Chemokines: Chemistry, Biochemistry and Biological Function

Daniel F. Legler* and Marcus Thelen

Abstract: The in vitro synthesis of correctly folded functional proteins remains challenging. Chemokines, which consist of only 70–100 amino acids, are accessible through solid-phase synthesis and easily fold into a thermally stable tertiary structure. From the time of their discovery in the late 1980s chemokines could therefore be synthesized using biochemical and chemical protocols for structure-function analyses and for exploring the chemokine system in vitro and in vivo. In this short overview aimed at a chemistry-oriented readership we will introduce chemokines in general, and then discuss their structure, their isolation from biological materials, as well as the different methods to produce chemokines in the laboratory and finally we will present some examples of their functions in vivo.

Keywords: Chemokines - Cytokines

Introduction

Cytokines are secreted proteins, which activate and mediate communication between immune cells, regulate hematopoiesis, and immune responses during inflammation. Chemotactic cytokines, a large subfamily with approx. 50 members, are functionally characterized by their ability to stimulate cell migration through cognate G-protein coupled receptors (GPCRs). From an evolutionary point of view, the chemokine system is conserved from jawed vertebrates to humans. Some chemokines are highly conserved and can stimulate cells in fish as well as in mammals. Chemokines are small proteins of 8–12 kDa, which share four cysteines forming two characteristic disulfide bonds that are critical for their conserved structure, which is referred to as the chemokine fold. Hence, chemokines are defined through their function as chemoattractants with a common secondary and tertiary structure. The spacing of the first two cysteines has been used as a lead for systematic nomenclature: in the CC group the cysteines are adjacent, in the CXC group they are separated by any amino acid while CX,C ligand 1 (CX,C1) contains three separating amino acids. The chemokines XCL1 and XCL2 fall somewhat out of the rule as they possess only one disulfide bridge, but share the chemokine fold.

The primary sequences of chemokines are highly divergent, but almost all give rise to a marked alkaline isoelectric point. In fact this property has been used to isolate chemokines from biological material (see below). For the receptors, a nomenclature was introduced based on the structural features of the CXC, CC and CX,C chemokines, corresponding to CXC receptors (CXC), CC receptors (CCR) and the CX,C receptor (CX,CR1), respectively.

All chemokines possess a flexible (unfolded) N-terminus, preceding the first cysteine followed by a rigid loop that ends in three antiparallel β-strands (Fig. 1).

The C-terminus is helical and at its end is also unfolded. The disulfide bridges [1-3] and [2-4] connect the N-terminus of the rigid loop with the β-strands. Functionally relevant domains of chemokines are the rigid loop, the N-terminus, and glycosaminoglycan (GAG) binding sites located within the β-strands or in the C-terminus. The rigid loop and the N-terminus are important for receptor binding and activation. In the two-step model the chemokine first binds with its rigid loop to the N-terminus of a cognate receptor followed by the insertion of its N-terminus into the activation pocket of the receptor (Fig. 2). Instead, the GAG binding sites are in opposition to the receptor-interacting domains enabling presentation of the chemokine by proteoglycans in an active conformation. Indeed retention of chemokines on cell surfaces or matrices is critical for the formation of chemotactic gradients required for directional cell migration. The N-terminus of the chemokine is most critical for receptor activation and substitution of single amino acids or truncations can lead to a total loss of activity or changes in receptor specificity. The notion is consistent with the N-terminus falling deep into the binding pocket formed by the transmembrane spanning helices of the GPCRs and thereby inducing receptor conformational changes.

At physiological concentrations chemokines can oligomerize, a process that is often assisted by sulfated sugars, such as heparan sulfate. However, monomeric chemokines are fully active and may be the major active entity in vivo. In fact, in some cases it was shown that monomeric and dimeric forms can induce different responses on the same receptor. Moreover, in addition to homomers, che-
Chemokines can also oligomerize with other chemokines and trigger responses, which are often synergistic, i.e. more pronounced than elicited by the individual chemokines alone.[14–16] The physiological relevance of the synergism is given at sites of inflammation where multiple chemokines are produced concomitantly. Finally, chemokines can interact with heterologous proteins, e.g. CXCL12 binds the alarmin high mobility group box 1 (HMGB1), which is secreted by immune cells at sites of inflammation, leading to markedly altered receptor responses.[17]

