al. Neuronal growth cone retraction relies on proneurotrophin receptor signaling through Rac. Sci Signal 2011;4(202):ra82.
[16] Siao CJ, Lorentz CU, Kermani P, Marinic T, Carter J, McGrath K, et al. ProNGF, a cytokine induced after myocardial infarction in humans, targets pericytes to promote microvascular damage and activation. J Exp Med 2012 ;20912:2291–305.
[17] Westergaard UB, Kirkegaard K, Sørensen ES, Jacobsen C, Nielsen MS, Petersen CM, et al. SorCS3 does not require propeptide cleavage to bind nerve growth factor. FEBS Lett 2005;579:1172–6.
[18] Feng D, Kim T, Özkan E, Light M, Torkin R, Teng KK, et al. Molecular and structural insight into proNGF engagement of p75NTR and sortilin. J Mol Biol 2010;396(4):967–84.
[19] He X, Garcia KC. Structure of nerve growth factor complexed with the shared neurotrophin receptor p75. Science 2004;304:870–5.
[20] Mcdonald NQ, Lapatto R, Murray-Rust J, Gunning J, Wlodawer A, Blundell TL. New protein fold revealed by a 2.3-A resolution crystal structure of nerve growth factor. Nature 1991;354:411–4.
[21] Leloup N, Chataigner LMP, Janssen BJC. Structural insights into SorCS2-Nerve Growth Factor complex formation. Nat Commun 2018;91:2979–89.
[22] Rattenholl A, Lilie H, Grossmann A, Stern A, Schwarz E, Rudolph R. The pro-sequence facilitates folding of human nerve growth factor from Escherichia coli inclusion bodies. Eur J Biochem 2001;268:3296–303.
[23] Kliemannel M, Ratteholl A, Ralph G, Balbach J, Lilie H, Rudolph R, et al. The mature part of proNGF induces the structure of its pro-peptide. FEBS Lett 2004;556:207-12.
[24] Kliemannel M, Golbik R, Rudolph R, Schwarz E, Lilie H. The pro-peptide of proNGF: structure formation and intramolecular association with NGF. Protein Sci 2007;16(3):411–9.
[25] Hauburger A, Kliemannel M, Madsen P, Rudolph R, Schwarz E. Oxidative folding of nerve growth factor can be mediated by the pro-peptide of neurotrophin-3. FEBS Lett 2007;581:4159–64.
[26] Paoletti F, Covaceuszach S, Konarev PV, Gonfloni S, Malerba F, Schwarz E, et al. Intrinsic structural disorder of mouse proNGF. Proteins 2009;75 (4):990–1009.
[27] Paoletti F, Malerba F, Kelly G, Noinville S, Lamba D, Cattaneo A, et al. Conformational Plasticity of proNGF. PLoS ONE 2011;6(7):e22615.
[28] Paoletti F, Malerba F, Bruni Ercole B, Lamba D, Cattaneo A. Cattaneo A comparative analysis of the structural, functional and biological differences between Mouse and Human Nerve Growth Factor. Biochim Biophys Acta 2015;1854(3):187–97.
[29] Anastasia A, Deinhardt K, Chao MV, Will NE, Irmady K, Lee FS, et al. Val66Met polymorphism of BDNF alters prodomain structure to induce neuronal growth cone retraction. Nat Commun 2013;4(1). https://doi.org/10.1038/ncomms3490.
[30] Trabjerg E, Abu-Asad N, Wan Z, Kartberg F, Christensen S, Rand KD. Investigating the conformational response of the sortilin receptor upon binding endogenous peptide- and protein ligands by HDX-MS. Structure 2019;27(7):1103–1113.e3.
[31] Wang J, Bains H, Anastasia A, Bracken C. NMR backbone resonance assignments of the prodomain variants of BDNF in the urea denatured state. Biomol NMR Assign 2018;12(1):43–5.
[32] Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, et al. D2P2: Database of Disordered Protein Predictions. Nucl Acids Res 2013;41:D508-16.
[33] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins: Structure. Funct Genet 2000;41(3):415–27.
[34] Arai M, Sugase K, Dyson HJ, Wright PE. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. Proc Natl Acad Sci USA 2015;112:9614–9.
[35] Dyson HJ. Making Sense of Intrinsically Disordered Proteins. Biophys J 2016;110(5):1013–6.