**Identification of CXCL8 and its First Recombinant Expression**

Neutrophils are the most abundant white blood cells in the circulation and constitute the front line defense of the innate immune system of vertebrates. In the late 1980s almost concomitantly four laboratories, three in Europe and one in the USA, reported the sequence of a monocyte-derived neutrophil activating peptide. [18] In Switzerland the laboratory of Marco Baggiozini at the Theodor Kocher Institute of the University of Bern isolated CXCL8, which had been originally called Neutrophil Activating Factor (NAF), and was later renamed to NAP-1 and IL-8. In all laboratories CXCL8 was purified from supernatants of stimulated monocytes by classical biochemical protein purification procedures including ammonium sulfate precipitation, different steps of column chromatography, such as size exclusion, ion-exchange and reverse-phase high pressure liquid chromatography (HPLC).[19–22] The protocols used chemotaxis and enzyme release from intracellular granules of neutrophils as bio-assisys to follow the activity in the different fractions during purification. In those days the human genome was not sequenced and therefore classical amino acid sequencing by Edman degradation was performed to reveal the primary sequence. It turned out that the sequence was identical to that of a predicted secreted protein deduced from cDNA, which was obtained by reverse transcription of mRNA from mitogen-stimulated leukocytes. However, the hypothetical product was not characterized.[19]

The laboratory in Bern together with the Sandoz (later Novartis) Research Institute in Vienna went on and synthesized the gene of mature CXCL8 through hybridization of six overlapping synthetic oligonucleotides. The artificial gene was inserted into a plasmid and transduced into E. coli for protein production.[24] Chemokines, when expressed in bacteria, are usually not soluble, but form densely packed amorphous inclusion bodies.[25,26] A possible explanation for the formation of inclusion bodies is that many chemokines are bactericidal either in their full-length conformation or as cleavage products and bacteria may store the toxic material in an inactive conformation.[27]

Chemokines can easily be extracted from inclusion bodies and folded into their functional conformations. The amorphous insoluble structures are a convenient source for 80–90% pure chemokine. Inclusion bodies are isolated by differential centrifugation, rigorously washed with mild chaotropic reagents, e.g. 2M urea, and subsequently solubilized in 6M guanidine HCl under strong reducing conditions to break disulfide bridges. For convenience, recombinant chemokines are often tagged with histidines so that they can be easily isolated on immobilized metal ion affinity columns (IMAC). This chromatography step usually yields >95% pure chemokines which are in a linear and unfolded conformation. Removal of guanidine by dialysis in most cases causes the precipitation of the crude material, but is a convenient method for its isolation. Pellets are then solubilized in a small volume of 6M guanidine under strong reducing conditions at high protein concentrations (5–10 mg/ml). Due to their basic isoelectric points unfolded chemokines are soluble at low pH. Therefore, simple buffer exchange on a desalting column equilibrated at pH 3 allows removal of the guanidine and the reducing agent. Subsequent rapid dilution into a redox-balanced buffer at physiological pH, which favors disulfide bridge formation in the presence of low molecular weight chaperones, e.g. arginine, permits efficient folding of chemokines. The diluted chemokines can be recovered by cation chromatography or by reverse phase chromatography on neutral, pH resistant matrices, such as POROS® R2 (Thermo Fisher Scientific).[25]

A critical step for the preparation of recombinant chemokines is the formation of a correct N-terminus. Natural chemokines are translated from their genes with a leader sequence at the N-terminus, which is required for secretion and is cleaved off after membrane translocation. This natural truncation gives no preference for amino acids at the N-terminus, whereas bacterial expressed chemokines start with a formylated methionine, which can significantly alter receptor activation. The starting amino acid of recombinantly produced chemokines can be enzymatically removed after purification with methionine aminopeptidases. Alternatively, if an affinity tag for purification was added at the N-terminus, insertion of a consensus sequence for proteolytic cleavage allows the removal of this part. A convenient protease for cleavage is enterokinase, because the enzyme cuts right after its specific consensus sequence leaving no undesired amino acids at the N-terminus of the chemokine.[25,26] Mature chemokines can easily be separated from uncleaved material or from the clipped N-terminus by reversed-phase HPLC.

Alternatively, fusing an N-terminal poly-histidine-SUMO (small ubiquitin-related modifier) double tag to the chemokine entails two advantages for generating functional chemokines. First, the addition of a SUMO-tag renders the recombinant chemokine-fusion protein soluble and prevents its accumulation in inclusion bodies. The soluble poly-histidine-tagged-chemokine-fusion protein can be easily purified on a classical IMAC. Second, the poly-histidine-SUMO tag can be cleaved off with an impeccable specificity and efficiency using the SUMO protease 1 (Ulp1) liberating the native chemokine with its correct mature N-terminus.[20,29] If Ulp1 is used as poly-histidine-tagged version, both Ulp1,
as well as the poly-histidine-SUMO tag can simply be removed by re-running the protein over an IMAC.\textsuperscript{[31]}

**Chemical Synthesis**

An alternative method to generate chemokines is full chemical solid-phase synthesis. Due to the relatively short amino acid sequence and the favorable folding properties, chemical synthesis is an elegant and efficient method, however, this method requires specialized laboratory equipment and chemical knowhow, which are not commonly found in life science-dedicated laboratories. In the early 90s Ian Clark-Lewis in Vancouver was amongst the first starting to synthesize chemokines.\textsuperscript{[32]} His protocol is based on automated solid-phase synthesis of chemokines, which are then cleaved off from the resin, chemically deprotected and precipitated from organic solvents. Guanidine and strong reducing conditions were used to solubilize the crude material in aqueous buffers at pH 3. After separation on a reversed-phase column, chemokines were subjected to folding. Correctly folded proteins, which possess typically a more hydrophilic surface bind markedly less well to reverse phase, can easily be separated from unfolded material. Chemical synthesis provides several advantages. Only one product with a defined sequence is obtained and no cleavage of amino acids from the N-terminus is required. Chemical synthesis also allows the protein sequence to be manipulated and non-natural amino acids to be introduced. Because random chemical labeling interferes with receptor binding, site-specific modifications are often required for biological studies, including the determination of receptor affinity and of receptor surface expression (in the absence of available receptor-specific antibodies). Fluorophore-labeled or biotinylated amino acids are often introduced near the C-terminus and were shown not to modify receptor interactions. Chemical synthesis was also used for a systematic truncation and modification of the amino-terminal sequence preceding the first cysteine of CXCL8 to unveil the importance of this domain for receptor binding and activation. The extensive studies of structure–function relations included a large number of synthetic analogs with single amino acid substitutions and showed that, with the exception of the cysteines and the glutamic acid-leucine-arginine motif (ELR motif), no other residue appeared to be required for functional CXCL8 receptor interaction.\textsuperscript{[33]}

**Chemokine Function**

Chemokines are best known for their regulation of leukocyte trafficking.\textsuperscript{[34,35]} Leukocyte migration is required for immune homoeostasis, surveillance and responses to infiltrating pathogens. Consequently, chemokines were functionally subdivided into two groups; on one side there are those that primarily govern homeostatic immune cell trafficking, but can eventually become upregulated under pathological conditions. On the other side are chemokines that are mainly induced and expressed under inflammatory conditions, often called ‘inflammatory’ chemokines.\textsuperscript{[36]}

The receptors of chemokines phylogenetically map to a group of GPCRs within the gamma subfamily of rhodopsin-like receptors that share the ability to mediate cell migration.\textsuperscript{[37]} Chemokine receptors in general couple to heterotrimeric G-proteins, therefore most responses can be fully inhibited by treatment of cells with Bordetella pertussis toxin. Today, a total of 19 receptors have been identified: 7 CXCRs, (CXCR1-6 and CXCR8), which is also known as GPR35, 10 CCRs (CCR1-10), CXCR1 and CCR1.\textsuperscript{[31]} Compared with the about 50 chemokines it becomes evident that the 19 receptors usually bind more than one chemokine, however, several chemokines can also bind to multiple receptors. The resulting promiscuity of the chemokine system has important consequences for immune responses, since a given receptor may induce different responses depending on the chemokine that triggers its activation. For example it was reported that CXCR3 when stimulated with its high affinity ligand CXCL11 skew cells towards a Th2 phenotype, whereas CXCL10, which binds with lower affinity to the same receptor, induces Th1 differentiation.\textsuperscript{[38]} Th1 and Th2 T cells are known to mediate distinct immune responses.\textsuperscript{[39]}