[36] Shammas SL, Crabtree MD, Dahal L, Wicky BIM, Clarke J. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. J Biol Chem 2016;291 (13):6689–95.
[37] Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A 2000;97(16):8868–73.
[38] Huang Y, Liu Z. Liu K Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. J Mol Biol 2009;393(5):1143–59.
[39] Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci 2008;33(1):2–8.
[40] Toth-Petroczy A, Palmedo P, Ingraham J, Hopf TA, Berger B, Sander C, et al. Structured States of Disordered Proteins from Genomic Sequences. Cell 2016;167(1):158–170.e12.
[41] Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and Disorder. Curr Opin Struct Biol 2011;21(3):441–6.
[42] Bellay J, Han S, Michaut M, Kim TaeHyung, Costanzo M, Andrews BJ, et al. Bringing order to protein disorder through comparative genomics and genetic interactions. Genome Biol 2011;12(2):R14. https://doi.org/10.1186/gb-2011-12-2-r14.
[43] Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. Chem Rev 2004;104(8):3607–22.
[44] Jensen MR, Ruigrok RWH, Blackledge M. Describing intrinsically disordered proteins at atomic resolution by NMR. Curr Opin Struct Biol 2013;23 (3):426–35.
[45] Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering SAXS combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 2007;40(3):191–285.
[46] Yan R, Yalinca H, Paoletti F, Gobbo F, Marchetti L, Kuzmanic A, et al. The Structure of the Pro-domain of Mouse proNGF in Contact with the NGF Domain. Structure 2019;27(1):78–89.e3.
[47] Doniach S. Changes in biomolecular conformation seen by small angle X-ray scattering. Chem Rev 2001;101(6):1763–78.
[48] Bernadó P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol Biosyst 2012;8(1):151–67.
[49] Receveur-Brechot V, Durand D. How random are intrinsically disordered proteins? A small angle scattering perspective. Curr Protein Pept Sci 2012;13:55–75.
[50] Bernadó P, Blackledge M. Structural Biology: Proteins in dynamic equilibrium. Nature 2010;468(7327):1046–8.
[51] Fahmi M, Ito M. Evolutionary Approach of Intrinsically Disordered CIP/KIP Proteins. Sci Rep 2019;9(1):1575.
[52] Mizianty MJ, Peng Z, Kurgan L. MFDp2 – accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. Intrinsically Disordered. Proteins 2013;1(1):e24428. https://doi.org/10.4161/idp.24428.
[53] Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005;21(16):3433–4.
[54] Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42(1):38–48.
[55] Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 2005;61(S7):176–82.
[56] Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins 2003;53(S6):566–72.
[57] Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 2010;1804(4):996–1010.
[58] Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
[59] Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucl Acids Res 2015;43(W1):W389–94.
[60] Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein—an open resource for online prediction of protein structural and functional features. Nucl Acids Res 2014;42:W337–43.
[61] Yang Y, Gao J, Wang J, Heffernan R, Hanson J, Paliwal K, et al. Sixty-five years of the long march in protein secondary structure prediction: the final stretch?. Brief Bioinformatics 2018;19(3):482–94.
[62] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10.
[63] The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucl Acids Res 2019;47:D506–15.

[64] Madeira F, Park YM, Lee J, Buso N, Gur T, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucl Acids Res 2019;47W1:W636-41.
[65] Mizianty MJ, Zhang T, Xue B, Zhou Y, Dunker A, Uversky VN, et al. In-silico prediction of disorder content using hybrid sequence representation. BMC Bioinf 2011;12(1):245. https://doi.org/10.1186/1471-2105-12-245.
[66] Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucl Acids Res 2016;44(W1):W344–50.
[67] Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. GLOOME: gain loss mapping engine. Bioinformatics 2010;26(22):2914–5.
[68] van den Ent F, Löwe J. RF cloning: A restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods 2006;67(1):67–74.
[69] Covaceuszach S, Capsoni S, Ugolini G, Spirito F, Vignone D, Cattaneo A. Development of a non invasive NGF-based therapy for Alzheimer's disease. Curr Alzheimer Res 2009;62:158–70.
[70] Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM, et al. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline PETRA III. DESY J Appl Crystallogr 2015;48(2):431–43.