CCR7 binds two chemokines, CCL19 and CCL21, and is a key regulator for the homing of leukocytes, such as dendritic cells, B and T cells, to secondary lymphoid organs. Although CCL19 and CCL21 both bind to CCR7, the two ligands can induce distinct signaling pathways. The binding of CCL19 to CCR7 leads to receptor internalization, while CCL21 promotes cell adhesion facilitated by tightly binding to GAGs through its extended C-terminus building haptotactic gradients for cell migration.\textsuperscript{[40]}

The chemokine primarily responsible for the homing and maintenance of hematopoietic stem cells in the bone marrow is CXCL12, the sole ligand of CXCR4. The CXCL12/CXCR4 axis is critical for B cell development and is required during lymphopoiesis, during differentiation in the germinal center reaction and drives homing of long-lived plasma cells to niches in the bone marrow. The CXCL12/CXCR4 couple is also necessary for myelopoiesis in the bone marrow. CXCL12, although considered mostly as a homeostatic chemokine, can be involved in inflammatory responses.\textsuperscript{[11]}

Another monogamous couple is constituted by the CXCL13/CXCR5 pair. The key function of the pair is the regulation of leukocyte trafficking within secondary lymphoid organs under healthy and inflammatory conditions. Thus CXCL13 guides CXCR5\textsuperscript{+} leukocytes, such as follicular helper T cells and B cells, to follicles. Furthermore, CXCL13 together with CXCL12 are the main chemokines for the segregation of the light and dark zone of germinal centers of secondary lymph nodes, respectively.\textsuperscript{[1]} In the germinal centers, B cells undergo affinity maturation and receptor class switch, a critical step in generating high affinity antibodies. In the dark zone, B cells proliferate, while in the light zone selection of high affinity antigen-specific B cells occurs.

Seven of the inflammatory CXC chemokines contain the ELR motif just in front of the first cysteine. All chemokines containing the ELR motif bind with similar affinity to CXCR2, the second receptor initially identified to bind CXCL8, while CXCR1 shows high affinity only for CXCL8.\textsuperscript{[32]} Interestingly, CXCL8, which binds to CXCR1 and CXCR2, induces stronger responses through CXCR1. The receptors are expressed on neutrophils and mediate the recruitment of front line immune cells to sites of inflammation.\textsuperscript{[41]}

Of the CC group the inflammatory chemokine receptors, CCR1, CCR2, CCR2 and CCR5 are most important for leukocyte recruitment during inflammation. With partially overlapping specificity, they share multiple CC chemokines as agonists. Locally produced cognate chemokines recruit lymphocytes and myelocytes to sites of infection. By contrast, expression of CCR4 is mainly restricted to T cells and appears to be involved in skin homing of the T lymphocytes. CCR4 binds two chemokines namely CCL17 and CCL22.\textsuperscript{[1]}

Next to their profound role in orchestrating leukocyte trafficking in homeostasis and inflammation, some chemokines are critical for development and neoplastic pathologies. A particular role is given for CXCL12, because its genetic ablation in mice leads to their perinatal death with pronounced defects in central nervous system development, heart function, B cell lymphopoiesis and vascularization. In addition, CXCR4, the sole functional receptor for CXCL12, is expressed on many cancer cells and often responsible for their metastatic behavior, such as homing
to bone marrow and lymphoid organs, respectively.[1]

In conclusion, the production – and the chemical synthesis in particular – of natively folded and pure chemokines, is instrumental for a variety of basic, translational, as well as clinical research programs. Researchers in Switzerland have taken a pioneering role in chemokine research which includes the co-discovery of the first chemokine[10] and substantially contributed to the understanding of chemokine functions in biomedicine. The field would greatly profit by integrating synthetic chemistry (again) in chemokine research programs.

Acknowledgements
We greatly acknowledge funding from the Swiss National Science Foundation (Sinergia grant 160719).