[71] Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Cryst 2017;50(4):1212–25.
[72] Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr 2003;36(5):1277–82.
[73] Guinier AL. diffraction des rayons X aux tres petits angles: applications a l'etude de phenomenes ultramicroscopiques. Ann Phys Paris 1939;12:161–237.
[74] Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 1992;25 (4):495–503.
[75] Porod G. In: Small Angle X-ray Scattering. Academic Press; 1982. p. 17–51.
[76] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195–201.
[77] Robinson RC, Radziejewski C, Stuart DI, Jones EY. Structure of the brain-derived neurotrophic factor/neurotrophin 3 heterodimer. Biochemistry 1995;34(13):4139–46.
[78] Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–47.
[79] Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Cryst 2012;45(2):342–50.
[80] Butte MJ, Hwang PK, Mobley WC, Fletterick RJ. Crystal structure of neurotrophin-3 homodimer shows distinct regions are used to bind its receptors. Biochemistry 1998;37(48):16846–52.
[81] Banfield MJ, Naylor RL, Robertson AGS, Allen SJ, Dawbarn D, Brady RL. Specificity in Trk-Receptor: neurotrophin Interaction: the crystal structure of Trkb-D5 in complex with neurotrophin-4/5. Structure 2001;9:1191–9.
[82] Covaceuszach S, Konarev P, Cassetta A, Paoletti F, Svergun D, Lamba D, et al. The conundrum of the high-affinity NGF binding site formation unveiled?. Biophys J 2015;108(3):687–97.
[83] Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. J Appl Crystallogr 2001;34(1):33–41.
[84] Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr 2003;36(3):860–4.
[85] Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 2007;129(17):5656–64.
[86] Svergun D, Barberato C, Koch MHJ. CRYSOL – a Program to Evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 1995;28(6):768–73.
[87] Atkins J, Boateng S, Sorensen T, McGuffin L. Disorder prediction methods, their applicability to different protein targets and their usefulness for guiding experimental studies. Int J Mol Sci 2015;16(8):19040–54.
[88] Chen ZY, Ieraci A, Teng H, Dall H, Meng CX, Herrera DG, et al. Sortilin controls intracellular sorting of brain-derived neurotrophic factor to the regulated secretory pathway. J Neurosci 2005;25:6156–66.
[89] Hallböök F. Evolution of the vertebrate neurotrophin and Trk receptor gene families. Curr Opin Neurobiol 1999;9(5):616–21.
[90] Mertens HDT, Svergun DI. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J Struct Biol 2010;172(1):128–41.
[91] Rambo RP, Tainer JA. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. Curr Opin Struct Biol 2010;20(1):128–37.
[92] Schrödinger LLC. The PyMOL molecular graphics system. Version 2010;1.3r1.
[93] Sehgal RNM, Lovette IJ. Molecular evolution of three avian neurotrophin genes: implications for proregion functional constraints. J Mol Evol 2003;57 (3):335–42.
[94] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 2015;16(1):18–29.
[95] Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, et al. Characterization of molecular recognition features, MoRFs, and their binding partners. J Mol Biol 2006;362:1043–59.
[96] Pancsa R, Fuxreiter M. Interactions via intrinsically disordered regions: What kind of motifs?. IUBMB Life 2012;64:513–20.
[97] Schneider R, Maurin D, Communie G, Kragelj J, Hansen DF, Ruigrok RWH, et al. Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. J Am Chem Soc 2015;137(3):1220–9.
[98] Kailainathan S, Piers TM, Yi JH, Choi S, Fahey MS, Borger E, et al. Activation of a synapse weakening pathway by human Val66 but not Met66 pro-brain-derived neurotrophic factor (proBDNF). Pharmacol Res 2016;104:97–107.
[99] Trabjerg E, Kartberg F, Christensen S, Rand KD. Conformational characterization of nerve growth factor-β reveals that its regulatory pro-part domain stabilizes three loop regions in its mature part. J Biol Chem 2017;292(40):16665–76.
[100] van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. Chem Rev 2014;114(13):6589–631.
[101] Tettamanti G, Cattaneo AG, Gornati R, de Eguileor M, Bernardini G, Binelli G. Phylogenesis of brain-derived neurotrophic factor (BDNF) in vertebrates. Gene 2010;450(1-2):85–93.